ATLASGAL – properties of a complete sample of Galactic clumps[★]

J. S. Urquhart,[1,2][★] C. König,[2] A. Giannetti,[2,3] S. Leurini,[2,4] T. J. T. Moore,[5]
D. J. Eden,[5] T. Pillai,[2] M. A. Thompson,[6] C. Braiding,[7] M. G. Burton,[7,8] T. Csengeri,[2]
J. T. Dempsey,[9] C. Figura,[10] D. Froebrich,[1] K. M. Menten,[2] F. Schuller,[2,11]
M. D. Smith[1] and F. Wyrowski[2]

Affiliations are listed at the end of the paper

ABSTRACT

The APEX Telescope Large Area Survey of the Galaxy (ATLASGAL) is an unbiased 870 µm submillimetre survey of the inner Galactic plane ($|\ell| < 60°$ with $|b| < 1°.5$). It is the largest and most sensitive ground-based submillimetre wavelength Galactic survey to date and has provided a large and systematic inventory of *all* massive, dense clumps in the Galaxy ($\geq 1000\,M_\odot$ at a heliocentric distance of 20 kpc) and includes representative samples of all of the earliest embedded stages of high-mass star formation. Here, we present the first detailed census of the properties (velocities, distances, luminosities and masses) and spatial distribution of a *complete* sample of ∼8000 dense clumps located in the Galactic disc ($5° < |\ell| < 60°$). We derive highly reliable velocities and distances to ∼97 per cent of the sample and use mid- and far-infrared survey data to develop an evolutionary classification scheme that we apply to the whole sample. Comparing the evolutionary subsamples reveals trends for increasing dust temperatures, luminosities and linewidths as a function of evolution indicating that the feedback from the embedded protoclusters is having a significant impact on the structure and dynamics of their natal clumps. We find that the vast majority of the detected clumps are capable of forming a massive star and 88 per cent are already associated with star formation at some level. We find the clump mass to be independent of evolution suggesting that the clumps form with the majority of their mass *in situ*. We estimate the statistical lifetime of the quiescent stage to be ∼5×10^4 yr for clump masses ∼$1000\,M_\odot$ decreasing to ∼1×10^4 yr for clump masses >$10000\,M_\odot$. We find a strong correlation between the fraction of clumps associated with massive stars and peak column density. The fraction is initially small at low column densities, but reaching 100 per cent for column densities above 10^{23} cm^{-2}; there are no clumps with column densities above this value that are not already associated with massive star formation. All of the evidence is consistent with a dynamic view of star formation wherein the clumps form rapidly and are initially very unstable so that star formation quickly ensues.

Key words: stars: formation – stars: massive – ISM: clouds – Galaxy: kinematics and dynamics – Galaxy: structure.

1 INTRODUCTION

Star formation is a Galaxy-wide phenomenon that takes place within the densest parts of giant molecular clouds (GMCs), which are themselves subject to a wide range of physical and environmental conditions (Heyer & Dame 2015). Observations of nearby spiral galaxies have revealed that their molecular gas and star formation are both tightly correlated and trace a well-defined spiral pattern (e.g. M51; Colombo et al. 2014). Analysis of the distribution of molecular gas and star formation in the Milky Way has revealed significant peaks at specific Galactocentric radii (e.g. Moore et al. 2012; Urquhart et al. 2014b) and found to be concentrated along loci in longitude–velocity space (Urquhart et al. 2014c), both of which confirm the spiral structure of the Milky Way and its connection to the ongoing star formation in the Galaxy. Establishing the statistical properties of star-forming regions and their connection with

---

Galactic structures are the aim of several programmes based on large-scale surveys of the Galactic plane (e.g. Hi-GAL; Molinari et al. 2010; Elia et al. 2017).

Although there is a strong correlation between the spiral arms, the molecular material and star formation, it is not currently clear how these are connected and what role the spiral arms play in the star formation process. Theoretical models of the spiral arms predict that they play a role in the formation of GMCs and enhance the overall star formation efficiency (SFE, Dobbs, Bonnell & Pringle 2006). The first point is supported by Moore et al. (2012) who found an increase in the density of molecular clouds and star formation in the spiral-arm regions. However, these authors found that the SFE, as measured from the luminosity-to-mass ratio (L/M), was not significantly enhanced in the spiral arms compared to interarm regions once two of the most extreme star-forming regions in the Galaxy were excluded (i.e. W49 and W51). This led Moore et al. (2012) to conclude that the increase in star formation found in the spiral arms was the result of source crowding rather than any direct influence from the arms themselves. This conclusion has been supported by Eden et al. (2013, 2015) who found little variation between the L/M ratio of star-forming clumps in the spiral arm and interarm locations along two different lines of sight.

Eden et al. (2012) also estimated the efficiency with which clouds in the spiral arms and interarm regions convert their mass into dense clumps, which is the first stage in the star formation process. This is known as the clump formation efficiency (CFE), and was calculated from the ratio of dense gas as determined from the dust emission mapped by the Bolocam Galactic Plane Survey (BGPS; Aguirre et al. 2011) and the cloud mass as calculated from the ^{13}CO (1–0) data provided by the Galactic Ring Survey (GRS; Jackson et al. 2006). Comparisons of the spiral arms and interarm regions revealed no significant variations between the two regions. Another more recent study investigated the star formation fraction (SFF; Ragan et al. 2016) as a function of Galactic position. This parameter is simply the ratio of the number of Hi-GAL clumps associated with star formation (i.e. a 70 μm point source) and the total number of Hi-GAL clumps. Analysis of the distribution of the SFF as a function of Galactocentric distance revealed a modest decrease in the SFF with increasing distance from the Galactic Centre (GC), but no significant increases in the SFF were observed towards the locations of the spiral arms. Elia et al. (2017) find no significant differences between the average evolutionary status of sources in locations within and between arms.

The absence of any significant variations in the CFE, SFF, SFE or trend as a function of Galactocentric position suggests that although the spiral arms may play an important role in the formation and/or concentration of molecular clouds, they do not play a role in the conversion of the cloud mass into dense clumps or in the subsequent star formation associated with the clumps. However, many of the studies discussed are averaging these parameters on kiloparsec scales, and so it may transpire that there are significant underlying localized variations in the SFE measures that are largely washed out when averaged over the large segments of the spiral arms; indeed, Moore et al. (2012) found significantly higher SFEs towards two of the most intense star-forming regions in the Galaxy (i.e. W49 and W51, Urquhart et al. 2014b).

Recent studies have revealed that a significant fraction of the star formation taking place in our Galaxy is concentrated in 18–30 star-forming complexes (e.g. ~50 per cent – Murray & Rahman 2010 and ~30 per cent – Urquhart et al. 2014b, respectively). These have very different physical properties and environmental conditions,

including some of the most extreme in the Galaxy. The remaining star formation is more evenly distributed across the Galactic plane, and while the physical properties and local environments tend to be less extreme, these regions are still responsible for a large fraction of the overall Galactic star formation rate (50–70 per cent). In order to obtain a clear picture of the role of the spiral arms and the influence of different environments on star formation, we need to conduct a comprehensive study of the star-forming properties of a significant fraction of the Galactic plane. The recent multiwavelength surveys of the Galactic plane (e.g. Hi-GAL, Molinari et al. 2010; WISE, Wright et al. 2010; MSX, Price et al. 2001; GLIMPSE, Benjamin et al. 2003; CORNISH, Hoare et al. 2012) provide the data to conduct just such a study.

The Atacama Pathfinder Experiment (APEX) Telescope Large Area Survey of the Galaxy (ATLASGAL) is the largest and most sensitive ground-based submillimetre survey of the inner Galactic Plane (Schuller et al. 2009; see Fig. 1 for survey coverage and sensitivity). The primary goal of this survey is to provide a large and systematic inventory of dense molecular clumps that includes representative samples of sources in all of the early embedded evolutionary stages associated with high-mass star formation. The Galactic distribution of dense gas has been investigated using ATLASGAL and reported by Beuther et al. (2012) and Csengeri et al. (2014). The ATLASGAL compact source catalogue (CSC; Contreras et al. 2013; Urquhart et al. 2014a) includes ~10 000 dense clumps that have been extracted from the processed emission maps. At a distance of 20 kpc, the ATLASGAL coverage includes all of the compact[1] located within the Solar circle and ~90 per cent of all dense molecular gas in the Galaxy (Urquhart et al. 2014a). Assuming a dust temperature of 15 K,[2] a distance of 20 kpc and a flux density of 0.3 Jy beam^{-1} (~5σ) corresponds to a minimum mass sensitivity of ~1000 M$_\odot$ beam^{-1} or ~700 M$_\odot$ pc^{-2}. The ATLASGAL CSC should therefore include the vast majority of all current and future high-mass star-forming clumps ($M_{\rm clump} > 1000$ M$_\odot$) in the Galaxy (assuming a standard initial mass function (IMF) and a typical SFE of ~30 per cent, Lada & Lada 2003).

Although the ATLASGAL CSC provides a large and statistically representative sample of massive dense clumps, analysis is limited as a priori distances are unknown, and these are required to derive important physical properties such as the clump mass, luminosity and Galactic distribution. In this paper, we derive distances and physical properties to nearly all ATLASGAL sources with a peak flux above 5σ and that are located away from the GC region (300° < ℓ < 355° and 5° < ℓ < 60°). We determine distances to all of the sources using a combination H I analysis, maser parallax and spectroscopic measurements taken from the literature and a clustering analysis to group clumps into associations. We combine this catalogue of clumps with mid- and far-infrared (IR) data (Hi-GAL, WISE, MSX and GLIMPSE) to determine their dust temperature from a fit to their spectral energy distribution (SED), masses and SFE (L/M). We used these parameters to investigate variations in the SFE as a function of Galactic position

---

[1] ATLASGAL is sensitive to angular scales up to 2.7 arcmin due to spatial filtering in the reduction process; this means that it is sensitive to different physical scales at different distances [i.e. cores within a kiloparsec, clumps out to 4–5 kpc and small clouds at larger distances given the typical clump full width at half-maximum (FWHM) size of 1 arcmin; see fig. 11 of Contreras et al. 2013], and high column density molecular gas ($N(H_2) \gtrsim 10^{22}$ cm^{-2}).

[2] This temperature is the average kinetic temperature determined from ammonia observations of quiescent clumps (i.e. Wienen et al. 2012).
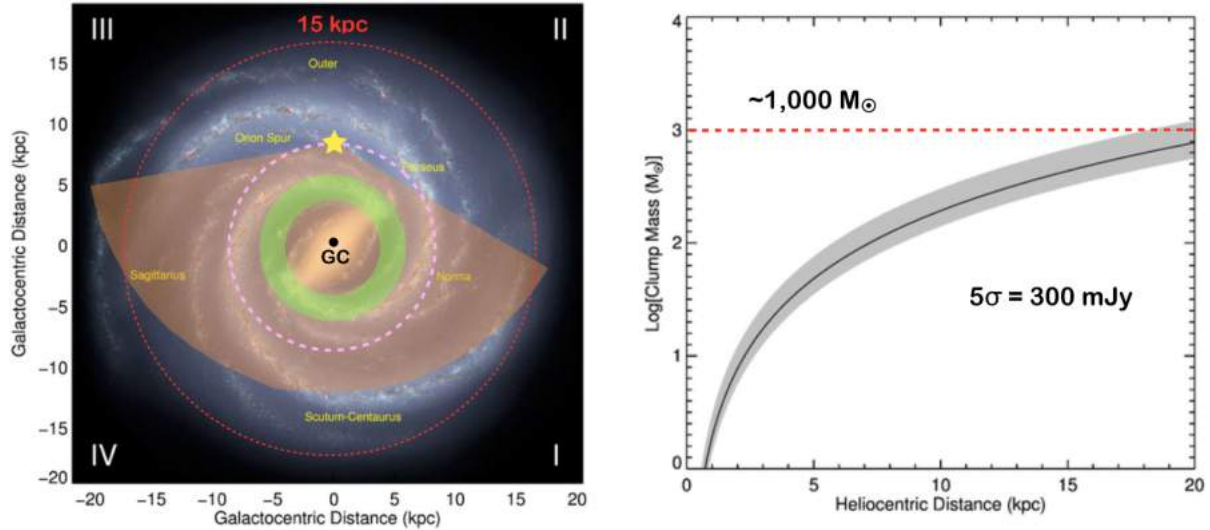
Figure 1. The image shown in the left-hand panel gives an overview of the coverage of the Galactic mid-plane provided by the ATLASGAL survey. The background is an artist's impression of how the Milky Way would appear to an external observer looking down from the Northern Galactic Pole [courtesy of NASA/JPL-Caltech/R. Hurt (SSC/Caltech)]. The yellow star indicates the position of the Sun, while the orange shaded region shows the coverage of the survey out to a heliocentric distance of 20 kpc. The Solar circle and the 5 kpc molecular ring are shown by the pink dashed line and the light green band centred on the GC. The right-hand panel shows the mass sensitivity limits of the survey as a function of heliocentric distance for a 5σ flux sensitivity of 300 mJy beam^{-1}. The black curve shows the mass sensitivity assuming a dust temperature of 20 K, while the grey shaded region shows the uncertainty, allowing ± 5 K in temperature.

and environment on physical scales of individual clumps and whole complexes.

This is the fourth paper in a series that focuses on the properties of the Galactic population of dense clumps identified by ATLASGAL. The three previous papers have focused on clumps associated with massive star formation tracers: methanol masers (Urquhart et al. 2013a; hereafter Paper I), UC HII regions (Urquhart et al. 2013b; hereafter Paper II) and massive young stellar objects (Urquhart et al. 2014c; hereafter Paper III). Combined these three papers identified ∼1300 massive star-forming (MSF) clumps, however, this sample only make up a relatively small fraction of the total population of clumps identified by ATLASGAL (∼15 per cent).

In this paper, we derive the physical properties for the whole population of clumps and investigate their star-forming properties and Galactic distribution. The structure of the paper is as follows: in Section 2, we give an overview of the molecular line surveys used to assign radial velocities to the clumps and describe the APEX observations that were made to increase the completeness of the velocity measurements for the ATLASGAL sample of clumps; in Section 3, we describe the methods used to determine distances to the clumps and describe the clustering analysis used to group clumps into star-forming complexes; in Section 4, we briefly discuss the method used to fit the SED and estimate their current level of star formation activity; in Section 5, we derive the physical properties of the clumps and in Section 6, we investigate the Galactic distribution of clumps and complexes and their correlation with the spiral arms; in Section 7, we derive empirical star formation relationships and discuss what these can tell us about the star formation process; in Section 8, we investigate the evolution of the clumps over their lifetimes; in Section 9, we compare the properties of the most active regions in the Galaxy and assess their contribution to the Galactic dense gas reservoir and star formation; and finally in Section 10, we summarize our results and present our main conclusions.

2 RADIAL VELOCITY MEASUREMENTS

A first step to determining a source's distance and physical properties is its radial velocity with respect to the local standard of rest (LSR, v_{lsr}). This velocity can be used in conjunction with a model of the Galactic rotation to obtain a kinematic distance (see Section 3). The radial velocities of molecular clumps can be measured from molecular line observations (e.g. CO, NH$_3$, CS, etc.) and these are readily available for many of the ATLASGAL clumps from a number of Galactic plane surveys [e.g. GRS, Mopra CO Galactic plane Survey (MGPS; Burton et al. 2013; Braiding et al. 2015), the Three-mm Ultimate Mopra Milky Way Survey (ThrUMMS) (Barnes et al. 2015), SEDIGISM (Schuller et al. 2017), COHRS (Dempsey, Thomas & Currie 2013) and CHIMPS (Rigby et al. 2016)] and large targeted observational programmes towards selected samples (e.g. MALT90 (Jackson et al. 2013), the Red MSX Source (RMS) survey (Urquhart et al. 2007, 2008, 2011, 2014b), BGPS (Dunham et al. 2011a; Schlingman et al. 2011; Shirley et al. 2013) as well as dedicated ATLASGAL follow-up observations (e.g. Wienen et al. 2012; Csengeri et al. 2016a; Kim et al. 2017).

2.1 Archival molecular line surveys

We began this process of assigning velocities by matching ATLAS-GAL clumps with all large molecular line catalogues reported in the literature and assigning a velocity to a source where the pointing centre of the observation was found to overlap with the structure of the source. Where multiple observations were available the transition with the highest critical density (e.g. NH$_3$ and HNC over CO) was preferred as these are less affected by multiple components arising from diffuse molecular clouds lying along the same line of sight as the target source. In Table 1, we present a summary of the various surveys used to assign velocities to the vast majority of ATLASGAL sources of interest to this study.

For sources where a velocity was not already available, we extracted spectra directly from the various survey data cubes

Table 1. Surveys used to assign radial velocities to 7809 ATLASGAL CSC sources.

