

# ATLASGAL – properties of a complete sample of Galactic clumps<sup>★</sup>

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

*Affiliations are listed at the end of the paper*

Accepted 2017 August 30. Received 2017 August 29; in original form 2017 July 15

## ABSTRACT

The APEX Telescope Large Area Survey of the Galaxy (ATLASGAL) is an unbiased 870  $\mu\text{m}$  submillimetre survey of the inner Galactic plane ( $|\ell| < 60^\circ$  with  $|b| < 1.5^\circ$ ). 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  $\sim 8000$  dense clumps located in the Galactic disc ( $5^\circ < |\ell| < 60^\circ$ ). We derive highly reliable velocities and distances to  $\sim 97$  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  $\sim 5 \times 10^4$  yr for clump masses  $\sim 1000 M_\odot$  decreasing to  $\sim 1 \times 10^4$  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  $10^{23} \text{ cm}^{-2}$ ; 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

<sup>★</sup>The full version of Tables 2, 5, 8 and A2, and Figs 6, 8 and A1 are only available in electronic form at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.125.5) or via <http://cdsweb.u-strasbg.fr/cgi-bin/qcat?J/MNRAS/>.

\* E-mail: [j.s.urquhart@gmail.com](mailto:j.s.urquhart@gmail.com)

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  $^{13}\text{CO}$  (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  $70\ \mu\text{m}$  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.  $\sim 50$  per cent – Murray & Rahman 2010 and  $\sim 30$  per cent – Urquhart et al. 2014b, respectively). These have very different physical properties and environmental conditions,

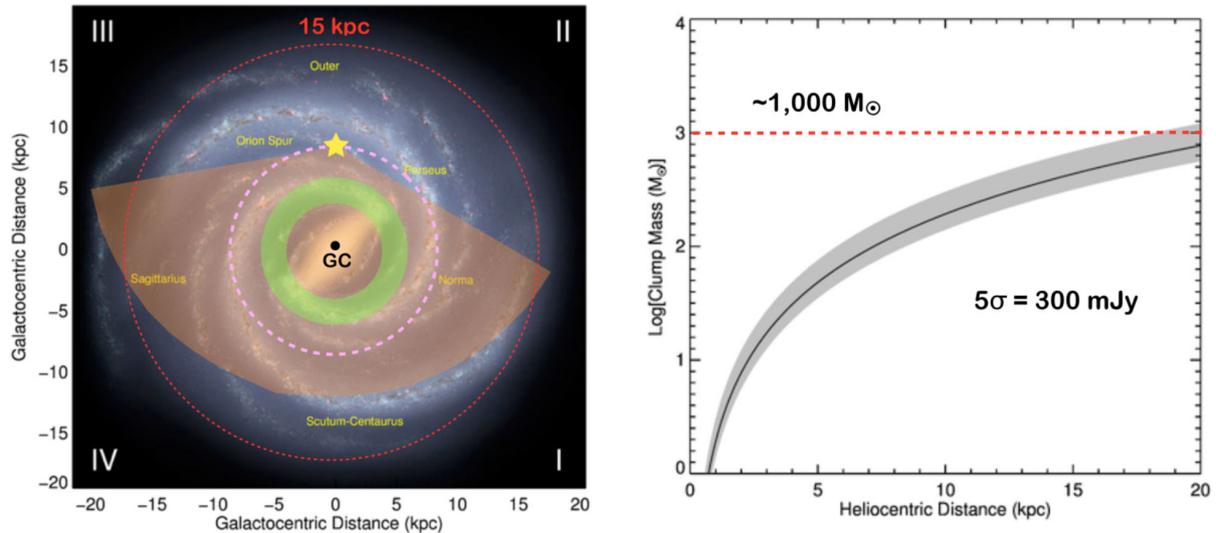
including some of the most extreme in the Galaxy. The remaining star formation is more evenly distributed across the Galactic plane, and while the physical properties and local environments tend to be less extreme, these regions are still responsible for a large fraction of the overall Galactic star 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  $\sim 10\ 000$  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<sup>1</sup> located within the Solar circle and  $\sim 90$  per cent of all dense molecular gas in the Galaxy (Urquhart et al. 2014a). Assuming a dust temperature of 15 K,<sup>2</sup> a distance of 20 kpc and a flux density of  $0.3\ \text{Jy beam}^{-1}$  ( $\sim 5\sigma$ ) corresponds to a minimum mass sensitivity of  $\sim 1000\ M_{\odot}\ \text{beam}^{-1}$  or  $\sim 700\ M_{\odot}\ \text{pc}^{-2}$ . The ATLASGAL CSC should therefore include the vast majority of all current and future high-mass star-forming clumps ( $M_{\text{clump}} > 1000\ M_{\odot}$ ) in the Galaxy (assuming a standard initial mass function (IMF) and a typical SFE of  $\sim 30$  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  $5\sigma$  and that are located away from the GC region ( $300^{\circ} < \ell < 355^{\circ}$  and  $5^{\circ} < \ell < 60^{\circ}$ ). 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

<sup>1</sup> 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 ( $N(\text{H}_2) \gtrsim 10^{22}\ \text{cm}^{-2}$ ).

<sup>2</sup> This temperature is the average kinetic temperature determined from ammonia observations of quiescent clumps (i.e. Wielen 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  $5\sigma$  flux sensitivity of  $300 \text{ mJy beam}^{-1}$ . The black curve shows the mass sensitivity assuming a dust temperature of 20 K, while the grey shaded region shows the uncertainty, allowing  $\pm 5 \text{ K}$  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  $\sim 1300$  massive star-forming (MSF) clumps, however, this sample only make up a relatively small fraction of the total population of clumps identified by ATLASGAL ( $\sim 15$  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,  $v_{\text{LSR}}$ ). 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,  $\text{NH}_3$ , 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 ATLASGAL 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.  $\text{NH}_3$  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	$^{13}\text{CO} (2-1)/\text{C}^{18}\text{O} (2-1)$	This paper
ATLASGAL	693	$\text{NH}_3 (1,1)$	Wienen et al. (2012)
ATLASGAL	299	$\text{NH}_3 (1,1)$	Wienen et al. (2017)
ATLASGAL	154	$\text{N}_2\text{H}^+ (1-0)$	Urquhart et al. (in preparation)
ATLASGAL (70 $\mu\text{m}$ dark)	50	$\text{C}^{18}\text{O} (2-1)/\text{N}_2\text{H}^+ (1-0)$	Pillai et al. (in preparation)
ATLASGAL	62	$\text{N}_2\text{H}^+ (1-0)$	Csengeri et al. (2016a)
MALT90 <sup>a</sup>	1205	$\text{HNC}/\text{N}_2\text{H}^+ (1-0)$	Jackson et al. (2013)
BGPS	959	$\text{N}_2\text{H}^+/\text{HCO}^+ (1-0)$	Shirley et al. (2013)
ThrUMMS <sup>b</sup>	960	$\text{CO} (1-0)$	Barnes et al. (2015)
COHRS	580	$^{12}\text{CO} (3-2)$	Dempsey et al. (2013)
MSGPS	499	$\text{CO} (1-0)$	Burton et al. (2013)
RMS	407	$^{13}\text{CO}/\text{NH}_3$	Urquhart et al. (2007, 2008, 2011, 2014b)
GRS <sup>c</sup>	293	$^{13}\text{CO} (1-0)$	Jackson et al. (2006)
BGPS	153	$\text{NH}_3 (1,1)$	Dunham et al. (2011b)
UC H II cand.	148	$\text{CS} (2-1)$	Bronfman, Nyman & May (1996)
CHIMPS	83	$^{13}\text{CO} (3-2)$	Rigby et al. (2016)
HOPS	60	$\text{NH}_3 (1,1)$	Purcell et al. (2012)
IRDCs	55	$\text{CS} (2-1)$	Jackson et al. (2008)
SEDIGISM	43	$^{13}\text{CO} (2-1)$	Schuller et al. (2017)
M8	3	$\text{CO} (3-2)$	Tohill et al. (2002)
IR-quiet clumps	2	$\text{NH}_3 (1,1)$	Hill et al. (2010)

<sup>a</sup>Spectra were extracted towards the submillimetre peak of each ATLASGAL source from the 3 arcmin  $\times$  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://atofa.atnf.csiro.au/>).

<sup>b</sup>Source: <http://www.astro.ufl.edu/~peterb/research/thrumms/>

<sup>c</sup>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  $10 \text{ km s}^{-1}$  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  $\sim 1000$  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  $\mu\text{m}$  dark sources presented here was obtained from a dedicated follow-up of 70  $\mu\text{m}$  dark sources with the APEX 12 m ( $\text{C}^{18}\text{O} 2-1$ ) and IRAM 30 m ( $\text{N}_2\text{H}^+ 1-0$ ) telescopes (the APEX and IRAM 30 m telescopes have similar beam sizes at these frequencies; FWHM  $\sim 28$  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 ( $\sim 98$  per cent of the sample). In Table 2, we give the source names, velocities, molecular transition used to determine the velocity and

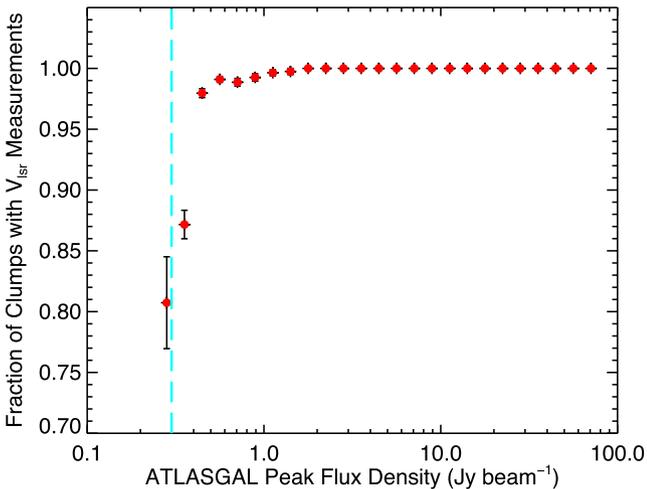
**Table 2.** Summary of the kinematic distance analysis.

ATLASGAL CSC name	$v_{\text{lsr}}$ ( $\text{km s}^{-1}$ )	$v_{\text{lsr}}$ references	Reid distances			Adopted kinematic solution			Cluster	
			Bayesian (kpc)	Near (kpc)	Far (kpc)	Distance (kpc)	Solution Flag <sup>a</sup>	Distance reference <sup>b</sup>	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/>.

<sup>a</sup>The distance solution flags refer to the different steps described in Appendix B1 and Table B2.

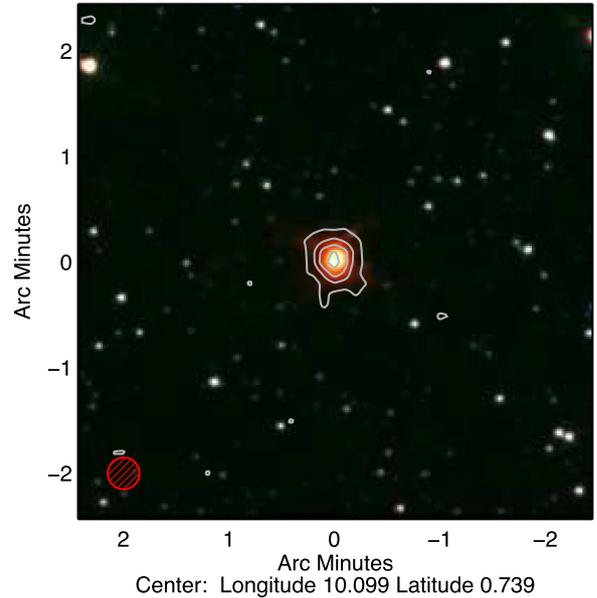
<sup>b</sup>References for distance solutions: (1) Anderson & Bania (2009), (2) Araya et al. (2002), (3) Battisti & Heyer (2014), (4) Urquhart et al. (2013b), (5) Dunham et al. (2011b), (6) Fish et al. (2003), (7) Reid et al. (2014), (8) Sanna et al. (2014), (9) Wu et al. (2014), (10) Xu et al. (2009), (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  $5\sigma$  sensitivity limit of the ATLASGAL survey.