Survey	Number of sources	Molecular transitions	References
ATLASGAL	1101	^{13}CO (2–1)/C^{18}O (2–1)	This paper
ATLASGAL	693	NH$_3$ (1,1)	Wienen et al. (2012)
ATLASGAL	299	NH$_3$ (1,1)	Wienen et al. (2017)
ATLASGAL	154	N$_2$H$^+$ (1–0)	Urquhart et al. (in preparation)
ATLASGAL (70 μm dark)	50	C^{18}O (2–1)/N$_2$H$^+$ (1–0)	Pillai et al. (in preparation)
ATLASGAL	62	N$_2$H$^+$ (1–0)	Csengeri et al. (2016a)
MALT90^a	1205	HNC/N$_2$H$^+$ (1–0)	Jackson et al. (2013)
BGPS	959	N$_2$H$^+$/HCO$^+$ (1–0)	Shirley et al. (2013)
ThrUMMS^b	960	CO (1–0)	Barnes et al. (2015)
COHRS	580	^{12}CO (3–2)	Dempsey et al. (2013)
MSGPS	499	CO (1–0)	Burton et al. (2013)
RMS	407	^{13}CO/NH$_3$	Urquhart et al. (2007, 2008, 2011, 2014b)
GRS^c	293	^{13}CO (1–0)	Jackson et al. (2006)
BGPS	153	NH$_3$ (1,1)	Dunham et al. (2011b)
UC H II cand.	148	CS (2–1)	Bronfman, Nyman & May (1996)
CHIMPS	83	^{13}CO (3–2)	Rigby et al. (2016)
HOPS	60	NH$_3$ (1,1)	Purcell et al. (2012)
IRDCs	55	CS (2–1)	Jackson et al. (2008)
SEDIGISM	43	^{13}CO (2–1)	Schuller et al. (2017)
M8	3	CO (3–2)	Tothill et al. (2002)
IR-quiet clumps	2	NH$_3$ (1,1)	Hill et al. (2010)

^{a}Spectra were extracted towards the submillimetre peak of each ATLASGAL source from the 3 arcmin×3 arcmin maps of the brightest four transitions covered by this survey. These maps have been downloaded from the reduced and calibrated data products available from the Australia Telescope Online Archive (http://atoa.atnf.csiro.au/).
^{b}Source: http://www.astro.ufl.edu/~peterb/research/thrumms/
^{c}Source: http://www.bu.edu/galacticring/

(i.e. MALT90, ThrUMMS and the GRS) or from spectra which were provided by the survey teams (i.e. COHRS, MGPS and CHIMPS); these data are fully reduced and calibrated and so no further processing was required. The extracted spectra were fitted with Gaussian profiles using an automatic routine. The spectra and resulting fits were inspected to ensure the data quality was good and the reliability of the resulting fits. Poor quality data (e.g. those contaminated by emission in the off-source positions or strong baseline ripples) were discarded and poor reliability fits to the emission profile were refit by hand (this tended to occur in complex regions where multiple velocity components overlap with each other, however, this only affected approximately 5 per cent of sources).

When the quality of the data and fits were considered reliable they were used to assign velocities to the clumps using the following criteria: (1) if only a single component was detected, then the peak velocity of the molecular line was assigned to the source; (2) if multiple components are detected, but all of the strongest components (integrated line intensity) are within 10 km s^{-1} of each other, then the velocity of the strongest component was selected; (3) if multiple components were detected, then the velocity of the component with the largest integrated line intensity was used, provided it was at least twice as large as the next strongest component. If none of these criteria were satisfied, then no velocity was allocated and additional observations were obtained (as described in the following subsection). In Table 1, we give a summary of the surveys used and number of velocities each one provides.

2.2 ATLASGAL follow-up line surveys

There have been a number of follow-up molecular line surveys that have provided velocity information for ~1000 sources (e.g.

ammonia observations: Wienen et al. 2012 and Wienen et al. 2017; 3-mm line surveys: Csengeri et al. 2016a; Kim et al. 2017; Giannetti et al. 2017). Velocity information for a further sixty-seven 70 μm dark sources presented here was obtained from a dedicated follow-up of 70 μm dark sources with the APEX 12 m (C^{18}O 2–1) and IRAM 30 m (N$_2$H$^+$ 1–0) telescopes (the APEX and IRAM 30 m telescopes have similar beam sizes at these frequencies; FWHM ~ 28 and 26 arcsec, respectively). The observations are part of a separate follow-up programme and a more detailed description will be presented in a forthcoming paper (Pillai et al. in preparation).

The results of these dedicated follow-up programmes and results taken from the literature provide velocity information to many thousands of ATLASGAL sources, however, there were still approximately 1200 sources for which no velocity measurement were available. It is necessary to obtain velocities for as many sources as possible in order to improve the completeness of our velocity coverage and the reliability of any statistical analysis based on this sample. We therefore instigated a bad-weather backup programme on the APEX telescope that observed all sources with a missing velocity assignment. These observations are reported in Appendix A1 and provide radial velocities for 1115 clumps of which 1101 were previously unknown.

2.3 Completeness

There are 8002 ATLASGAL sources located in the region of interest considered here and through the steps described in the previous subsection we have determined a velocity to 7809 sources (~98 per cent of the sample). In Table 2, we give the source names, velocities, molecular transition used to determine the velocity and

Table 2. Summary of the kinematic distance analysis.

ATLASGAL CSC name	v_{lsr} (km s^{-1})	v_{lsr} references	Reid distances			Adopted kinematic solution			Cluster	
			Bayesian (kpc)	Near (kpc)	Far (kpc)	Distance (kpc)	Solution Flag[a]	Distance reference[b]	name	Cluster distance (kpc)
AGAL005.001+00.086	2.1	This paper	2.9	0.7	17.0	17.0	(vi)	–	–	–
AGAL005.041−00.097	46.3	This paper	10.8	6.6	10.0	10.8	(vi)	–	G005.087−00.097	10.8
AGAL005.049−00.192	6.1	This paper	2.9	1.7	15.4	2.9	(vi)	–	–	–
AGAL005.076−00.091	45.1	Jackson et al. (2013)	10.8	6.5	10.1	10.8	(vi)	–	G005.087−00.097	10.8
AGAL005.094−00.104	45.7	Jackson et al. (2013)	10.8	6.6	10.1	10.8	(vi)	–	G005.087−00.097	10.8
AGAL005.139−00.097	44.1	This paper	10.8	6.4	10.2	10.8	(vi)	–	G005.087−00.097	10.8
AGAL005.184+00.159	–	–	–	–	–	–	(x)	–	–	–
AGAL005.192−00.284	8.0	This paper	2.9	2.0	14.9	2.9	(vi)	–	–	–
AGAL005.202−00.036	0.7	This paper	16.0	0.1	17.7	16.0	(vi)	–	–	–
AGAL005.321+00.184	–	–	–	–	–	–	(x)	–	–	–

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/.
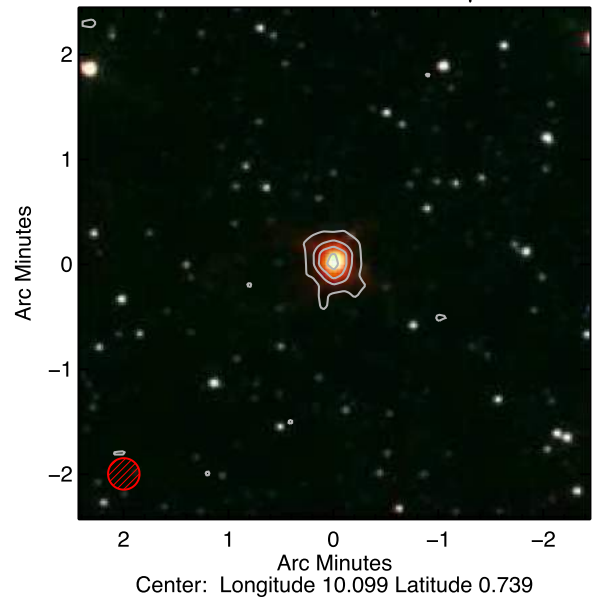[a]The distance solution flags refer to the different steps described in Appendix B1 and Table B2.
[b]References for distance solutions: (1) Anderson & Bania (2009), (2) Araya et al. (2002), (3) Battisti & Heyer (2014), (4) Urquhart et al. (2013b), (5) Dunham et al. (2011b), (6) Fish et al. (2003), (7) 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).
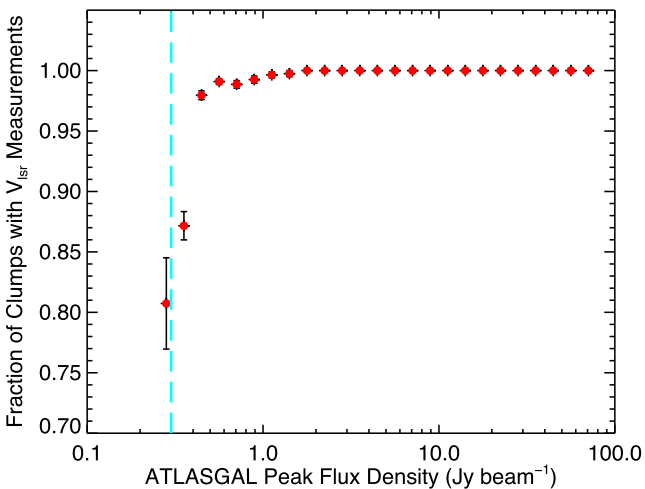
Figure 2. Plot of the ratio of sources with velocities as a function of their peak flux densities. The bin size for both plots is 0.1 dex. The errors shown in the lower panel have been estimated using binomial statistics. The vertical dashed line indicates the 5σ sensitivity limit of the ATLASGAL survey.

AGAL010.099+00.739 IRAC + 870 μm Contours
Center: Longitude 10.099 Latitude 0.739

Figure 3. Three colour image of an evolved star candidate (this is constructed using the GLIMPSE 3.6, 5.8 and 8 μm images, which are coloured blue, green and red, respectively). The contours show the submillimetre dust emission. The IR and dust emission both show a point-like distribution and there is no ^{13}CO (2–1) emission detected.

the reference from where these data originated. In Fig. 2, we show the proportion of sources with determined radial velocities as a function of the peak flux distribution. This plot shows that we are complete for nearly all sources with a peak flux density $\simeq$0.5 Jy beam^{-1}. Visual inspection of IR and submillimetre maps for the ~200 weak sources for which we have not been able to assign a velocity reveals that many of them have a wispy structure and often appear to be part of the diffuse outer envelopes of larger regions. Molecular line emission is actually detected towards many of these diffuse sources; however, this emission is normally in the ^{12}CO and ^{13}CO (1–0) transitions where multiple peaks of approximately equal intensity are detected. It therefore seems likely that many of the clumps with undetermined velocities are of little interest in terms of their current star formation potential.

While the vast majority of sources identified in the ATLAS-GAL CSC are dense molecular clumps, the CSC is also sensitive to the submillimetre emission from the dusty envelopes that surround evolved stars. Although these dusty envelopes are warm enough to produce significant dust emission, they do not generally have sufficiently high column densities to be detected in molecular line observations [^{12}CO (1–0) and (2–1) typically less than 1 K; Loup et al. 1993, see also Fig. 3 for an example of an evolved star candidate]. The possible contamination of the ATLASGAL catalogue
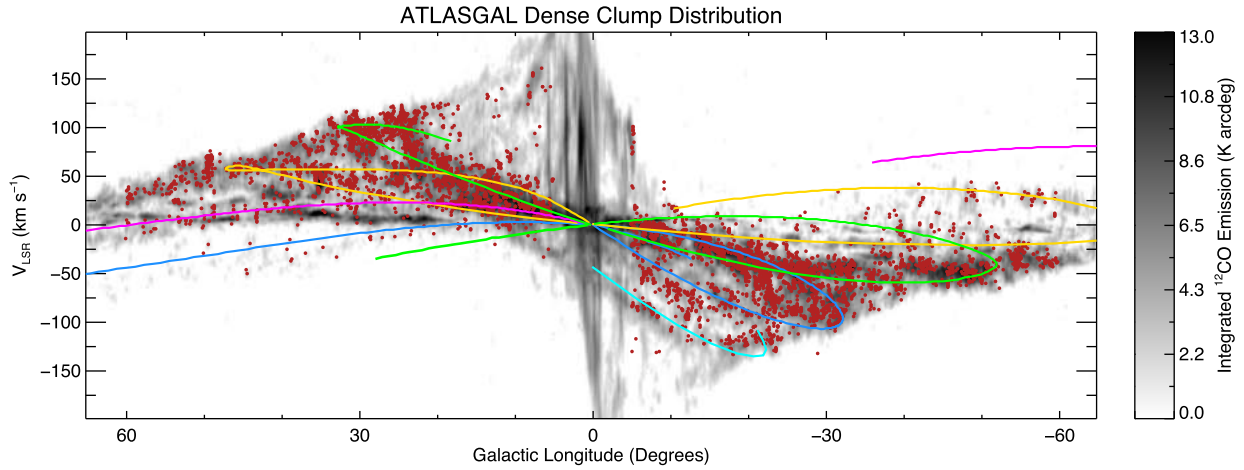
Figure 4. Longitude–velocity distribution of all ATLASGAL source for which we have been able to assign a velocity. The grey-scale image shows the distribution of molecular gas as traced by the integrated ^{12}CO (1–0) emission for comparison (Dame et al. 2001). The location of the spiral arms are shown as curved solid lines, coloured to identify the individual arms. The positions of the four main spiral and local arms have been taken from model by Taylor & Cordes (1993) and updated by Cordes (2004), while the position of the near 3-kpc arm has been taken from Bronfman et al. (2000). Colours: Norma/Outer → blue; Perseus → magenta; Sagittarius → yellow; Scutum-Centaurius → green and near 3-kpc arm → cyan.

by diffuse gas and warm envelopes of evolved stars suggests that our sample is likely as complete as possible in terms of velocity determination.

Fig. 4 presents a longitude–velocity (ℓv) diagram that shows the positions of all of the ATLASGAL CSC for which we have determined a velocity. The background image shown in this figure is the integrated ^{12}CO (1–0) map from Dame, Hartmann & Thaddeus (2001) that traces the large-scale Galactic distribution of molecular gas. Comparison between the CO emission and dense gas traced by the dust emission reveals a good correlation between the two. This plot also shows the loci of the spiral arms from Taylor & Cordes (1993) and Cordes (2004) and the loci of the 3 kpc expanding arm from Bronfman et al. (2000). The majority of ATLASGAL sources are seen to be tightly clustered around the four-arm Milky Way model spiral-arm loci. We will revisit the Galactic distribution in Section 6.

3 DISTANCES AND ASSOCIATIONS

A key element required for determining the physical properties of dense clumps identified by ATLASGAL is their heliocentric distance. This has been achieved using a multistep approach that includes adopting maser parallax and spectroscopic distances where available, determining kinematic distances to all remaining sources for which we have derived a radial velocity, and resolving any resulting distance ambiguities using archival H I data. These steps are described in detail in Appendix B and outlined in the flow chart presented in Fig. 5.

Using these steps, we have determined kinematic distance solutions to ~7000 ATLASGAL clumps, with ~80 per cent being considered to be reliable. Comparing these with distances solutions recently published in the literature, we have found agreement in 73 per cent of cases (see Table B3 for a detailed breakdown). This analysis has resulted in 78 per cent of sources being placed at the near distance and the remaining 22 per cent being placed at the far distance, which is consistent with other studies (cf. Eden et al. 2012). We have not performed a direct comparison with the recent work presented by Elia et al. (2017) as only ~40 per cent

of their distances were the result of distance ambiguity resolution: the remainder are arbitrarily placed at the far distance and so the samples are not comparable as their results are likely to be biased to systematically larger distances.

Although the steps outlined in Fig. 5 have been able to apply distances to almost 90 per cent of the sample, there are still a significant number for which we have been unable to determine a velocity, or for which we have been unable to resolve the distance ambiguity. In an effort to determine distances to these sources and to provide a consistency check on the kinematic solutions derived from the H I data, we have performed a friends-of-friends clustering analysis; this is described in detail in Appendix C1. This analysis has identified 776 clusters of ATLASGAL sources, many of the largest of which, are associated with some of the most well-known star-forming regions in the Galaxy.

In Fig. 6, we present a few examples of the clusters identified by the friends-of-friends analysis; the distributions of clumps are overlaid on 8 μm *Spitzer* images taken as part of the GLIMPSE legacy survey. The 8 μm images are very sensitive to emission from polycyclic aromatic hydrocarbons (PAHs; e.g. Urquhart et al. 2003) which are excited in photo-dominated regions often associated with H II regions and are, therefore, excellent tracers of star-forming regions. Areas of extinction are very useful for tracing the colder quiescent dense clumps where future star formation is expected to take place. These mid-IR images are therefore able to trace both the large-scale environments of active star formation regions and trace the distribution of dense clumps identified by ATLASGAL. The examples presented in Fig. 6 nicely illustrate these points and demonstrate the reliability of the friends-of-friends analysis.