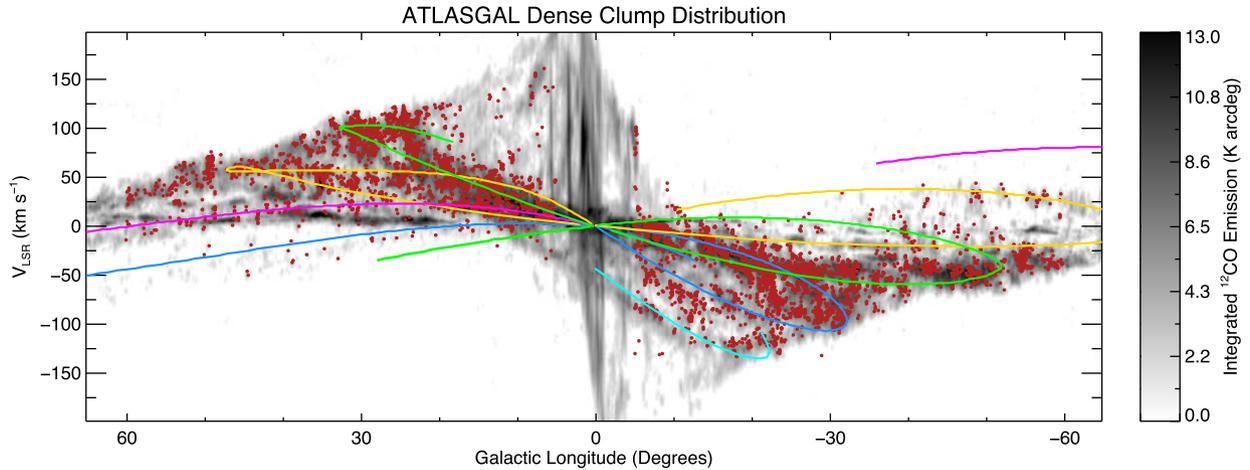
the reference from where these data originated. In Fig. 2, we show the proportion of sources with determined radial velocities as a function of the peak flux distribution. This plot shows that we are complete for nearly all sources with a peak flux density  $\approx 0.5 \text{ Jy beam}^{-1}$ . Visual inspection of IR and submillimetre maps for the  $\sim 200$  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  $^{12}\text{CO}$  and  $^{13}\text{CO}$  (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.

AGAL010.099+00.739 IRAC +  $870 \mu\text{m}$  Contours



**Figure 3.** Three colour image of an evolved star candidate (this is constructed using the GLIMPSE 3.6, 5.8 and  $8 \mu\text{m}$  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  $^{13}\text{CO}$  (2–1) emission detected.

While the vast majority of sources identified in the ATLASGAL 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 [ $^{12}\text{CO}$  (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  $^{12}\text{CO}$  (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  $\rightarrow$  blue; Perseus  $\rightarrow$  magenta; Sagittarius  $\rightarrow$  yellow; Scutum-Centaurus  $\rightarrow$  green and near 3-kpc arm  $\rightarrow$  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 ( $\ell v$ ) 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  $^{12}\text{CO}$  (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  $\text{H I}$  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  $\sim 7000$  ATLASGAL clumps, with  $\sim 80$  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  $\sim 40$  per cent

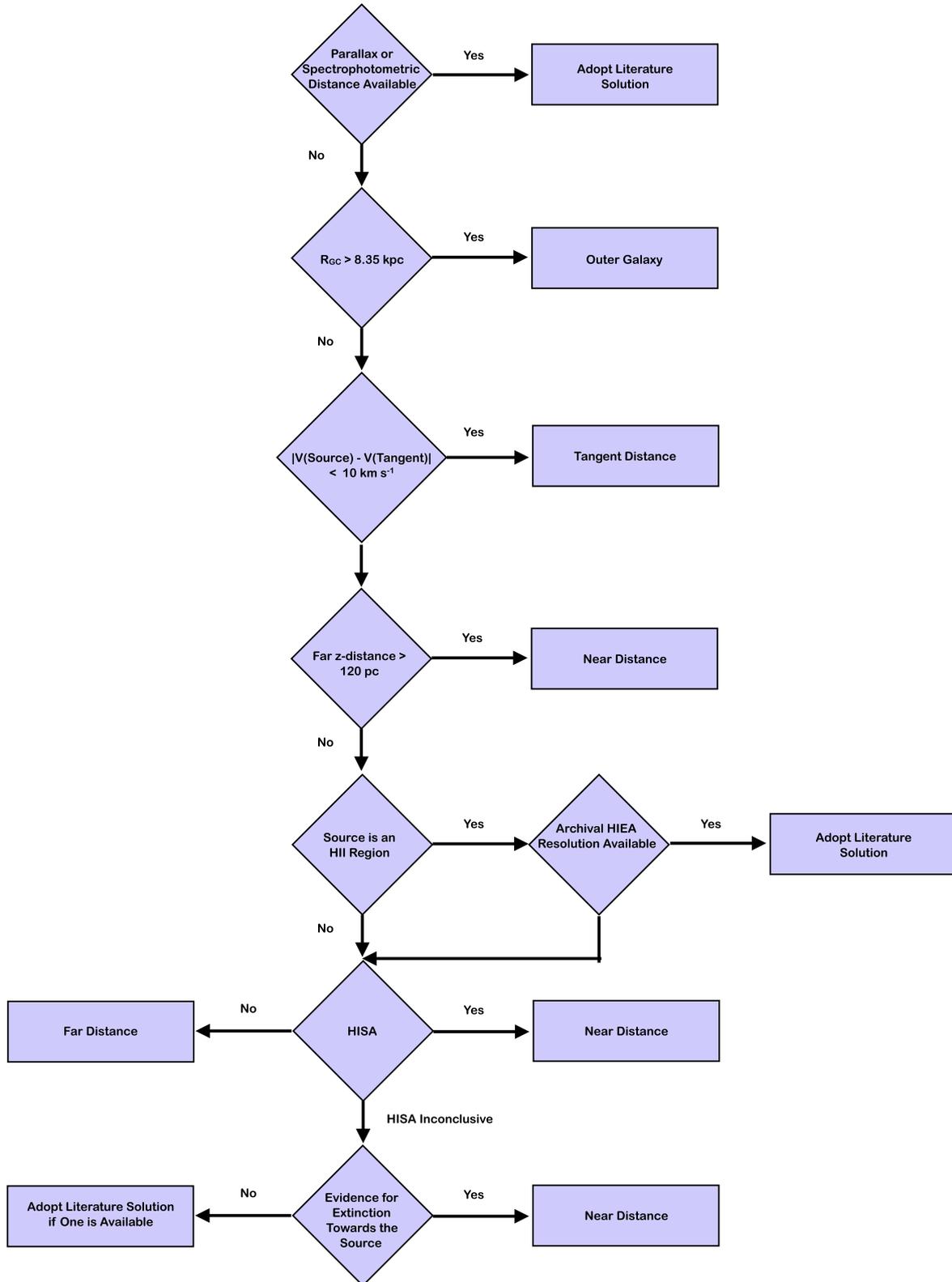
of their distances were the result of distance ambiguity resolution: the remainder are arbitrarily placed at the far distance and so the samples are not comparable as their results are likely to be biased to systematically larger distances.

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  $\text{H I}$  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  $8\ \mu\text{m}$  *Spitzer* images taken as part of the GLIMPSE legacy survey. The  $8\ \mu\text{m}$  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  $\text{H II}$  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  $\ell v$  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  $\text{H I}$  self-absorption ( $\text{H I SA}$ ) technique to solve the distance ambiguity or extracted more reliable distances from the

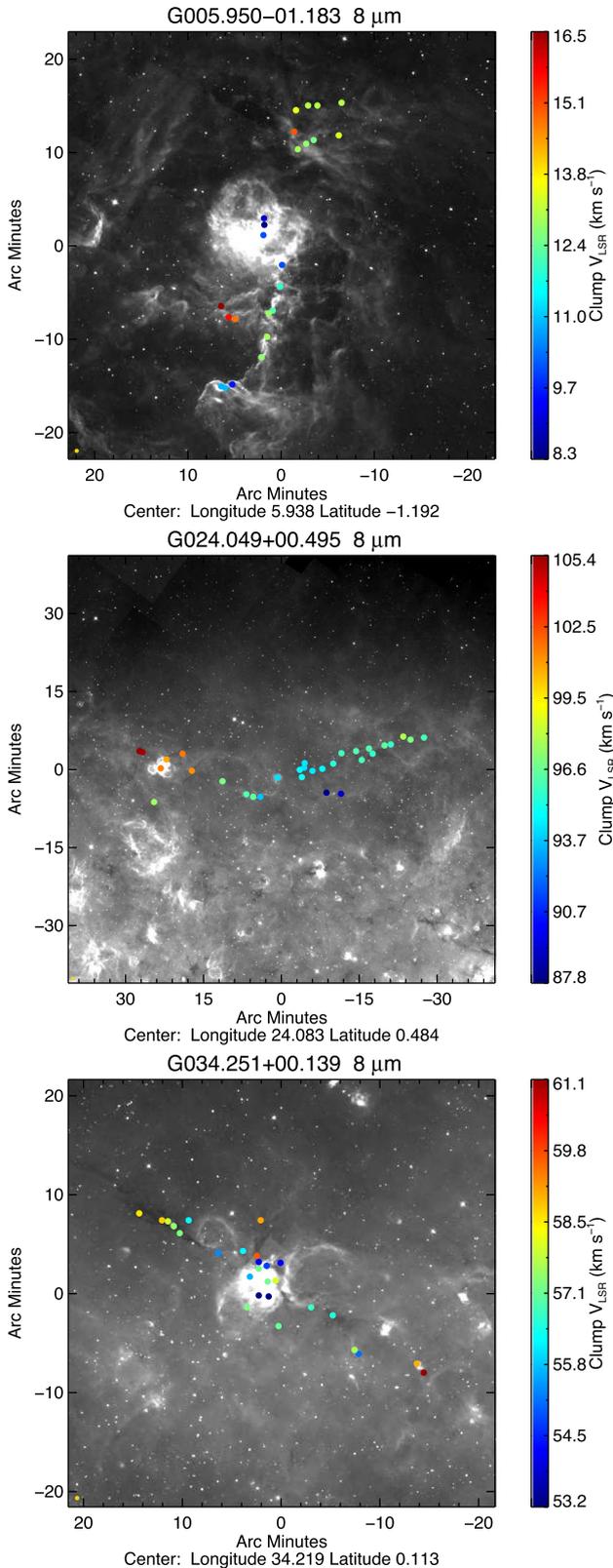


**Figure 5.** Flow chart showing the criteria used to determine distance to ATLASGAL CSC clumps.

literature for an additional 1150 sources. In total, this combination of methods has provided distances to 7770 clumps ( $\sim 97$  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<sup>-1</sup> 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.

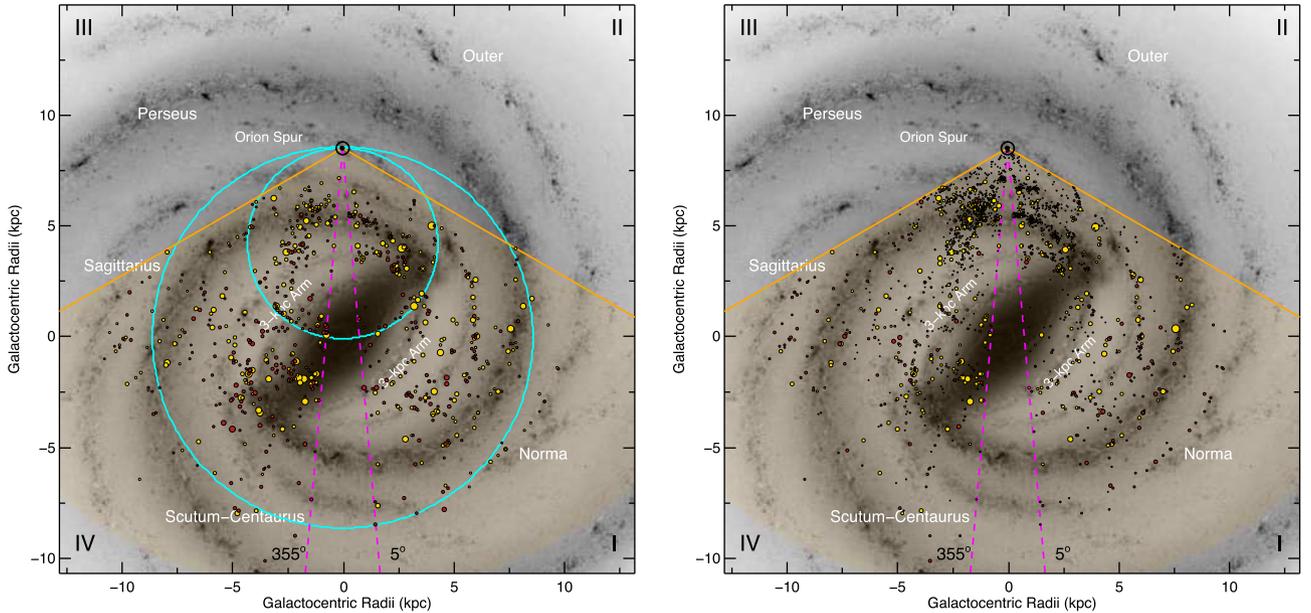
Taken together, these four subsamples represent an evolutionary sequence for the formation of massive stars and clusters (König et al. 2017) and therefore comparing the physical properties of these different phases will provide some insight into the processes associated with star formation.