3.1 Distance summary

In total, we have identified 776 clusters that are correlated in $\ell b v$ and distance; these are associated with 6620 clumps, including 75 clumps for which no velocity was available and a further 549 clumps for which we were previously unable to determine a distance. We have either used the H I self-absorption (H I SA) technique to solve the distance ambiguity or extracted more reliable distances from the
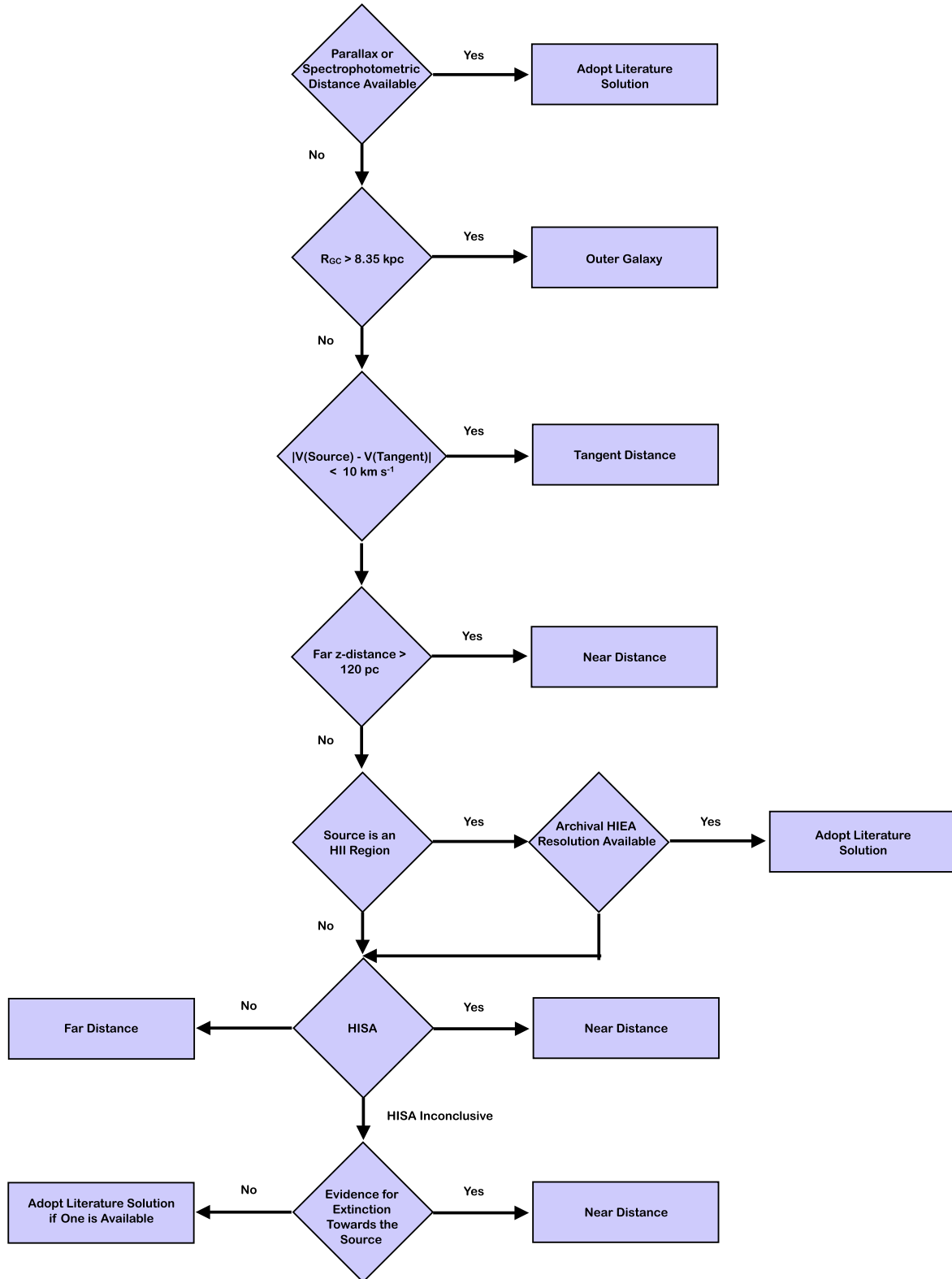
Figure 5. Flow chart showing the criteria used to determine distance to ATLASGAL CSC clumps.

literature for an additional 1150 sources. In total, this combination of methods has provided distances to 7770 clumps (∼97 per cent of the full CSC sample in the region covered here). In Fig. 7, we show the Galactic distribution of the clusters and isolated clumps that is derived from the distances discussed in this section.

4 SPECTRAL ENERGY DISTRIBUTIONS

We follow the general procedures of König et al. (2017) to obtain the SEDs of the clumps, but apply some changes where necessary to be able to obtain the SEDs for the complete CSC in a fully
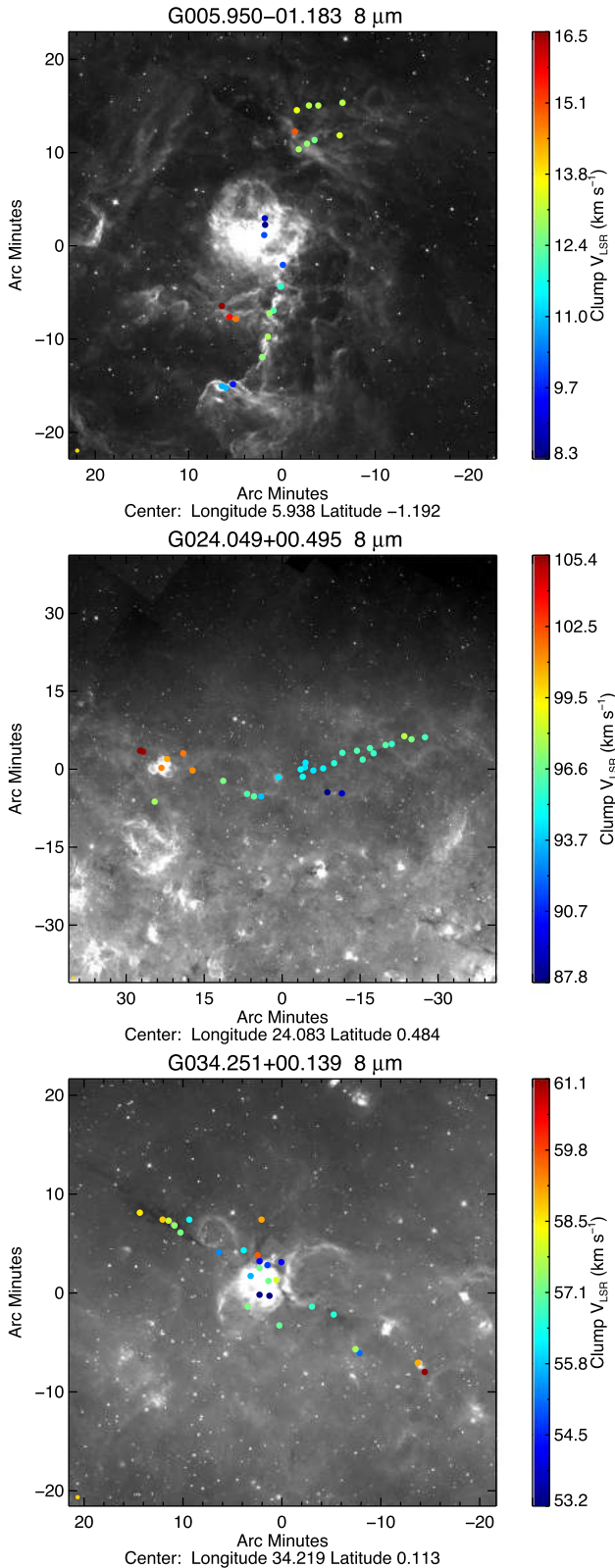
Figure 6. Examples of clusters identified by the friends-of-friends analysis. The background image is the 8 μm emission taken from the GLIMPSE survey. The filled circles show the positions of ATLASGAL CSC objects, the colours of which provide an indication of their velocities (see colour bars for values). In the upper, middle and lower panels, we show the Lagoon nebula (M8), the GMF G023.985+0.479 (Li et al. 2016); see also CFG024.00+0.48 (Wang et al. 2015) and G34.

automated way. We use the ATLASGAL (Schuller et al. 2009) 870 μm emission maps as well as the Hi-GAL (Molinari et al. 2010) images available for the PACS (Poglitsch et al. 2010) 70 and 160 μm and the SPIRE (Griffin et al. 2010) 250, 350 and 500 μm bands to reconstruct the SEDs cold dust component. We estimate the emission from a hot and embedded component, likely to be associated with a more evolved part of the clump, using MSX (Egan et al. 2003) emission maps at 8, 12, 14 and 21 μm, as well as images from the WISE (Wright et al. 2010) 12 and 24 μm bands. After reprojecting the images into the same projection as the ATLASGAL images, we converted all maps to Jy pixel^{-1} units and extracted 5×5 arcminute-sized images for each source. From these images, we obtained the photometric data in all bands and subsequently fit the SEDs by a single- or two-component model. Examples of these fits are shown in Fig. 8 and a detailed description of the photometric measurements and fitting procedure are provided in Appendix D.

4.1 Evolutionary sequence

Following the classification scheme introduced in König et al. (2017) and confirmed in Giannetti et al. (2017), we find that the majority of sources are actively forming stars with 5275 clumps (∼66 per cent) being classified as mid-IR bright (i.e. associated with a 21–24 μm point source with a flux > 2.6 mJy). We make a distinction in this sample between clumps previously associated with a massive star formation tracer (MSF clumps; i.e. radio bright H II regions, massive young stellar objects and methanol masers; see Paper III for details). We will refer to the other mid-IR bright clumps as young stellar object (YSO)-forming clumps: these subsamples contribute ∼16 and 52 per cent of the sample, respectively. Another 1640 sources (∼21 per cent) are classified as mid-IR weak but far-IR bright, making them likely to be in an earlier protostellar phase; we refer to these as protostellar clumps. The remaining 946 clumps (∼12 per cent) are classified as 70-μm weak, indicating that they are the youngest objects in the sample, likely to be in a starless or pre-stellar phase, however, a number of these 70 μm sources have been found to be associated with molecular outflows (e.g. Traficante et al. 2017) and so this fraction should be considered as an upper limit. In Table 3, we present a summary of evolutionary types identified and the numbers of sources associated with each. On the face of it, this suggests that 88 per cent of all ATLASGAL clumps are currently involved in the star formation process and the pre-stellar phase is relatively short.

Taken together, these four subsamples represent an evolutionary sequence for the formation of massive stars and clusters (König et al. 2017) and therefore comparing the physical properties of these different phases will provide some insight into the processes associated with star formation.

5 PHYSICAL PROPERTIES

In this section, we will use the distances assigned in Section 3 and the parameters derived from the analysis of the SEDs presented in Section 4 to determine physical properties for all of the clumps.

5.1 Analysis tools: sample comparison

We use the two-sample Kolmogorov–Smirnov (KS) test to compare the similarities between the different evolutionary types for the various derived properties. This is a non-parametric test that compares
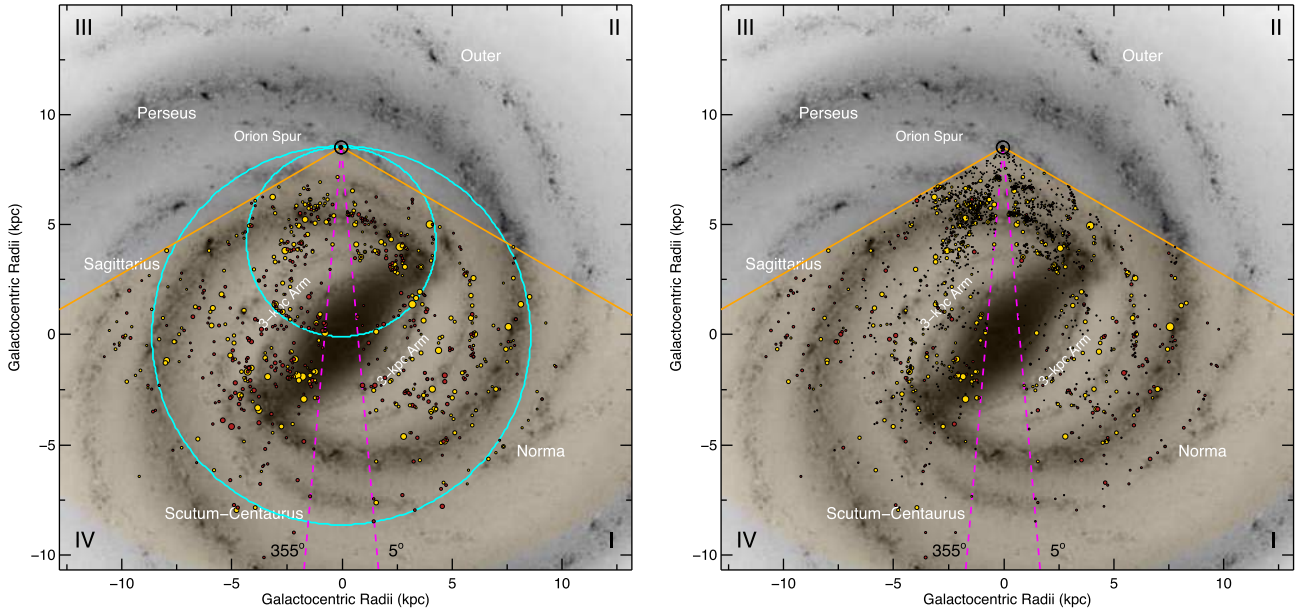
Figure 7. 2D images showing expected large-scale features of the Milky Way as viewed from the Galactic pole. The positions of ATLASGAL sources have been overlaid to facilitate comparison of their distribution to the large-scale structure of the Galaxy. The yellow circles show positions of the clusters identified by the friends-of-friends analysis described in Appendix C1, while the red circles show the positions of individual clumps. The sizes of the circles in the left- and right-hand panels give an indication of the masses and luminosities of the clumps and clusters, respectively. 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 smaller of the two cyan dot–dashed circles represents the locus of tangent points, while the larger circle traces the solar circle. 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 towards the GC where distances are not reliable. The distances of the ATLASGAL sources have been determined using a mixture of the Reid et al. (2014) method to calculate the near and far distances, while taking advantage of the additional constraints provided by the Reid et al. (2016) Bayesian maximum likelihood to better constrain sources with velocities close to the Sun's, the tangent position and outside the Solar circle.

the empirical cumulative distribution functions for the two samples and measures the largest difference between them [this is referred to as the KS statistic (D)] and its associated confidence value referred to as the *p*-value. The null hypothesis is that both samples are drawn from the same parent population; however, this can be rejected if the *p*-value is smaller than 3σ (i.e. < 0.0013), allowing us to conclude that there is sufficient evidence to consider the samples to be drawn from different populations. The KS test is useful as it is sensitive to differences in both the location and shape of the cumulative distributions of the different samples.

5.2 Distance distribution

We present the distance distribution for the full sample and the subsample of MSF clumps in Fig. 9. The heliocentric distribution reveals that the vast majority of clumps are relatively nearby (<5 kpc), with the strongest and second-strongest peaks likely to be associated with the segments of the Sagittarius and Scutum-Centaurus arms that lie between us and the GC. There is another significant peak between ∼10 and 11 kpc, which coincides with the end of the far side of the Galactic bar and the far-side Scutum-Centaurus and Perseus arms. The distribution of the MSF clumps shows a similar distribution of peaks but the relative ratio of MSF clumps is much lower for the nearer distance bins, revealing that although the majority of clumps are located nearby, few are associated with MSF and are therefore likely to be predominantly lower mass clumps.

We note that the highest density of sources occurs between 2 and 4 kpc for both the full sample of clumps and for the MSF clumps, and so we define clumps in this range as the *distance-limited sample*. In the following subsection, we provide histograms of the full distribution and for the distance-limited samples, together with associated cumulative distribution plots to aid with comparison of the different evolutionary phases. These show the distributions of parameters for the distance-limited subsample of each evolutionary phase (Figs 10, 11, 13 and 14). A summary of the derived physical parameters for the whole sample and the different evolutionary types is presented in Table 4 and measurements of individual clumps are given in Table 5.

5.3 Clump sizes

The physical sizes of the clumps have been calculated using their effective angular radii and the distances discussed in the previous section. A significant fraction of the ATLASGAL CSC is unresolved, and so we are only able to determine sizes for ∼60 per cent of the sample. We show the radius distribution for the 4714 clumps that are resolved in the upper panel of Fig. 10. The mean size of the clumps is 0.72±0.01 pc.