## 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  $\odot$  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  $3\sigma$  (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  $\sim 10$  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.

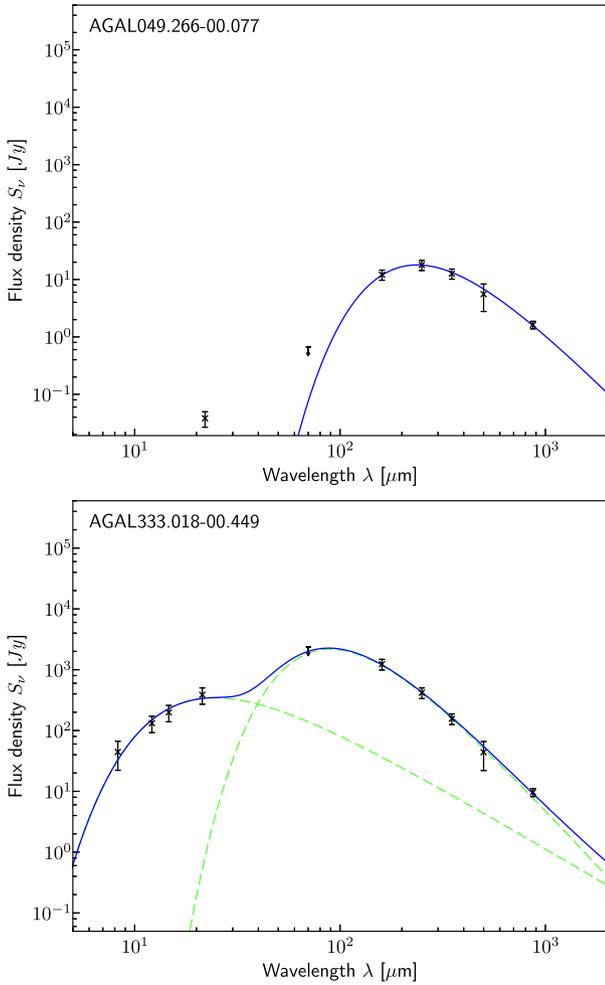
We note that the highest density of sources occurs between 2 and 4 kpc for both the full sample of clumps and for the MSF

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  $\sim 60$  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  $0.72 \pm 0.01$  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 envelopes 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.

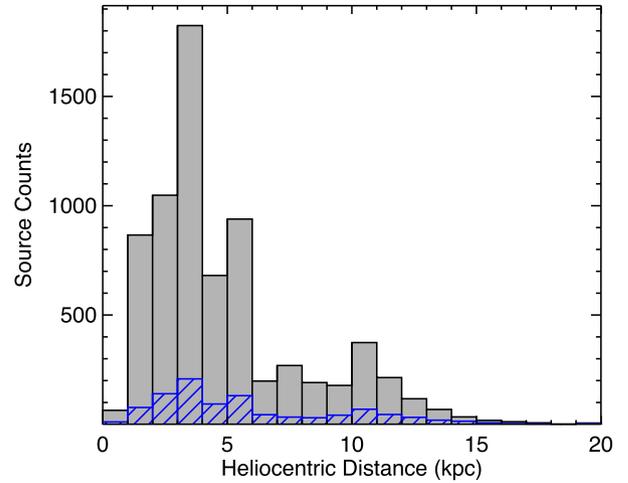
**Table 3.** Summary of evolutionary types identified from the SED analysis and our previous work (Paper III).

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

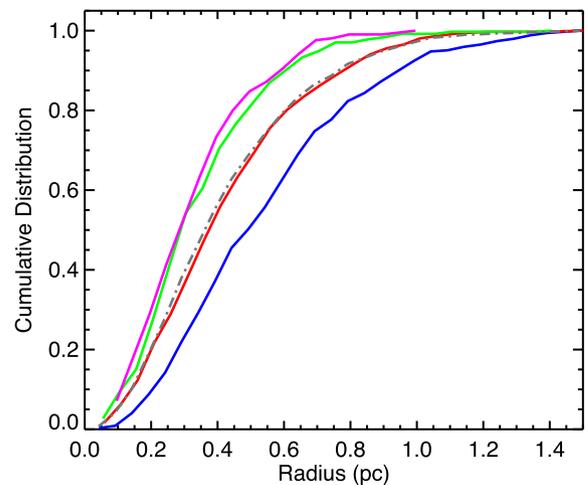
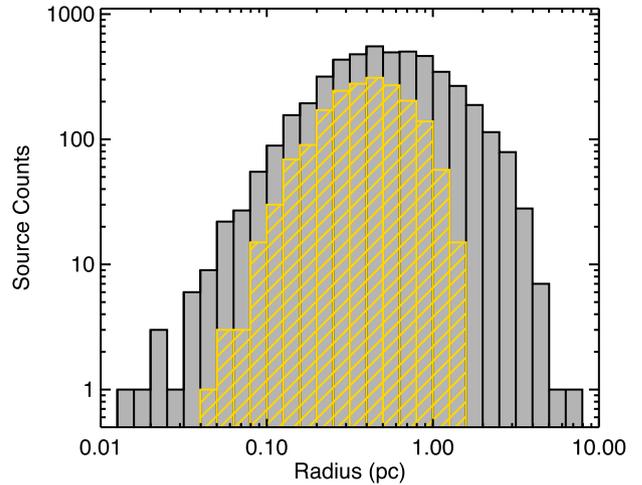
warmer and more evolved objects are more readily detectable (this is discussed in more detail in Section 8).

#### 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	$\bar{x}$	$\frac{\sigma}{\sqrt{N}}$	$\sigma$	$x_{\text{med}}$	$x_{\text{min}}$	$x_{\text{max}}$
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[ $N(\text{H}_2)$ ( $\text{cm}^{-2}$ )]	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
$L_{\text{bol}}/M_{\text{clump}}$ 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 ( $M_{\odot} \text{pc}^{-2}$ )]	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

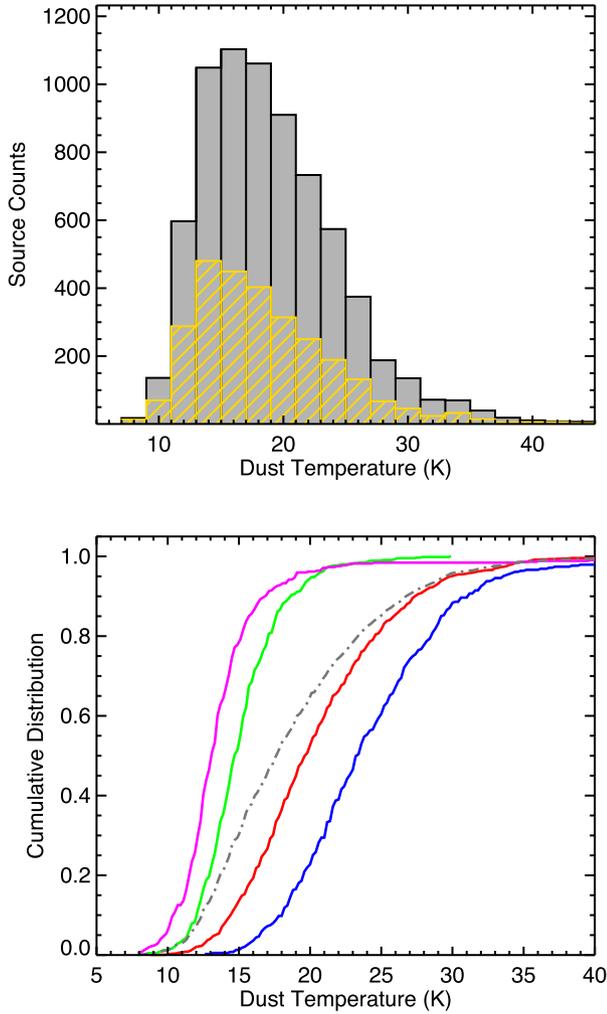
**Table 5.** Derived clump parameters.

CSC name	Evolution type	Peak flux ( $\text{Jy beam}^{-1}$ )	Integrated flux (Jy)	$V_{\text{LSR}}$ ( $\text{km s}^{-1}$ )	Distance (kpc)	$R_{\text{GC}}$ (kpc)	Radius (pc)	$T_{\text{dust}}$ (K)	log[ $L_{\text{bol}}$ ] ( $L_{\odot}$ )	log[ $M_{\text{clump}}$ ] ( $M_{\odot}$ )	log[ $N(\text{H}_2)$ ] ( $\text{cm}^{-2}$ )
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  $\sim 16$  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  $\ll 0.001$ ).

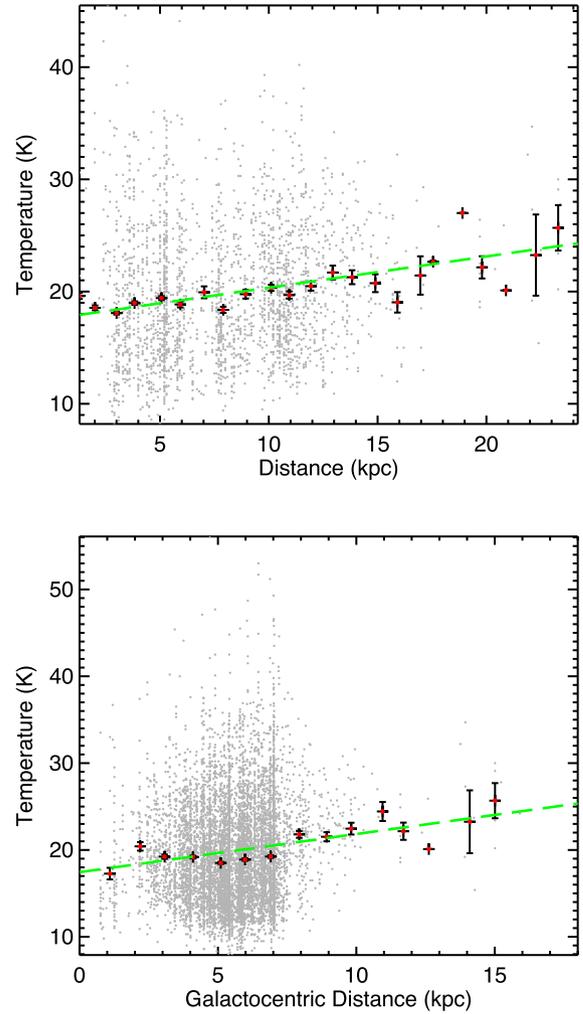


**Figure 11.** Upper panel shows the temperature distribution for all clumps (grey) and the distance-limited sample (yellow). Lower panel shows the temperature distribution of the distance-limited subsamples of the four evolutionary types described in Section 4 (colours as given in Fig. 10). The bin size used in the upper panel is 2 K.

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  $\ll 0.001$ .