In the lower panel of Fig. 10, we show the size distributions of the four evolutionary distance-limited subsamples. This plot reveals a trend of increasing clump radius as a function of advanced evolutionary stage. There is no significant difference between the sizes of the pre-stellar and protostellar clumps; however, there is a visible trend for more evolved clumps to be larger than the preceding phase: this is confirmed by the KS test. This difference in size is likely to be the result of an observational bias as the extended envelops of
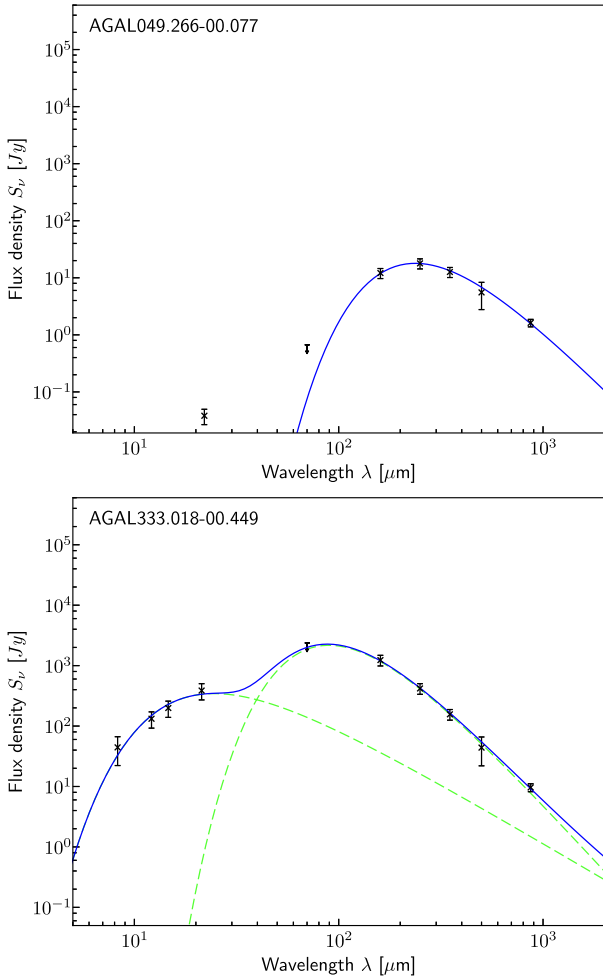
Figure 8. Example SEDs showing a source fitted with single-component greybody model (upper panel) and with a two-component model (lower panel). The photometry measurements are shown as crosses with their associated uncertainties indicated by the vertical limits dissecting them. In the upper panel, the greybody fit to the data is shown in blue while, in the lower panel, the blue curve shows the combined fits to these data of the individual cold greybody and warm blackbody functions; these are shown by the two green dashed curves.

Table 3. Summary of evolutionary types identified from the SED analysis and our previous work (Paper III).

| Evolutionary type	Number of sources	Fraction of total	Notes	
MSF	1222	0.16	} 0.88	Star forming
YSO	4053	0.52		
Protostellar	1640	0.21		
Quiescent	946	<0.12	} <0.12	Non-star forming

warmer and more evolved objects are more readily detectable (this is discussed in more detail in Section 8).

5.4 Dust temperature and bolometric luminosity

The dust temperature is derived directly from the greybody fit to the submillimetre dust emission, as described in Section 4. The distribution of these temperatures is shown in the upper and lower
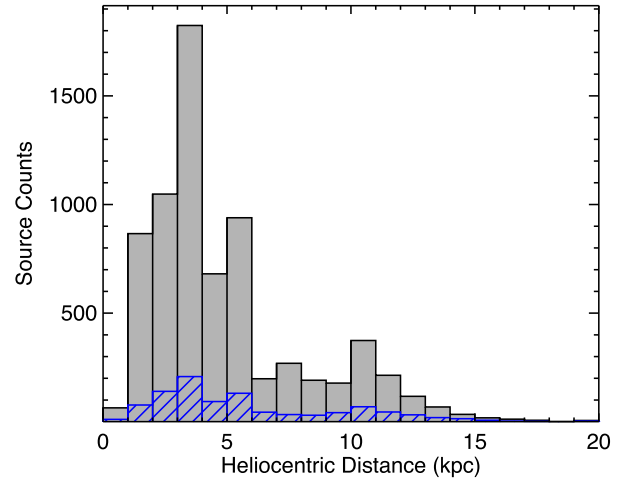
Figure 9. Heliocentric distance distribution of all clumps (grey) and clumps associated with massive star formation (blue). The bin size used is 1 kpc.
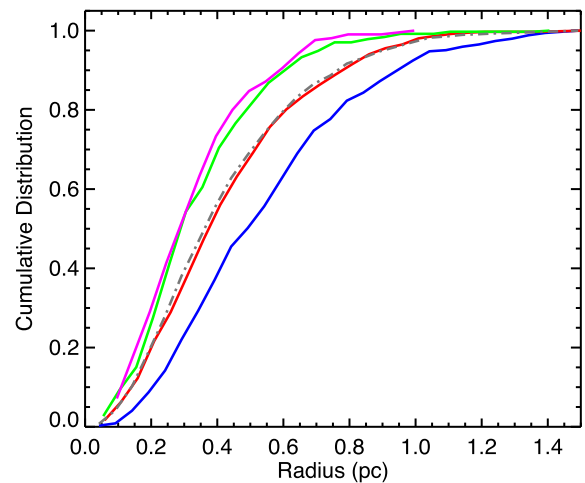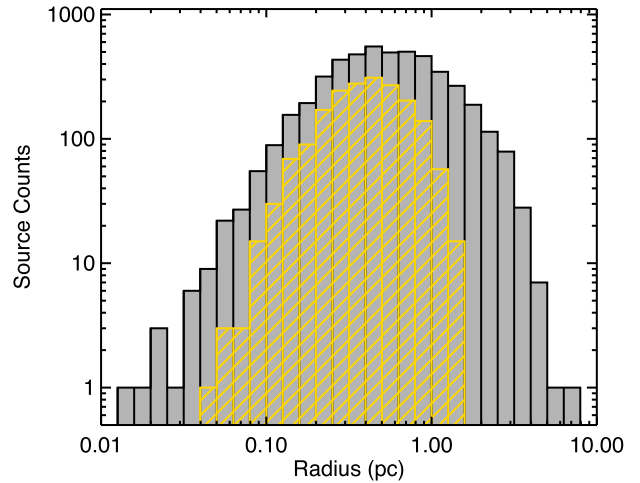
Figure 10. The radius frequency distribution for all clumps (grey) and for the distance-limited sample (yellow) is shown in the upper panel. The bin size is 0.1 dex. The lower panel shows the cumulative radius distribution for the whole sample as the grey dashed-dotted curve. The distance-limited subsamples of the four different evolutionary types (pre-stellar, protostellar, mid-IR bright and MSF – MMB, massive-YSO or H II region) are represented by the magenta, green, red and blue curves, respectively.

Table 4. Summary of physical properties of the whole population of clumps and the four evolutionary subsamples identified. 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, in Columns (6)–(8), we give the median and minimum and maximum values of the samples. Note we do not breakdown the results of the scaleheight analysis into evolutionary groups as there is no significant difference between them.

| Parameter	Number	$\bar{x}$	$\frac{\sigma}{\sqrt{(N)}}$	σ	x_{med}	x_{min}	x_{max}
Temperature (K)	7861	19.52	0.07	5.80	18.60	7.90	56.10
MSF	1222	24.58	0.14	5.06	24.00	12.60	56.10
YSO	4053	20.93	0.08	5.25	20.10	8.60	53.00
Protostellar	1640	15.52	0.07	2.80	15.20	8.30	30.20
Quiescent	946	13.88	0.12	3.64	13.30	7.90	48.70
Radius (pc)	4836	0.71	0.01	0.62	0.52	0.01	7.73
MSF	1017	1.00	0.02	0.77	0.78	0.01	7.73
YSO	2552	0.70	0.01	0.59	0.52	0.02	4.86
Protostellar	840	0.53	0.01	0.43	0.40	0.05	2.89
Quiescent	427	0.47	0.02	0.37	0.38	0.03	2.25
log[luminosity (L$_\odot$)]	7614	2.95	0.01	1.03	2.89	-0.30	6.91
MSF	1191	4.01	0.03	0.93	4.03	-0.30	6.91
YSO	3922	3.12	0.01	0.84	3.10	0.43	6.24
Protostellar	1580	2.29	0.02	0.69	2.24	-0.30	4.62
Quiescent	921	1.99	0.02	0.72	1.93	0.00	4.83
log[clump mass (M$_\odot$)]	7614	2.68	0.01	0.65	2.69	-1.00	5.04
MSF	1191	2.95	0.02	0.71	3.02	-1.00	5.04
YSO	3922	2.62	0.01	0.66	2.64	-0.40	4.72
Protostellar	1580	2.64	0.01	0.58	2.62	0.18	4.36
Quiescent	921	2.66	0.02	0.57	2.64	0.81	4.34
log[N(H$_2$) (cm^{-2})]	7861	22.33	0.00	0.29	22.30	21.58	24.02
MSF	1222	22.55	0.01	0.39	22.50	21.68	24.02
YSO	4053	22.25	0.00	0.26	22.22	21.58	23.54
Protostellar	1640	22.33	0.01	0.21	22.31	21.72	23.27
Quiescent	946	22.40	0.01	0.19	22.39	21.91	23.20
L_{bol}/M_{clump} ratio (L$_\odot$/M$_\odot$)	7614	18.76	0.75	65.17	4.47	0.03	2519.50
MSF	1191	38.90	2.75	94.88	19.03	0.44	2519.50
YSO	3922	22.61	1.11	69.75	7.30	0.08	1944.37
Protostellar	1634	2.42	0.12	5.02	1.27	0.04	95.08
Quiescent	933	6.65	1.58	48.37	0.58	0.03	833.62
log[mass surface density (M$_\odot$ pc^{-2})]	4838	2.92	0.00	0.35	2.87	1.86	4.34
MSF	1019	2.80	0.01	0.26	2.78	1.92	4.05
YSO	2552	2.85	0.01	0.33	2.81	1.86	4.07
Protostellar	840	3.10	0.01	0.34	3.05	2.30	4.34
Quiescent	427	3.21	0.02	0.35	3.17	2.45	4.32
Scaleheight (z) (pc)	7123	-4.34	0.36	30.63	-5.50	-345.90	235.40

Table 5. Derived clump parameters.

CSC name	Evolution type	Peak flux (Jy beam^{-1})	Integrated flux (Jy)	V_{LSR} (km s^{-1})	Distance (kpc)	R_{GC} (kpc)	Radius (pc)	T_{dust} (K)	log[L_{bol}] (L$_\odot$)	log[M_{clump}] (M$_\odot$)	log[N(H$_2$)] (cm^{-2})
AGAL006.264−00.506	Quiescent	0.44	1.49	22.8	3.0	5.4	0.35	12.7	1.607	2.186	22.361
AGAL006.268−00.749	YSO	0.48	3.29	21.5	3.0	5.4	0.70	19.4	2.502	2.231	22.100
AGAL006.339−00.746	YSO	0.50	2.77	21.7	3.0	5.4	0.42	18.5	2.312	2.187	22.148
AGAL006.368−00.051	MSF	1.46	4.97	141.3	14.3	6.0	3.49	24.3	4.700	3.627	22.443
AGAL006.393−00.166	Protostellar	0.44	2.84	14.6	3.0	5.4	0.35	16.9	2.189	2.258	22.153
AGAL006.404−00.039	Protostellar	0.53	1.59	141.4	14.3	6.0	1.66	17.0	3.436	3.359	22.230
AGAL006.441−00.524	YSO	0.45	2.45	160.1	9.2	1.3	1.07	16.6	3.285	3.185	22.175
AGAL006.469−00.229	YSO	0.59	6.55	16.7	3.0	5.4	0.87	15.1	2.331	2.700	22.359
AGAL006.479−00.251	YSO	0.60	5.03	16.8	3.0	5.4	0.80	17.6	2.723	2.479	22.260
AGAL006.484−00.547	YSO	0.52	1.91	14.8	3.0	5.4	0.35	19.9	2.573	1.978	22.118

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/.

panels of Fig. 11. The dust temperatures are between 10 and 40 K with a peak at ∼16 K, but is skewed towards warmer temperatures. The mean error of the temperature measurements is 1.9 K with a standard deviation of 1.9 K. Comparison of the temperature distribution between the whole sample and distance-limited subsample

reveals that more distant sources tend towards higher temperatures. The lower panel of Fig. 11 clearly shows that the evolutionary subsamples are well separated in temperature, which increases in line with the expected evolutionary sequence (KS tests comparing all subsamples returns p-values ≪0.001).
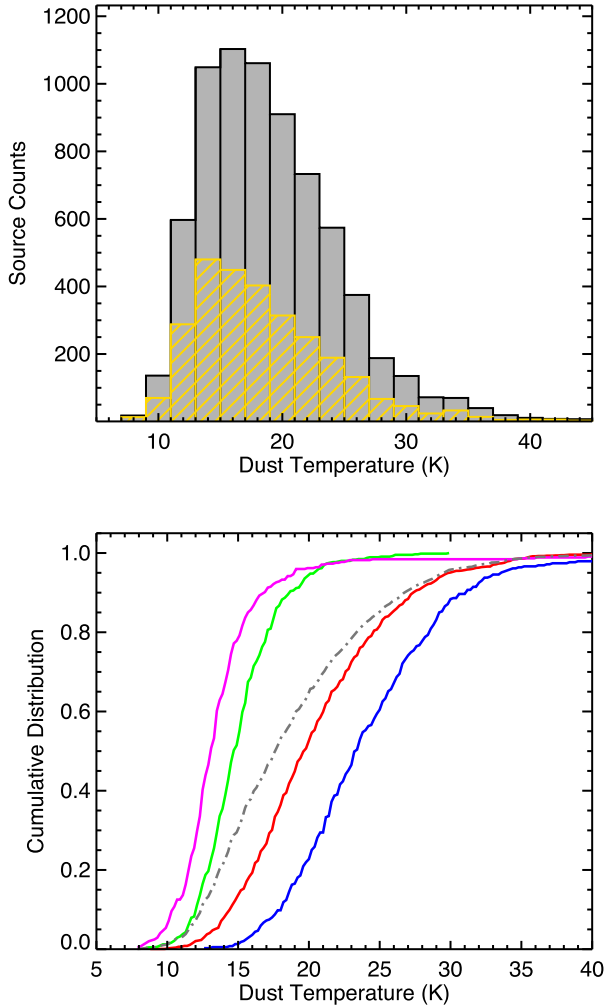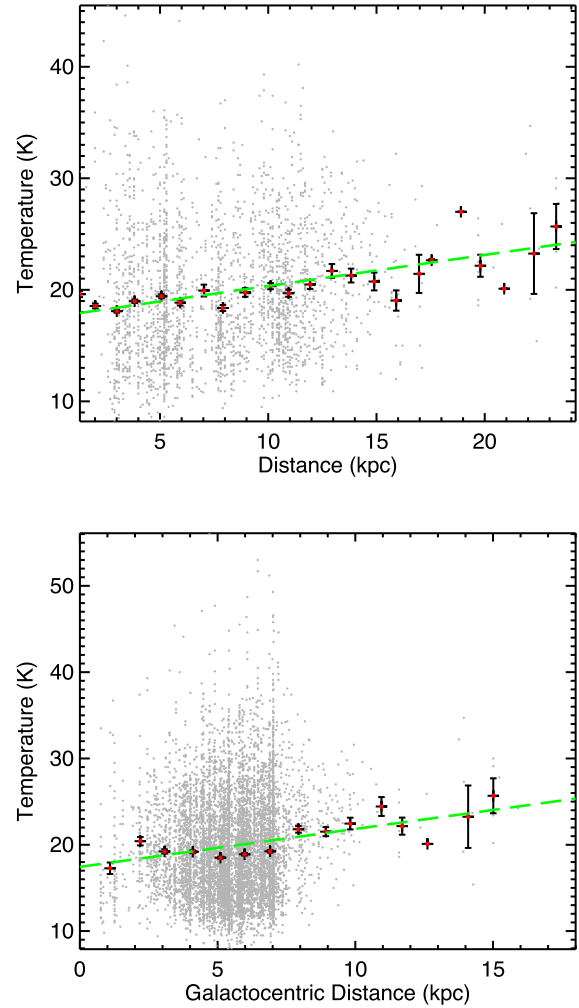
Figure 11. Upper panel shows the temperature distribution for all clumps (grey) and the distance-limited sample (yellow). Lower panel shows the temperature distribution of the distance-limited subsamples of the four evolutionary types described in Section 4 (colours as given in Fig. 10). The bin size used in the upper panel is 2 K.


Figure 12. Upper and lower panels show the temperature distribution as a function of heliocentric and Galactocentric distance, respectively. The red filled circles show the average temperatures and the error bars indicate the standard error of the mean; these are averaged in bins of 1 kpc. The dashed green lines show the result of linear fits to these data.

In the upper panel of Fig. 12, we show the temperature distribution as a function of heliocentric distance. This plot reveals a trend for increasing dust temperature with distance; however, this increase is quite modest, rising by only a few Kelvin between 1 and 15 kpc, which is comparable with the standard deviation in the temperature. A linear fit to the data shows the trend continues beyond 15 kpc although the statistics are quite low and the averaged values show a larger variation. The slope of the fit is 0.28 and the Spearman correlation coefficient is 0.2 with a *p*-value ≪0.001.

We show the temperature distribution as a function of Galactocentric distance in the lower panel of Fig. 12. The majority of sources are located between 3 and 7 kpc and, within this range, the temperature is relatively constant (∼20 K). A linear fit to the data reveals a general trend for an increase in clump temperatures with galactocentric distance that extends beyond the Solar circle. The slope of the fit is 0.44 and the Spearman correlation coefficient is 0.80 with a *p*-value = 0.0004 and so there is a stronger dependence on temperature with their distance from the GC than their heliocentric distance.

We show the luminosity distribution of the whole sample and the distance-limited subsample in the upper panel of Fig. 13. Both distributions have a similar shape, but the distance-limited sample is shifted to slightly lower luminosities. This shift is the result of the Marquis bias, reflecting that more luminous and massive sources tend to be located at greater distances. The lower panel of Fig. 13 shows the luminosity distributions of the four evolutionary subsamples, and clearly exhibits a strong increase in luminosity as a function of evolutionary stage.

5.5 Clump masses, peak column and mean surface densities

Detailed descriptions of the determination of clump mass and column density are presented in Paper I, and so we will only provide a brief summary of the procedure here. The isothermal clump masses are estimated using the Hildebrand (1983) method assuming that the total clump mass is proportional to the integrated submillimetre flux density measured over the source and assuming that the emission is optically thin:

$$M_{\text{clump}} = \frac{D^2 \, S_\nu \, \gamma}{B_\nu(T_{\text{dust}}) \, \kappa_\nu}, \qquad (1)$$
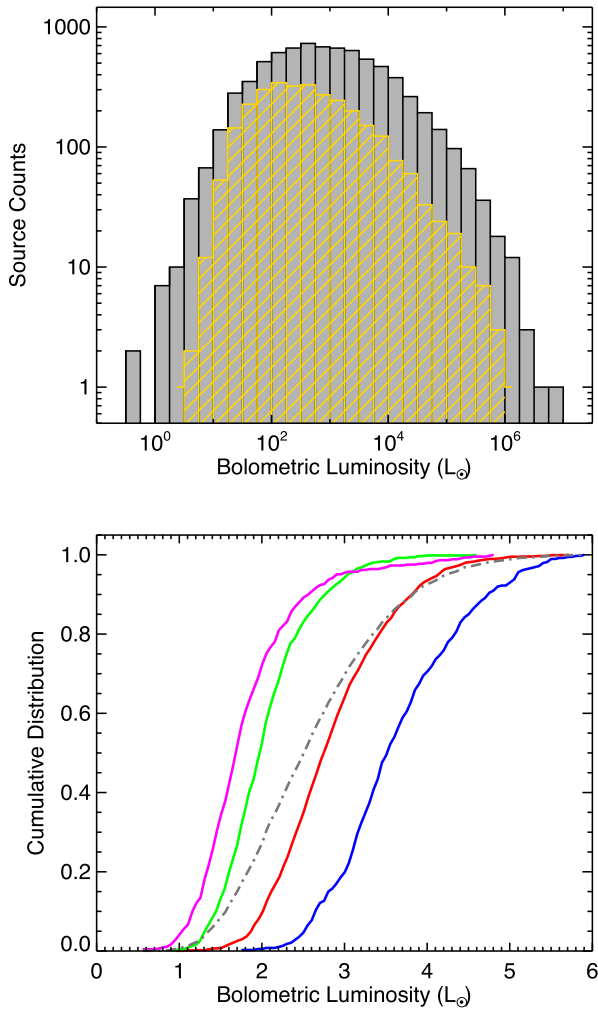
Figure 13. Upper panel shows the bolometric luminosity distribution for all clumps (grey) and the distance-limited sample (yellow). Lower panel shows the luminosity distribution of the distance-limited subsamples (colours as given in Fig. 10). The bin size used in the upper panel is 0.1 dex.

where S_ν is the integrated 870 μm flux density, D is the distance to the source, γ is the gas-to-dust mass ratio (assumed to be 100), 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 in Section 4 and assume that this is a reasonable estimate of the average temperature of the whole clump.

The peak column densities are estimated from the peak flux density of the clumps using:

$$N_{H_2} = \frac{S_{\nu,peak}\,\gamma}{B_\nu(T_{dust})\,\Omega\,\kappa_\nu\,\mu\,m_H}, \qquad (2)$$

where $S_{\nu,\,peak}$ is the peak flux, Ω is the beam solid angle, μ is the mean molecular weight of the gas (which we take to be equal to 2.8), m_H is the mass of the hydrogen atom and κ_ν and γ are as previously defined.

The clump mass and column density distributions are shown in the left- and right-hand panels of Fig. 14, respectively. The clump masses range from as little as a few solar masses to several times $10^4\,M_\odot$, with a peak in the distribution at $\sim400\,M_\odot$, which is likely to indicate the completeness of the survey (we will use this clump mass value where appropriate in subsequent analysis). The

shape of the distribution of clump masses for the distance-limited sample (upper left panel of Fig. 14) is significantly different from the full sample and is lognormal; this is in contrast with many other studies (e.g. Eden et al. 2015; Moore et al. 2015; Csengeri et al. 2017) that have tended to report power-law clump mass profiles that have slopes that are similar to the IMF. The column density distributions for the distance-limited and full samples have similar shapes and peak at approximately the same value. This should be a relatively distance-independent parameter, and this is confirmed by the similarity of their shapes: the higher number of sources in each bin is simply a result of the larger volume probed by the full sample.

We have previously noted trends relating increasing temperatures and luminosities to the evolutionary sequence. We have also seen, however, that the clump radius seems to increase, which is a little harder to interpret. A comparison of distance-limited subsamples illustrates some interesting trends that might provide some insight into the increasing sizes. We note there is a trend for clump masses for the quiescent, protostellar and YSO phases to be higher in earlier stages, however, below a few $100\,M_\odot$ this is likely to be due to incompleteness; above this threshold the cumulative distribution curves begin to converge. This is also the reason for the differences seen for these three phases in the column density distributions. We do not, therefore, read too much into the differences in the column density and clumps masses for the quiescent, protostellar and YSO phases. We do, however, find that the clump mass and column densities of the MSF clumps are both significantly higher than for any of the other subsamples. Among the distance-limited samples, almost every clump above $1000\,M_\odot$ and/or possessing a column density above $\log(N(H_2)) = 23\,cm^{-2}$ is associated with massive star formation: the MSF clumps tend to be significantly more massive, have significantly higher column densities and are significantly larger than their non-MSF counterparts.

There are two possible explanations for the observed properties of the MSF clumps: (1) that they evolve from lower mass clumps associated with the earlier evolutionary stages by acquiring additional mass and growing in size through global infall, or (2) they are simply larger and more massive at their inception and evolve so rapidly that they are not likely to be observed in the early phases.

The first of these options seems less likely given the trend we see for decreasing clump masses and column densities between the quiescent and YSO phases. We can estimate the amount of material that a clump might gain via infall from the surrounding lower density envelope. Assuming a clump with a typical radius of 0.5 pc (see Fig. 10), infall velocity of $1.5\,km\,s^{-1}$ and particle density of $5\times10^3\,cm^{-3}$ (e.g. Wyrowski et al. 2016), the mass infall rate ($\dot{M}$) is $\sim1.5\times10^{-3}\,M_\odot\,yr^{-1}$. As the time-scales for massive star formation is in the region of several $10^5\,yr$ (Mottram et al. 2011; Davies et al. 2011), the total increase in clump mass due to infall (166–500 $M_\odot$ for 1–3× $10^5\,yr$) is likely to be a relatively small fraction of the mass of the MSF clumps but could potentially be more significant for low-mass clumps where the time-scales are likely to be longer.

Furthermore, if infall were having a significant impact, we would expect the mass surface density to increase as the clumps evolve due the influx of material and presumably the contraction of the clumps themselves. In Fig. 15, we show the mass surface densities for the four evolutionary subsamples. Rather surprisingly, this plot reveals that the pre-stellar clumps actually have the highest surface density, and this decreases with advancing evolutionary stage. So rather than the mass surface density increasing due to contraction and infall, we find the opposite trend. The decreasing mass surface
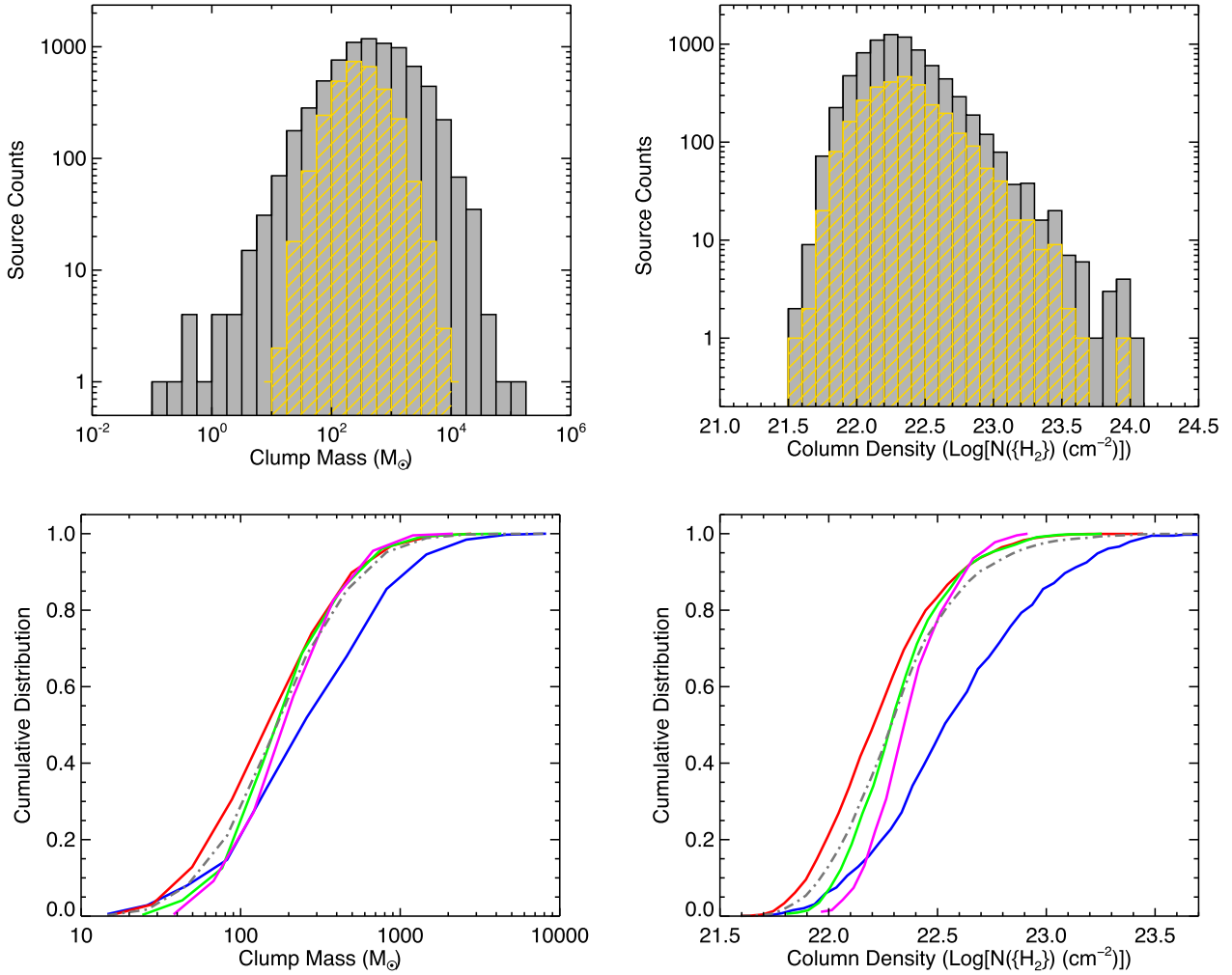
Figure 14. Clump mass and peak column density distribution. In the upper left and right panels, we present histograms of the whole sample (grey) and the distance-limited sample (yellow), respectively. Lower panels show the cumulative distributions of the distance-limited subsamples (colours as given in Fig. 10). The bin sizes used in the upper left and right panels are 0.25 and 0.1 dex, respectively.
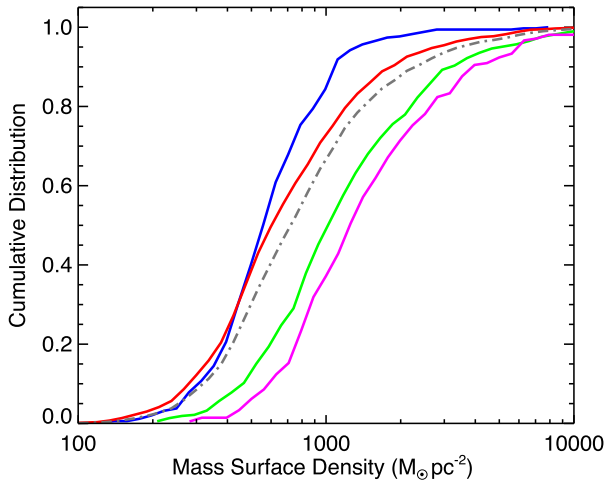
Figure 15. Cumulative distribution of the mean clump mass surface densities (M_{clump}/pc^2) for the four evolutionary distance-limited subsamples (colours as given in Fig. 10).

density as a function of evolution *may* be linked to the ongoing star formation processes within the clumps, which is something that we will investigate in more detail in a later section.

Given the lack of any evidence that might support the hypothesis that the lower mass clumps can evolve to have larger masses, sizes and column densities of clumps associated with massive star formation, the second option therefore seems more plausible and is something we will revisit in later sections.

5.6 Uncertainties in the derived parameters

The typical uncertainty on the distances estimated from Bayesian distance algorithm is ∼0.3 kpc (Reid et al. 2016; this is a 30 per cent error for clumps located at 1 kpc, but only a few per cent for sources located at distances of 10 kpc or more. At the mean distance of 5 kpc, the uncertainty in distance is ∼6 per cent. The mean uncertainty in the integrated flux determined from the SED fitting is $\Delta S_\nu / S_\nu$ ∼42 per cent. The uncertainty in the luminosity is determined by combining the uncertainties in distance and integrated flux in quadrature, but since the error in the latter is significantly larger than in the former, the typical luminosity uncertainty is totally

dominated by the uncertainty in the integrated flux, and effectively has the same level of uncertainty (i.e. $\Delta L/L \sim 42$ per cent).

The uncertainty in the clump mass is estimated from those on the dust temperature ($\Delta T/T \sim 10$ per cent), integrated 870-μm flux ($\Delta S_{\nu,\,peak}/S_{\nu,\,peak} \sim 15$ per cent, Schuller et al. 2009) and distance ($\Delta D/D \sim 6$ per cent), added in quadrature. The typical error in clump mass is thus estimated to be of order 20 per cent. The uncertainty on the column density is considered to be similar to that on the clump masses.

The uncertainties reflected here are the measurement errors; however, all of the calculated parameters are likely to be dominated by inaccuracies in the many assumptions used in determining them (such as the gas-to-dust ratio, the dust opacity and the value of β used in the SED fitting), all of which are poorly constrained. The actual uncertainties are therefore likely to be significantly larger than those determined above. These uncertainties should therefore be considered as lower limits, and the real errors in the measured values could differ by a factor of a few. As noted in the previous papers, however, these additional uncertainties are likely to be systematic and so are unlikely to have a significant impact on the overall distributions or the statistical analysis.

5.7 Comparison with previous ATLASGAL results

The masses and column densities used in the analyses presented in the previous papers in this series were determined using a temperature of 20 K. This was necessary because more reliable temperature estimates were not available. Furthermore, a number of distances have been updated and so it is necessary to determine how these improvements have affected the previous values and evaluate whether these have had a significant impact on our previous results.

The values that are likely to be most affected are the luminosities, clump masses and column densities. Comparing the luminosities, we find them to be in excellent agreement (previous mean value for $L_{bol} = 10^{4.01 \pm 0.03}$, compared to the new value of $L_{bol} = 10^{4.08 \pm 0.03}$). The previous luminosities were determined using the Robitaille et al. (2007) models and Hi-GAL fluxes, and given that the photometry has been determined using two different methods and the SEDs have been determined using two independent models the agreement is very reassuring.

We show the bolometric luminosity distribution of the whole sample and the MSF clumps in the upper panel of Fig. 16. This reveals that the MSF clumps dominate the upper end of the luminosity distribution confirming that our previous efforts of identifying all of the most active sites of massive star formation have been very successful (Papers I, II and III). However, we do also note that there are some very luminous clumps that are currently not in our MSF clump sample. There are few reasons why these were missed: (1) these clumps are associated with evolved H II regions and are resolved in the MSX survey and so not included in the MSX point source catalogue (Price et al. 2001) and consequently not included in the RMS sample (Lumsden et al. 2013); (2) the clumps are located towards the GC in a region excluded from the RMS survey due to problems with source confusion (i.e. $350° < \ell < 355°$ or $5° < \ell < 10°$); (3) they were not located in the region covered in the Methanol Multibeam survey (MMB) catalogue used to identify embedded massive protostars in Paper I (i.e. $186° < \ell < 20°$).