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 ( $\sim 20$  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



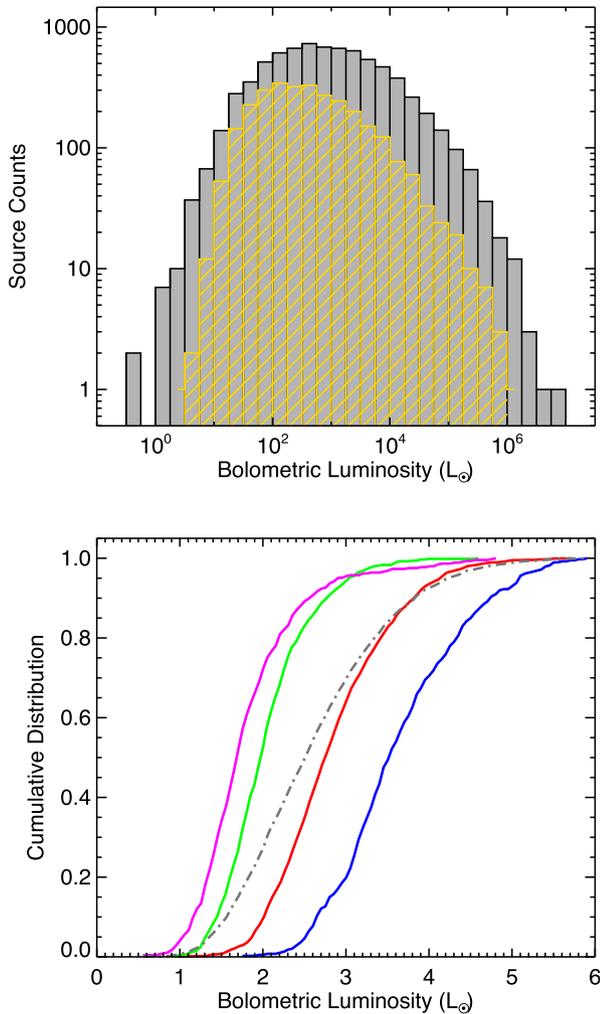
**Figure 12.** Upper and lower panels show the temperature distribution as a function of heliocentric and Galactocentric distance, respectively. The red filled circles show the average temperatures and the error bars indicate the standard error of the mean; these are averaged in bins of 1 kpc. The dashed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

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:

$$M_{\text{clump}} = \frac{D^2 S_{\nu} \gamma}{B_{\nu}(T_{\text{dust}}) \kappa_{\nu}}, \quad (1)$$



**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  $S_\nu$  is the integrated  $870\ \mu\text{m}$  flux density,  $D$  is the distance to the source,  $\gamma$  is the gas-to-dust mass ratio (assumed to be 100),  $B_\nu$  is the *Planck* function for a dust temperature  $T_{\text{dust}}$  and  $\kappa_\nu$  is the dust absorption coefficient taken as  $1.85\ \text{cm}^2\ \text{g}^{-1}$  (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:

$$N_{\text{H}_2} = \frac{S_{\nu, \text{peak}} \gamma}{B_\nu(T_{\text{dust}}) \Omega \kappa_\nu \mu m_{\text{H}}}, \quad (2)$$

where  $S_{\nu, \text{peak}}$  is the peak flux,  $\Omega$  is the beam solid angle,  $\mu$  is the mean molecular weight of the gas (which we take to be equal to 2.8),  $m_{\text{H}}$  is the mass of the hydrogen atom and  $\kappa_\nu$  and  $\gamma$  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  $\sim 400 M_\odot$ , which is likely to indicate the completeness of the survey (we will use this clump mass value where appropriate in subsequent analysis). The

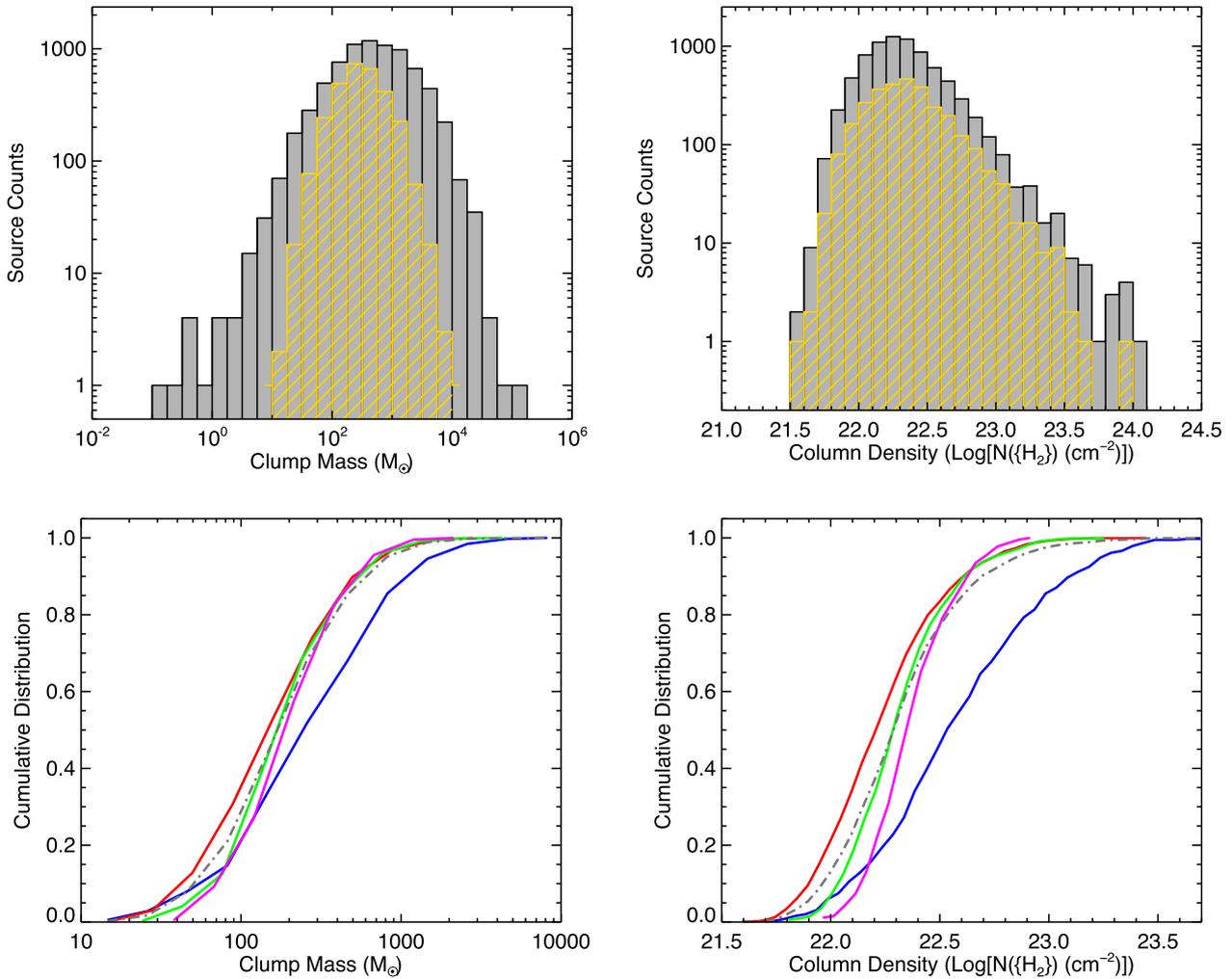
shape of the distribution of clump masses for the distance-limited sample (upper left panel of Fig. 14) is significantly different from the full sample and is lognormal; this is in contrast with many other studies (e.g. Eden et al. 2015; Moore et al. 2015; Csengeri et al. 2017) that have tended to report 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 clump 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  $\log(N(\text{H}_2)) = 23\ \text{cm}^{-2}$  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.