The 20 K temperature that was previously used to estimate the clump masses and column densities is significantly lower than the mean temperature determined from the SED fits to the MSF clumps (24 ± 0.16 K with a standard deviation of 5 K). We might therefore expect the newly estimated clump masses and column densities to
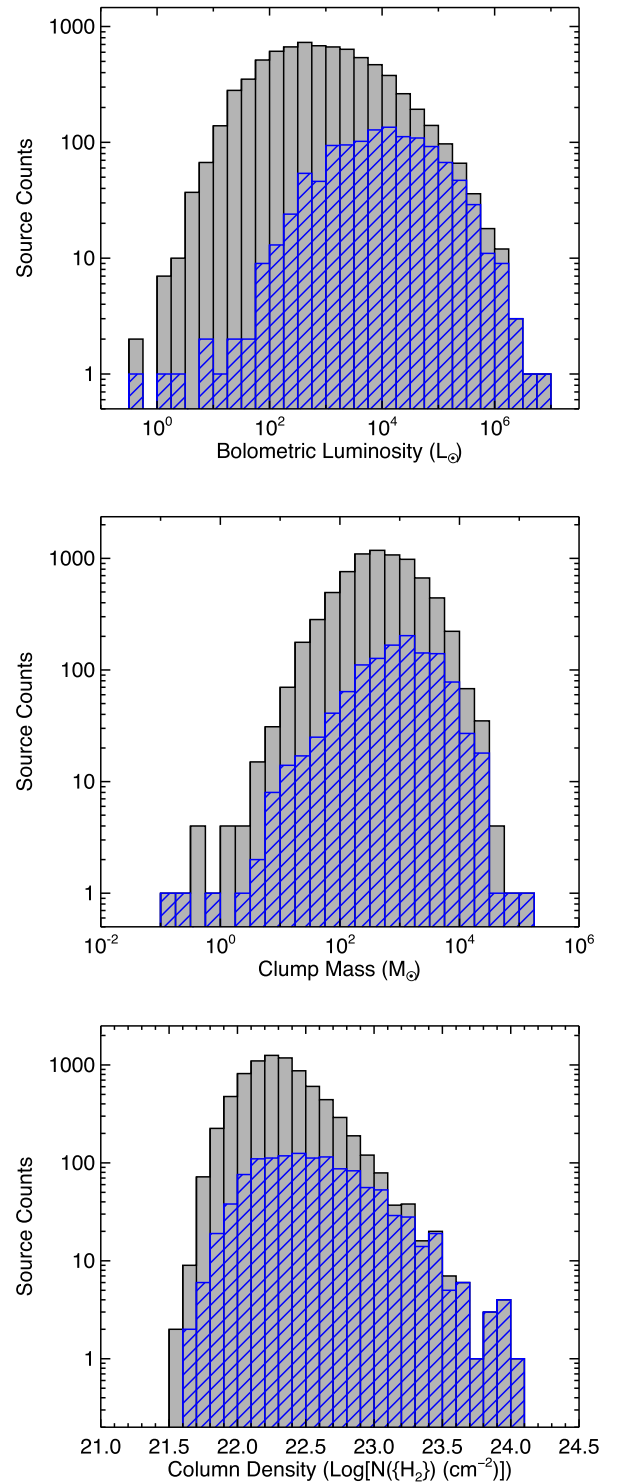
Figure 16. Luminosity, clump mass and peak column density distributions for the whole sample (grey) and the MSF subsample (blue) are presented in the upper, middle and lower panels, respectively. The bin sizes used in the upper, middle and lower panels is 0.25, 0.25 and 0.1 dex, respectively.

be systematically lower than those previously calculated (difference of 4 K corresponds to a decrease of $\sim$30 per cent). The prior average clump mass was $10^{3.29 \pm 0.02}\,M_{\odot}$, a factor of $\sim$2.2 times higher on average than the mean value of $10^{2.95 \pm 0.02}\,M_{\odot}$ calculated in this paper. Column densities differ by the same factor. This corresponds to a difference of $\sim$55 per cent, which is slightly higher than
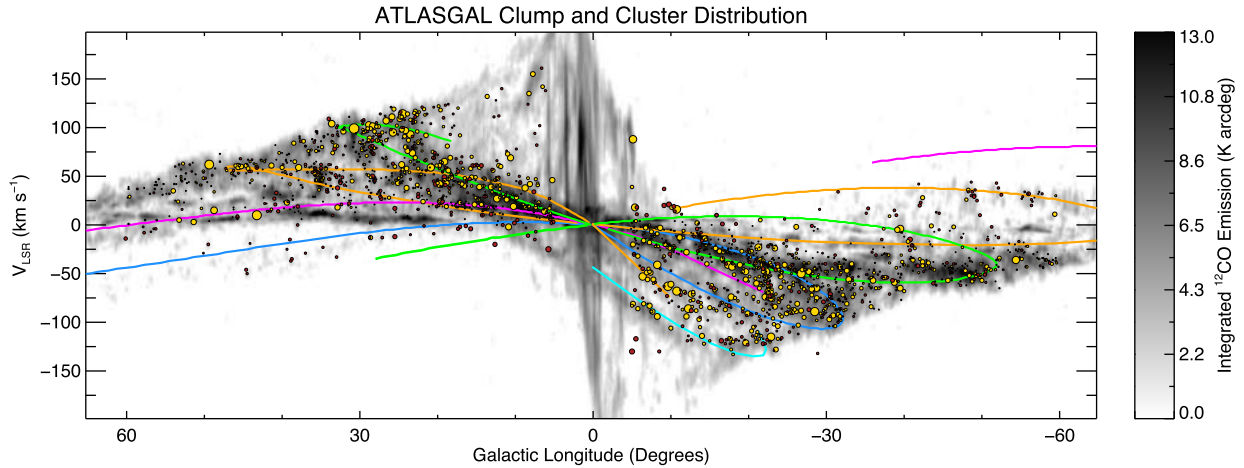
Figure 17. Longitude–velocity (ℓ-v) distribution of clumps and complexes. This figure is the same as Fig. 4 overlaid with the positions of the clusters; these are shown by the larger yellow filled circles. The sizes of the individual clumps and clusters give an indication of their masses.

expected but is likely to result from that the more massive clumps tend to be more active warmer and these skew the mean mass to a lower value.

One of the main conclusions from Urquhart et al. (2014a) was that massive star formation appeared to be associated with the most massive and highest column density clumps and that there were no clumps with column densities above $10^{23.5}$ cm^{-2} that were not already associated with star formation. The question is how has this finding been affected by the change to the column densities. In middle and lower panels of Fig. 16, we show the clump mass and column densities for the whole sample and the MSF clumps, respectively. It is clear from these plots that although the clump masses and column densities have been revised down, they are still significantly more massive and have higher column densities than the general population of clumps.

6 GALACTIC DISTRIBUTION

In Section 2, we compared the position of the ATLASGAL sources with the large-scale molecular gas in the longitude–velocity plane, and found a strong correlation between the locations of the dense gas traced by dust emission and the loci of the main spiral arms. In this section, we revisit this analysis using the mass distribution of clumps and clusters as this is less likely to be biased by the large numbers of nearby and predominately lower mass clumps.

In Fig. 17, we show the longitude–velocity distribution of isolated clumps and clusters. The background grey-scale image is the integrated ^{12}CO (1–0) emission (Dame et al. 2001), which traces the large-scale distribution of molecular gas. The loci of the spiral arms are overlaid (Taylor & Cordes 1993) to facilitate comparison between them and the ATLASGAL sources. It is clear from Fig. 17 that all of the ATLASGAL sources are located within the envelope of molecular gas traced by the ^{12}CO (1–0) emission. Furthermore, we find that the vast majority of the dense gas is correlated with the spiral-arm loci located within 30° of the GC. Detailed analysis of this correlation is difficult in the inner regions (i.e. $|\ell| < 40°$) due to the high density of sources and the overlapping of the spiral arms in ℓv-space. Beyond the inner 40° region, there are some clear correlations between the ATLASGAL clumps and some segments of the spiral arms located in less confused parts of the parameter space (such as the far sides of the Perseus and Norma arms in the

first quadrant and the far side of the Sagittarius arm in the fourth quadrant).

Although the vast majority of ATLASGAL sources are seen to be tightly clustered around the four-arm Milky Way model spiral-arm loci, there are a few small clusters located between the spiral-arm tangents (e.g. between the Scutum-Centaurus and Sagittarius tangents in the first quadrant and between the Sagittarius tangent and Perseus arm, again in the first quadrant). The cluster of sources located between the Scutum-Centaurus and Sagittarius arm tangents is correlated with significant amounts of molecular gas that is traced by CO studies (Lee et al. 2001; Stark & Lee 2006; Rigby et al. 2016), but the presence of the ATLASGAL clumps in these regions is the first indication that these regions are also associated with significant amounts of dense gas. Rigby et al. (2016) presented a very detailed ^{13}CO (3–2) high-resolution map of this region and discussed the origin for the emission (minor arm, an extension of the Scutum-Centaurus, a bridging structure or a spur), but at present its nature is unclear. Despite the presence of significant quantities of molecular material and dense clumps, the RMS survey (Urquhart et al. 2014b) finds no MYSOs or UCH II regions in this region, which suggests that massive star formation is lower than in the major arms.

In Fig. 7, we show the distribution of clusters and clumps with respect to the expected position of larger scale Galactic features, primarily the spiral arms and the Galactic bar. A comparison of the distribution of clumps and clusters with the spiral arms reveals some good agreement, particularly in the first quadrant where the rotation curve is more tightly constrained.

We show the distribution of clumps as a function of Galactocentric distance in the upper panel of Fig. 18. This plot reveals that the vast majority of sources are located between 2 and 8 kpc, with a strong peak at ∼5.5 kpc. The sample of MSF clumps is similarly distributed, with the peak of the distribution slightly farther from the GC. Although the distribution of the whole sample looks simple and suggests that the material in the inner Galaxy is rather smoothly dispersed, the combination of the clumps located in the first and fourth quadrants hides a significant amount of structure. To illustrate this, we have plotted the Galactocentric distributions of the first and fourth quadrant subsamples separately in the middle and lower panels of Fig. 18. We see two prominent peaks in the distribution of the MSF clumps in the first quadrant; these are at 4.5 and 6.5 kpc, which correspond to the Scutum-Centaurus arm and its intersection with the end of the Galactic long bar and the
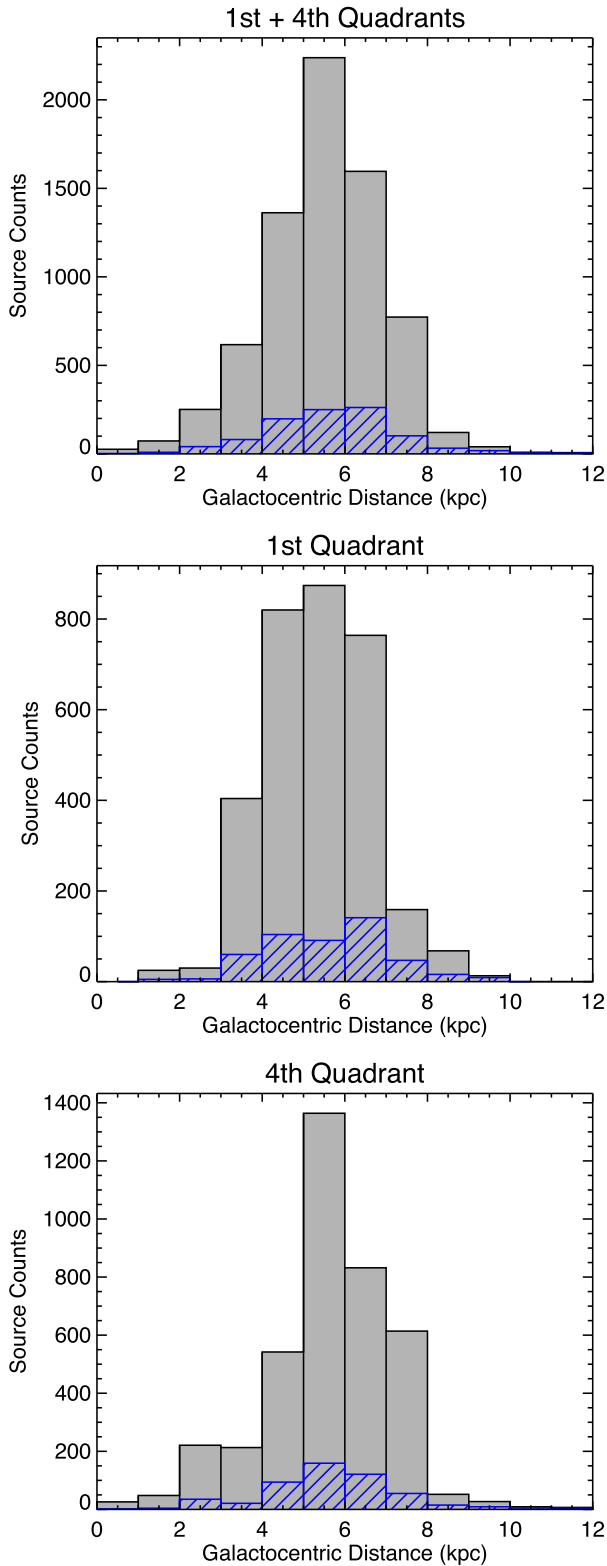
Figure 18. Galactocentric distance distribution of all clumps (grey) and the MSF clumps (blue). The upper panel shows the distribution for the whole sample, while the middle and lower panels show the galactocentric distribution of the first and fourth quadrants, respectively. The bin size used in all of the plots is 1 kpc.
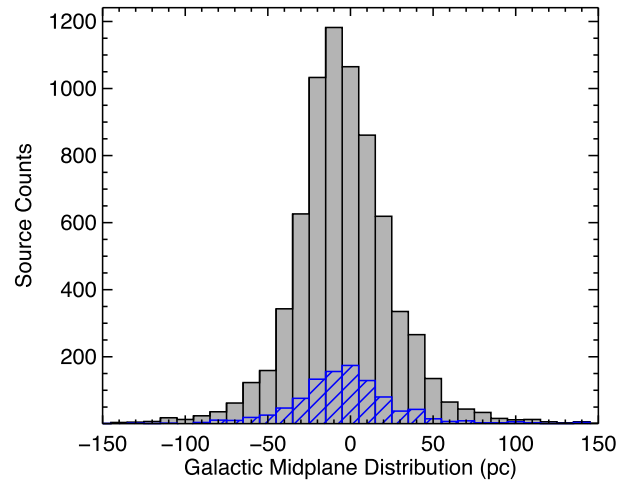
Figure 19. Galactic mid-plane distribution of all clumps (grey) and clumps associated with massive star formation (blue). The bin size used is 10 pc.

Sagittarius arm. Inspecting the fourth quadrant distribution, we note a peak located at 5.5 kpc, which is coincident with a combination of clumps located at the far side of the long bar and the Norma arm, and a second weaker peak at ∼2.5 kpc which is likely to be associated with the near side of the 3 kpc expanding arm.

The presence of peaks in the Galactocentric distribution at locations where spiral-arm segments and the Galactic bar are predicted to be support the spiral model of the Galaxy and that there are significant differences in the structure of the first and fourth quadrants. These structural features in the distribution of MSF clumps have been seen and commented on by many other studies (e.g. Urquhart et al. 2014b; Anderson et al. 2009) including a previous paper in this series (Urquhart et al. 2014c); however, the distances presented here have been estimated using a different rotation model, and approximately 20 per cent of the kinematic distance ambiguities (KDAs) have been revised. The continual presence of these features would suggest they are invariant with respect to the model used.

We show the latitude distribution of all ATLASGAL clumps and those associated with MSF clumps in Fig. 19. A KS test is unable to reject the null hypothesis that the two samples are drawn from the same parent population (p-value = 0.42). A fit of the absolute values of clump distance from the mid-plane finds that the scaleheight, determined from a $1/e$ fit to the data, is 26.04 ± 0.02 pc for the whole sample and 25.15 ± 0.13 pc for the MSF clumps, in excellent agreement with one another and other survey results (e.g. methanol masers, MYSOs and H II regions, Green & McClure-Griffiths 2011; Urquhart et al. 2014b and Paper II, respectively). We can conclude from this that the dense material is concentrated in a relatively narrow region of the Galactic disc, and that the dense gas and massive star formation are tightly correlated with each other.

6.1 Correlation with spiral arms

We have overlaid the spiral-arm loci derived by Taylor & Cordes (1993) updated by Cordes (2004) in Fig. 17, as these are readily available for the whole Galaxy. There is a more recent set of loci derived by Reid et al. (2014) from maser parallax work, but these are currently poorly constrained for the third and fourth quadrants and so have not been used in this figure.

Fig. 20 shows a zoom of Fig. 17 that focuses on the first quadrant and the spiral-arm loci derived by Reid et al. (2014). In addition to
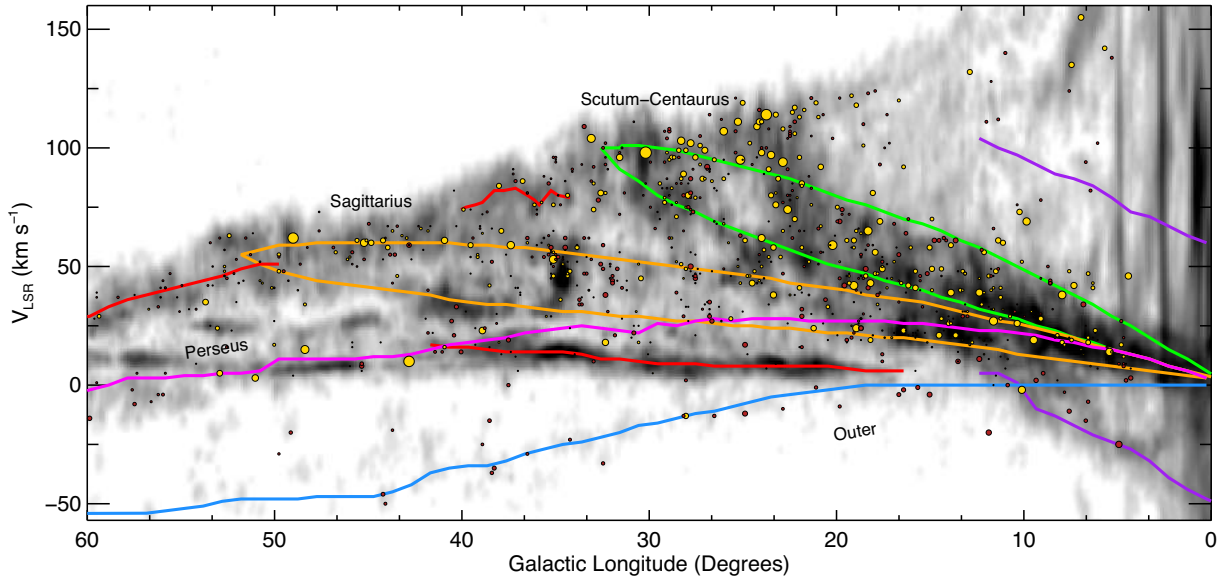
Figure 20. Longitude and velocity (ℓ-v) distribution of clumps and complexes located in the first quadrant. The greyscale image is the integrated ^{12}CO emission from Dame et al. (2001), and the red and yellow circles show the positions of the clumps and clusters, respectively. The sizes of the symbols give an indication of the masses of individual clumps and clusters. This map differs from the one presented in Fig 17 in that the spiral-arm loci are taken from the Reid et al. (2014) model. This map shows the loci of the four main arms (Outer, Scutum-Centaurus, Sagittarius and Perseus, colours are as in Fig. 4), but also includes the near and far 3 kpc arms (lower and upper purple curves, respectively), the local arm (left-most red curve), the Aquila Rift (longest red curve that runs between the Perseus and Outer arms) and the Aquila spur (red curve located between the Sagittarius and Scutum-Centaurus tangents).
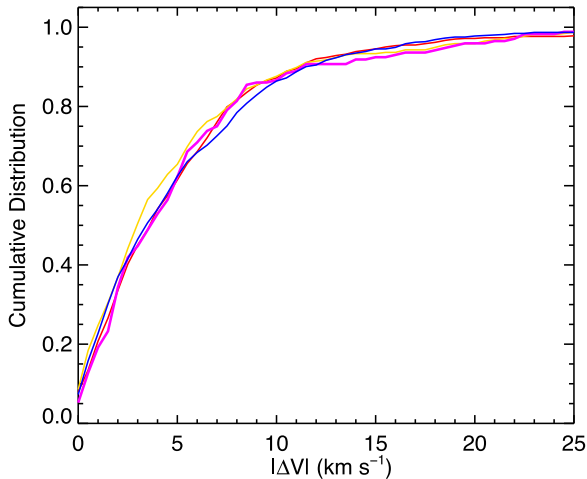
Figure 21. Absolute velocity offset between the ATLASGAL clumps and clusters and their nearest spiral-arm locus (red and yellow curves, respectively). The blue curve shows the results obtained by repeating the analysis for the GRS catalogue of molecular clouds.

the four main arms (which are also shown in Fig. 17), we include the near and far 3 kpc arms, the local arm, the Aquila Rift and the Aquila spur. Compared to Fig. 17, this image reveals a significantly better correlation between the ATLASGAL sources and the spiral-arm loci, particularly for the Outer and Sagittarius arms.

We have calculated the velocity difference between each source and all spirals arms along that source's line of sight to assess the correlation between the spiral-arm loci and the ATLASGAL sources. We then simply assume the source is associated with the spiral arm whose associated velocity difference is lowest. We show the results of this analysis in the form of a cumulative distribution plot in Fig. 21. It is clear from this plot that the velocities of the vast

majority of clumps and clusters are tightly correlated with the spiral arms (almost 90 per cent within 10 km s^{-1}) and that there is no significant difference between the clusters and isolated clumps. We note that the loci of the spiral arms in *lv*-space are not well determined, particularly towards the inner part of the Galaxy, however, the different models do not vary a great deal particularly towards the inner part of the disc (cf. Stark & Lee 2006). There is also an issue of non-circular motions due to streaming motions as clouds pass through the spiral arms, however, these tend to be of order ±7 km s^{-1} (Reid et al. 2014), which is close to the mean velocity difference seen in the distribution, and is probably responsible for the spread we find in the velocity differences.

We have repeated this analysis using the catalogue produced by the GRS (Rathborne et al. 2009) to compare the distribution of dense gas with larger molecular clouds; the results of this analysis are overlaid in blue. A comparison between the dense gas and larger more diffuse molecular cloud locations reveals no significant difference in their correlation with the spiral arms. We have also compared the distributions of the different samples (clusters, clumps and molecular clouds) broken down by mass and luminosity with the spiral-arm loci to determine if the more massive and luminous sources were more tightly correlated with the arms; however, this revealed no significant dependence on either of these two parameters. The distribution of star formation associated with respect to the arms is therefore similar to the distribution of molecular gas, which leads us to conclude that the star formation is not enhanced by the spiral arms and is likely simply the result of source crowding, as suggested by Moore et al. (2012).

7 EMPIRICAL STAR FORMATION RELATIONS

In the previous sections, we have determined the physical sizes, masses, column densities and luminosities for an almost complete

population of dense clumps. This is likely to also be complete to the whole embedded evolutionary sequence for massive stars, and allows for robust conclusions to be drawn from a detailed statistical analysis of these data. We have already found that all of the clumps appear to be rather spherical in structure and are centrally condensed (Paper III). We have found no significant differences in the structural properties of the MSF clumps compared with the rest of the clumps, which is perhaps not terribly surprising as ∼90 per cent appear to be in the process of forming stars. The clumps typically have masses of 500–1000 $M_\odot$ and radii of ∼0.5–1 pc, which is similar to the masses and physical sizes expected to form stellar clusters (Lada & Lada 2003); it is therefore likely that the majority of these are in the process of forming a stellar cluster.

It follows that the derived clump properties (e.g. mass, density, radius, etc.) and the bolometric luminosities of the embedded sources are therefore much more likely to be related to an embedded cluster than to a single star. The luminosity of a given MSF clump (which tend to be physically larger, more massive and have higher peak column densities than the other evolutionary clump samples) is likely to be dominated by an embedded massive protostellar object traced by the methanol maser, MYSO or H II region.

7.1 Analysis tools

In the following subsections, we will look at the correlations between the derived properties in an effort to gain some insight into the statistical nature of star formation in the Galaxy. We use a non-parametric partial Spearman correlation test to determine the level of correlation between pairs of parameters, as this removes their mutual dependence on the distance (Marquis bias, e.g. Yates, Miller & Peacock 1986), of the form $r_{AB,C}$, where

$$r_{AB,C} = \frac{r_{AB} - r_{AC}r_{BC}}{[(1 - r_{AC}^2)(1 - r_{BC}^2)]^{1/2}}, \tag{3}$$

where A and B are the two dependent variables and C is the independent variable (in our case the distance), and r_{AB}, r_{AC} and r_{BC} are the Spearman rank correlation coefficients for each pair of parameters. The significance of the partial rank correlation coefficients is estimated using $r_{AB,C}[(N - 3)/(1 - r_{AB,C}^2)]^{1/2}$ assuming it is distributed as Student's t-statistic (see Collins & Mann 1998 for more details). We check the significance of all correlations by their p-value, which is the probability that the absolute value of the correlation for uncorrelated data could be equal or higher than the measured value (similar to the KS test p-value). We consider a correlation to be significant if the p-value is lower than 0.13 per cent (i.e. difference $>3\sigma$).

If a significant correlation is found, we fit the data using a linear least-squares fit method. This is typically fit in log–log space excluding non-detections and assuming the errors are equal along both axes. These fits capture general trends in the data. As noted in the previous papers in this series, we caution that the mere existence of any correlations between the observed properties does not imply causation. The analysis of the positions of sources in the relevant parameter space can, however, highlight potentially interesting trends that might be present in the data, as well as draw attention to extreme objects, which in turn, may provide some insight into the underlying physics.

7.2 Relationship between luminosity, mass and temperature

Previous studies have found a positive correlation between the luminosity of the embedded massive protostars and the gas tempera-

tures of their natal clumps (e.g. Urquhart et al. 2011). We are able to extend this analysis to lower luminosity and less-evolved sources using the ATLASGAL sample. In the upper left panel of Fig. 22, we show the relationship between temperature and luminosity. This plot shows a strong correlation between these two parameters with an $r_{AB,C}$ of 0.80 (p-value $\ll 0.001$). The fit to these data yields a slope of 0.145±0.0014 for the distance-limited sample, which is equivalent to $L = (-6.0 \pm 0.08) \times T^{(6.9 \pm 0.06)}$. Despite the strong correlation, there is a significant amount of scatter that makes it difficult to assign a particular temperature from the luminosity or vice versa.

In the upper right panel of Fig. 22, we show the relationship between clump mass and temperature. We note that there is no evidence for a correlation between clump mass and temperature as illustrated in the upper right panel of Fig. 22. Given that the clump mass is calculated using the dust temperature, we might have expected these parameters should be strongly correlated; however, the flux is also used to determine the mass, and the flux and temperature are tightly correlated with each other. As the temperature rises so too does the emitted flux proportionally and since these are present as the denominator and numerator, respectively, the mass of a clump is effectively invariant to temperature.

We show the relation between the temperature and the L_{bol}/M_{clump} ratio in the lower left panel of Fig. 22. It is clear from this plot that these two parameters are very strongly correlated with each other ($r_{AB,C}$ of 0.94 and p-value $\ll 0.01$). They are sufficiently well correlated that it is possible to use one to reliably predict the other, and this relation holds over almost 6 orders of magnitude in L_{bol}/M_{clump} and the whole range of measured clump temperatures. The L_{bol}/M_{clump} ratio is a distance-independent parameter, and so the scatter and correlation coefficient are significantly better than that seen in the temperature–luminosity plot. The linear fit to the logs of these data gives a slope of 0.15235±0.00059 and an intercept of 1.231±0.00052; this results in the power-law relation $L_{bol}/M_{clump} = 10^{(-8.08 \pm 0.04)} \times T^{(6.564 \pm 0.035)}$.

We have found, then, that both luminosity and L_{bol}/M_{clump} ratio are strongly correlated with the dust temperature, although the large scatter in the data and strong power-law relationship of the luminosity–temperature distribution makes it a somewhat poor indicator of the stellar evolution taking place within the clumps, i.e. a particular temperature can correspond to a large range of luminosities (∼3 orders of magnitude). The L_{bol}/M_{clump}–temperature relation shows a stronger correlation and has lower power-law dependence; it is therefore less sensitive to small changes in temperature. This relation is independent of distance, removing one source of uncertainty. Furthermore, if the total luminosity is a measure of the energy output of an embedded cluster, then L_{bol}/M_{clump} is a measure of the energy output per unit mass (we are effectively normalizing the energy output by the clump mass to obtain an approximate measure of the instantaneous SFE, which is independent of distance and mass). The results of this L_{bol}/M_{clump} ratio versus temperature analysis is in excellent agreement with the results of an independent analysis recently reported by the Hi-GAL team (Elia et al. 2017).

The lower right panel of Fig. 22 shows the scatter plot of column density and the L_{bol}/M_{clump} ratio. There is no obvious correlation between these two parameters, which suggests that the column density does not evolve significantly during the star formation process. The test indicates that there is a weak negative correlation ($r_{AB,C} = -0.16$ with p-value $\ll 0.01$), suggesting that the column density decreases as the clumps evolve. This result may be an artefact of a significant observational bias: we are less sensitive to the lower column density
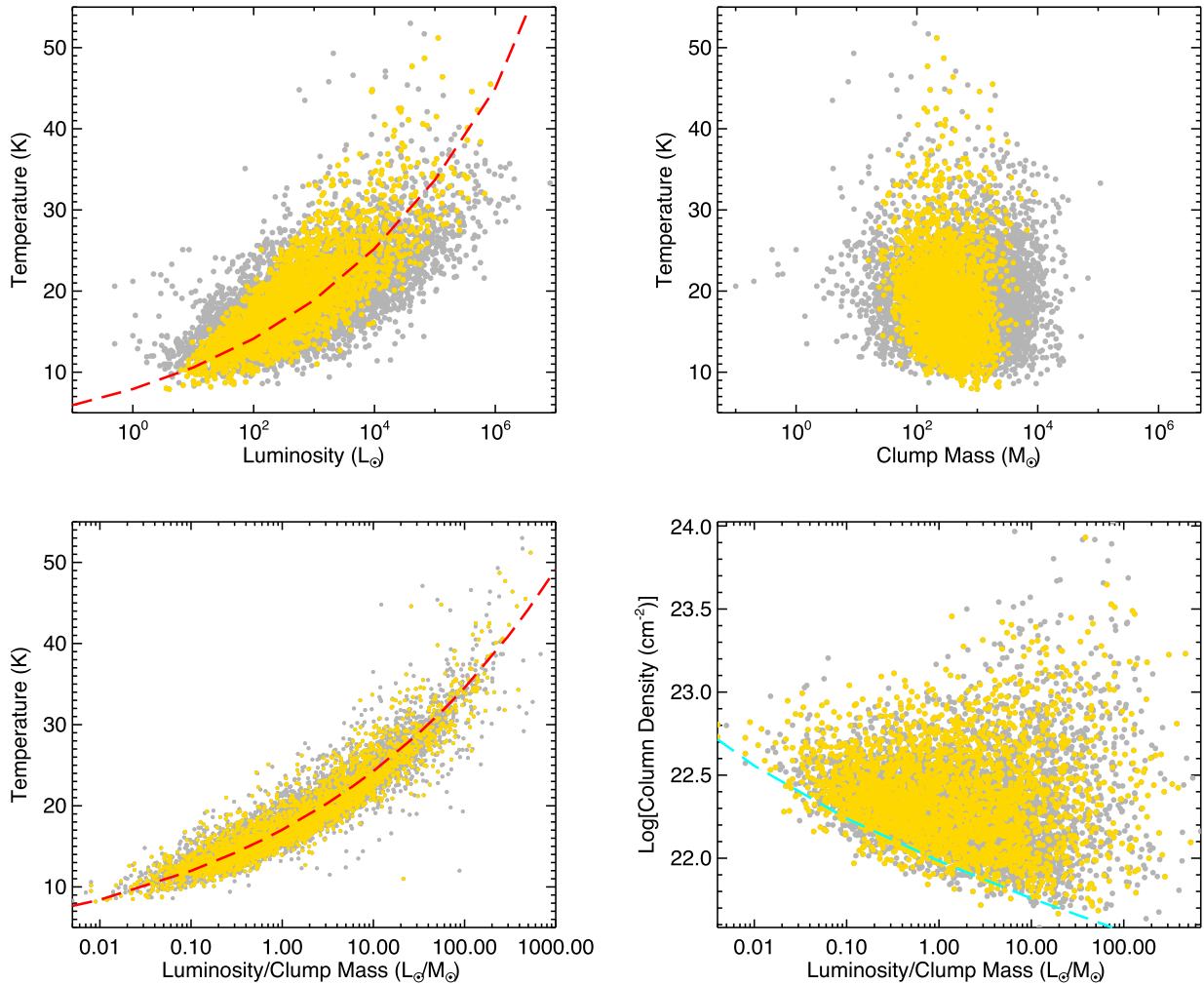
Figure 22. Scatter plots of temperature versus luminosity and temperature versus clump mass (upper left and right panels, respectively). The correlation between temperature and L_{bol}/M_{clump} ratio is shown in the lower left panel, while the lower right panel shows the column density and L_{bol}/M_{clump} ratio relationship. The whole sample is shown as grey circles, while the yellow circles show the distribution of the distance-limited sample. The red dashed lines on the left-hand panels are power-law fits to the data. No significant correlation is found in the distributions in the right-hand panels and so no fits have been made to these data. The dashed cyan line shown in the lower right panel indicates the lower limit to the column density imposed by the observational sensitivity limit.

pre-stellar clumps. The dashed cyan line illustrates the minimum column density detectable for the range of temperatures of interest: this limit is estimated using the ATLASGAL survey's 5σ sensitivity limit ($\sim$300 mJy beam^{-1}), and appears to be a hard boundary at the lower luminosity/column density region of the plot. This limit is likely to be skewing the distribution, and so we do not consider the negative correlation to be reliable.

We have already noted the lack of correlation between the clump mass and temperature. It follows that if temperature and L_{bol}/M_{clump} ratio are good measures of evolution that the clump mass is therefore independent of evolution. We can then conclude at this point that neither the clump mass nor the peak column density change significantly as the star formation evolves. This implies that the majority of clumps have not yet reached the envelope–dispersion stage of their evolution, and that our sample is primarily associated with the earliest evolutionary and pre-stellar stages. The utility of the L_{bol}/M_{clump} ratio as a diagnostic of evolution is illustrated in Fig. 23, which shows that the four evolution stages have clearly distinct profiles.
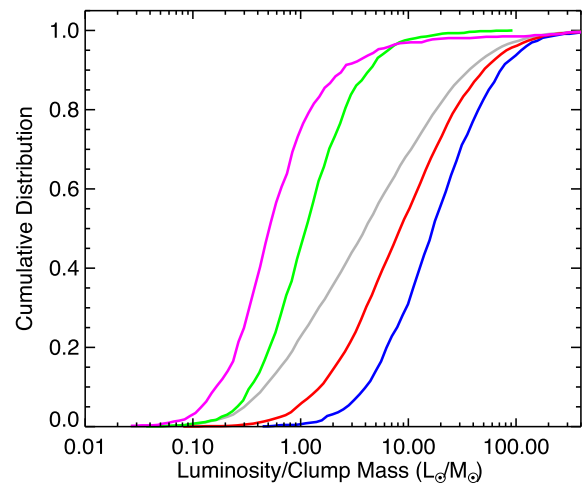
Figure 23. Cumulative distribution of the L_{bol}/M_{clump} ratio for the four evolutionary subsamples (colours as given in Fig. 10).
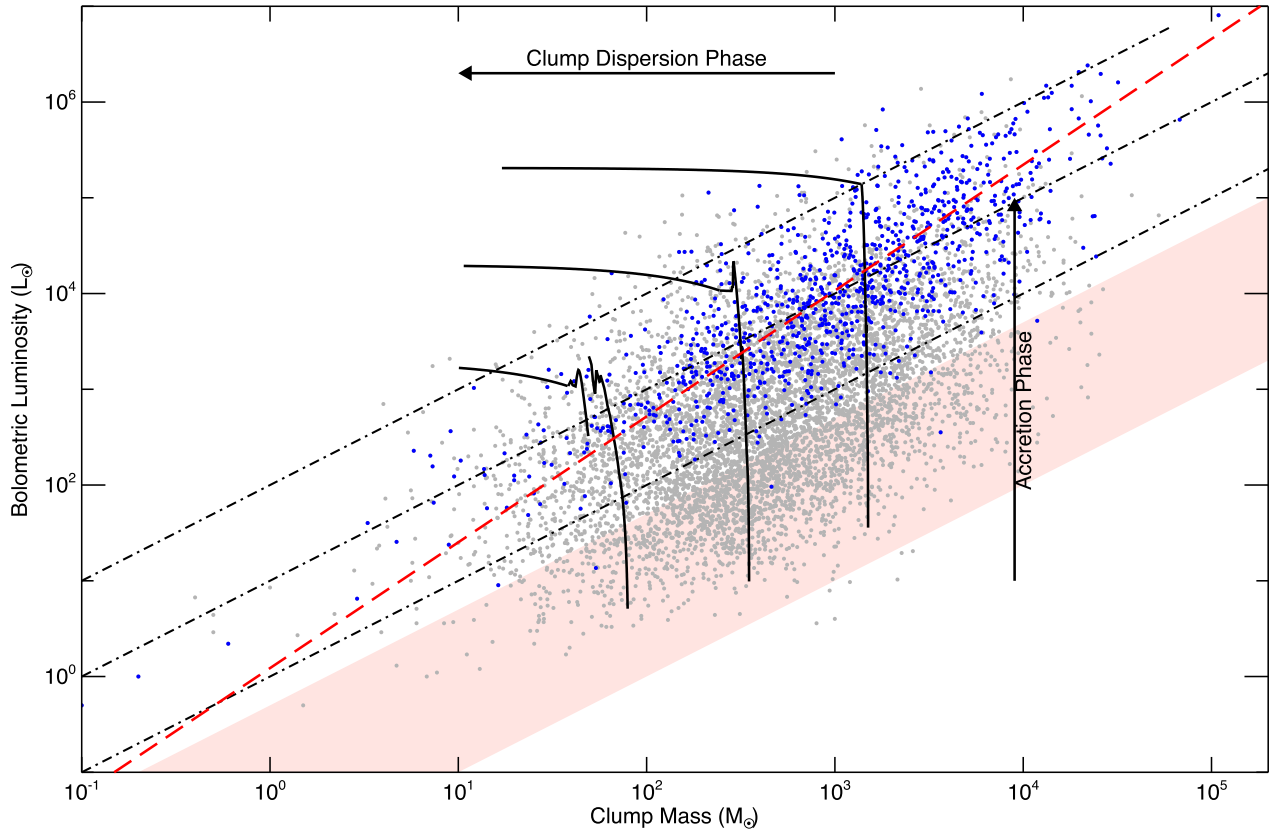
Figure 24. Clump mass–bolometric luminosity relationship for non-MSF clumps (grey circles) and the MSF clumps (blue). The lower, middle and upper diagonal dash–dotted lines indicate the $L_{bol}/M_{clump} = 1$, 10 and 100 $L_\odot/M_\odot$, respectively. The solid black curves, running respectively left to right, show the model evolutionary tracks calculated by Molinari et al. (2008) for stars with final masses of 6.5, 13.5 and 35 $M_\odot$, respectively. The light-red shaded area indicates the region of the parameter space we would expect to find quiescent clumps (temperatures between 10 and 15 K). The long-dashed diagonal red line show the results of a log–log, outlier resistant, linear fit to the MSF clumps; this has a slope of 1.314 ± 0.0189 and an intercept of 0.087 ± 0.056.

7.3 Luminosity–mass relation

We present the luminosity–mass (L_{bol}–M_{clump}) distribution of all ATLASGAL sources in Fig. 24 in order to examine the relationship between the physical properties of the clumps and their associated star formation. This type of diagram has been used in studies of low-mass (Saraceno et al. 1996) and high-mass star-forming regimes (Molinari et al. 2008; Giannetti et al. 2013), and is a useful diagnostic tool for separating different evolutionary stages. This figure shows the MSF and non-MSF clumps as blue and grey circles, respectively. For illustrative purposes, we include evolutionary tracks derived by Molinari et al. (2008) from a simple model for the formation of massive stars. This model is based on the observational evidence that their formation is a scaled-up version of the inside–out collapse model (which has been successfully applied to low-mass star formation). These show how the luminosity of the forming massive star and the mass of its natal clump changes as the star evolve over time. These tracks consist of vertical and horizontal components that Molinari et al. (2008) refer to as the *main accretion* and the *envelope clean-up* phases; these are indicated by arrows on the figure. The light-red shaded area indicates the region of the parameter space in which we would expect to find quiescent clumps, with temperatures of ~10–15 K.[3] The sample therefore covers all evolutionary stages from the quiescent starless/pre-stellar phase through to the formation of an H II region, when feedback from the massive star halts further accretion and begins to disperse the natal cloud. We also include three lines of constant L_{bol}/M_{clump} (i.e. 1, 10 and 100 $L_\odot/M_\odot$) in Fig. 24.

The MSF clumps are clearly the most luminous for any given clump mass: this is particularly true for the higher mass clumps. It is also clear that the most massive stars are forming in the most massive and dense clumps. As noted in Paper III, nearly all of the MSF clumps are found within a relatively broad range of L_{bol}/M_{clump} but the upper and lower envelopes of this range are well described by values of 100 and 1 $L_\odot/M_\odot$, respectively. It is particularly interesting to note the upper bound to the L_{bol}/M_{clump} range, as this may indicate that there is a mechanism that works to limit the conversion of mass into stars: this could be linked to the feedback from the embedded protoclusters. In Paper III, we compared the luminosity for the MSF clumps to what would be expected assuming an SFE of 10 per cent, and found that the most massive clumps were underluminous, while the lower mass clumps were more luminous than expected: this suggests that the SFE naturally decreases with the mass of the cluster.

---

[3] The lower end of the temperature range is determined by heating from the interstellar radiation field and cosmic ray heating (~10 K), and the upper

temperature from the mean temperature determined here from the SED fits to the quiescent clumps and from the mean ammonia temperature measured for IRDCs (e.g. Wienen et al. 2012).

The MSF clumps form a continuous distribution that is concentrated towards an $L_{bol}/M_{clump} \sim 10$ L$_\odot$/M$_\odot$ locus that extends over approximately 4 orders of magnitude in both axes. This clustering around 10 L$_\odot$/M$_\odot$ is likely to trace the transition between the main accretion phase and the clump dispersion phase. If the accretion increases with the mass of the protostar (as favoured by the population synthesis model, Davies et al. 2011), the protostar will accelerate towards the end of its vertical track, reaching the end of its main accretion phase with increasing rapidity. Once the main accretion phase has ended, it takes a long time before the embedded star begins to significantly disrupt its natal clump, retarding its progression along its horizontal track.

Similar thresholds in L_{bol}/M_{clump} are discussed by Giannetti et al. (2017) in the context of an evolutionary sequence for the process of high-mass star formation, confirming the above results. The authors compare different molecular tracers of physical conditions, and find that after an initial compression phase of the material in the clump ($L_{bol}/M_{clump} \lesssim 2$ L$_\odot$/M$_\odot$), the YSOs accrete material and grow in mass, reaching the zero age main sequence (ZAMS) at $L_{bol}/M_{clump} \sim 10$ L$_\odot$/M$_\odot$, and start dispersing the parent clump. Compact H II regions become common at $L_{bol}/M_{clump} \gtrsim 40$ L$_\odot$/M$_\odot$, however, we also note that few regions are found with L_{bol}/M_{clump} in excess of 100 L$_\odot$/M$_\odot$.

In the previous papers, we used a temperature of 20 K to estimate the masses for the MSF clumps, and fitting these data resulted in a relatively constant L_{bol}/M_{clump} of 10 L$_\odot$/M$_\odot$ (slope 1.03 ± 0.05) for all clump masses. We have repeated this analysis with the updated masses calculated using the dust temperatures, and find a slightly steeper relationship: ($\log(L) = 1.314 \times \log(M) + 0.0874$). The updated value for the slope is in better agreement with the value reported by Molinari et al. (2008) (~1.27) from a comparable fit to a sample of 27 IR-bright sources, and consistent with the results of a similar analysis presented by Giannetti et al. (2013) and Urquhart et al. (2015, 1.32±0.08). The luminosity and clump masses are strongly correlated, with $r_{AB,C} = 0.69$ and $p \ll 0.001$, verifying that the correlation is statistically significant.

7.3.1 Dependence of L/M on clump mass

We have previously noted that the MSF clumps are generally the most luminous and have higher L_{bol}/M_{clump} ratios: this is illustrated in the lower panel of Figs 13 and 23. We also observe from Fig. 24 that the ratio of the numbers of MSF to non-MSF clumps increases noticeably as the clump mass increases. This indicates that the L_{bol}/M_{clump} distribution is not the same for the whole mass range and may increase with clump mass. In Fig. 25, we compare the L_{bol}/M_{clump} distribution for three mass ranges (i.e. $M < 1000$ M$_\odot$, 1000 M$_\odot < M < 5000$ M$_\odot$ and $M > 5000$ M$_\odot$) for the whole population of clumps. This figure clearly shows that the L_{bol}/M_{clump} ratios are significantly different for the three mass bins, and this is confirmed by KS tests ($p \ll 0.001$). The mean values and error for the $\log(L_{bol}/M_{clump})$ ratios for the increasing mass bins are 0.60±0.012, 0.80±0.022 and 1.10±0.052 L$_\odot$/M$_\odot$, respectively. Closer inspection of the lowest mass range reveals a bimodal distribution with peaks a few tenths of a dex to either side of $L_{bol}/M_{clump} = 1$ L$_\odot$/M$_\odot$ (approximately 0.5 and 3.0 L$_\odot$/M$_\odot$). The upper panel of Fig. 25 includes error bars calculated assuming Poisson counting statistics (i.e. $\sqrt{N}$), and these confirm that the dip is significant. The distributions of the other two mass bins show only a single peak, although the distribution for the largest mass range is significantly flatter.
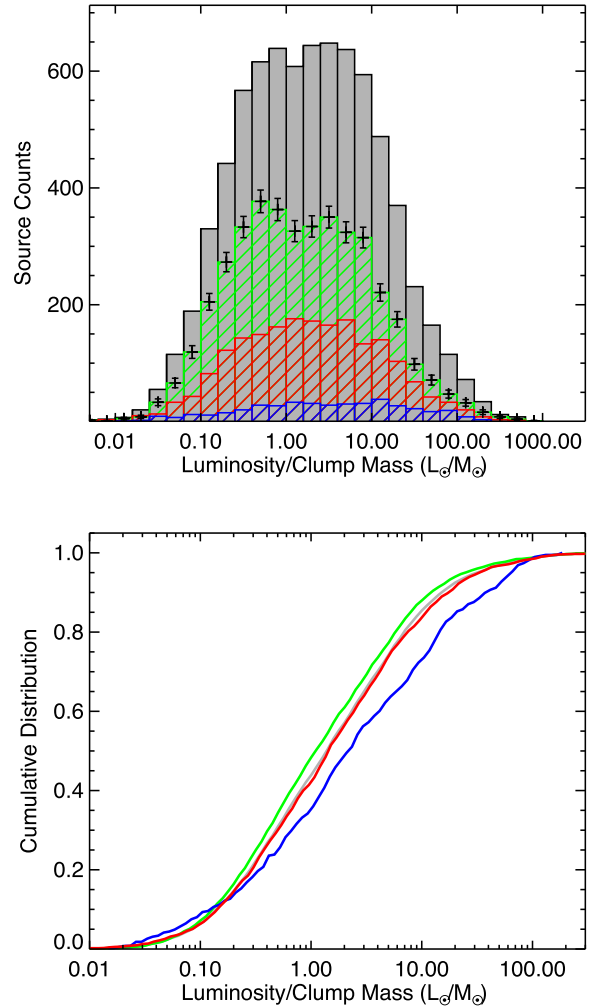
Figure 25. Luminosity to clump mass ratio. In the upper and lower panels, we show the frequency distribution of the L_{bol}/M_{clump} ratio for clumps with masses $M < 1000$ M$_\odot$, 1000 M$_\odot < M < 5000$ M$_\odot$ and $M > 5000$ M$_\odot$ in green, red and blue, respectively. The distribution of the whole clump population is shown in grey.

We show the cumulative distributions for the three mass ranges in the lower panel of Fig. 25. This plot reveals that the L_{bol}/M_{clump} ratio for the most massive clumps (i.e. $M > 5000$ M$_\odot$) is significantly higher than for the other clump mass ranges. This suggests that the massive clumps (>5000 M$_\odot$) have a very short pre-stellar lifetime and evolve more rapidly than lower mass clumps. This shift in the L_{bol}/M_{clump} ratio may also be due to the steeper-than-linear slope in the $L - M$ plot, and is therefore not necessarily evidence of evolution. In Fig. 26, we show the fractions of clumps found to be quiescent (70 μm dark), protostellar (24 μm dark) and YSOs (mid-IR bright). This plot reveals a clear trend for an increasing fraction of YSO-forming clumps with increasing mass, and a decrease in the fraction of quiescent and protostellar clumps. The fraction of clumps harbouring a YSO increases from 60 per cent to nearly 90 per cent over a range of clump masses that extends over almost 2 orders of magnitude, while the fraction of quiescent clumps decreases from ~15 per cent to 5 per cent over the same mass range. Furthermore, the fractions derived for the quiescent clumps are likely to be upper limits, as a significant number may host low-mass protostars that are not yet sufficiently luminous to be detected at 70 μm. This supports the hypothesis that the statistical lifetime of the earliest stages