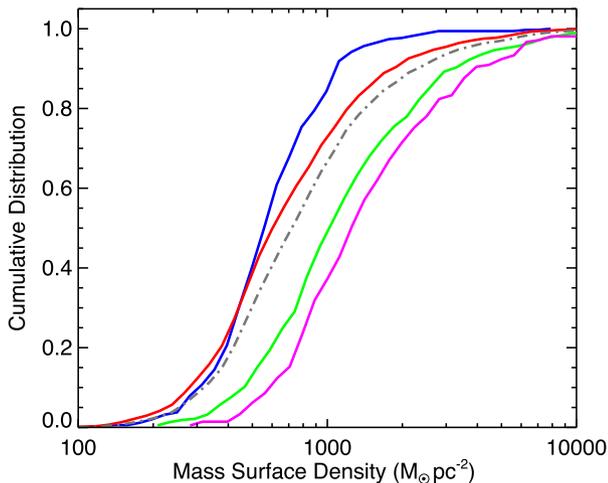
There are two possible explanations for the observed properties of the MSF clumps: (1) that they evolve from lower mass clumps associated with the earlier evolutionary stages by acquiring additional mass and growing in size through global infall, or (2) they are simply larger and more massive at their 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  $0.5\ \text{pc}$  (see Fig. 10), infall velocity of  $1.5\ \text{km s}^{-1}$  and particle density of  $5 \times 10^3\ \text{cm}^{-3}$  (e.g. Wyrowski et al. 2016), the mass infall rate ( $\dot{M}$ ) is  $\sim 1.5 \times 10^{-3} M_\odot\ \text{yr}^{-1}$ . As the time-scales for massive star formation is in the region of several  $10^5\ \text{yr}$  (Mottram et al. 2011; Davies et al. 2011), the total increase in clump mass due to infall ( $166\text{--}500 M_\odot$  for  $1\text{--}3 \times 10^5\ \text{yr}$ ) 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 ( $M_{\text{clump}}/\text{pc}^2$ ) 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  $\sim 0.3$  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  $\sim 6$  per cent. The mean uncertainty in the integrated flux determined from the SED fitting is  $\Delta S_{\nu}/S_{\nu} \sim 42$  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- $\mu\text{m}$  flux ( $\Delta S_{\nu, \text{peak}}/S_{\nu, \text{peak}} \sim 15$  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  $\beta$  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.

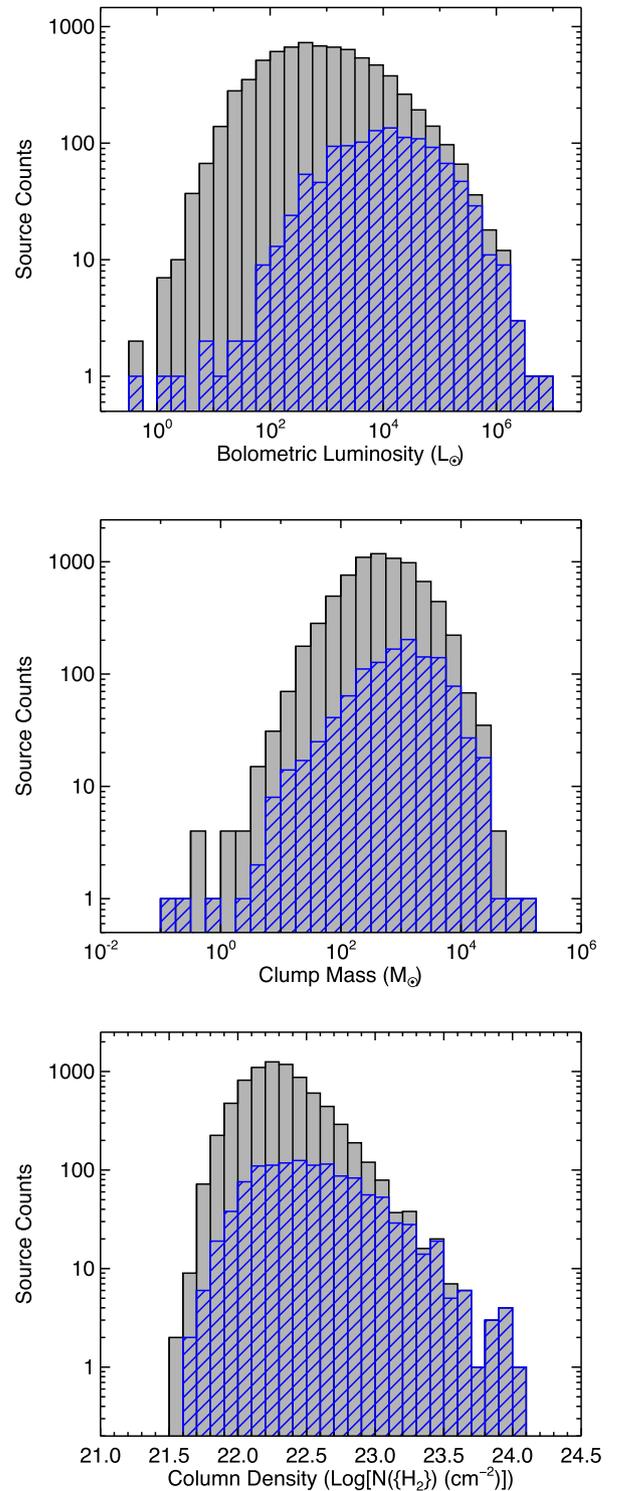
### 5.7 Comparison with previous ATLASGAL results

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  $L_{\text{bol}} = 10^{4.01 \pm 0.03}$ , compared to the new value of  $L_{\text{bol}} = 10^{4.08 \pm 0.03}$ ). 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.

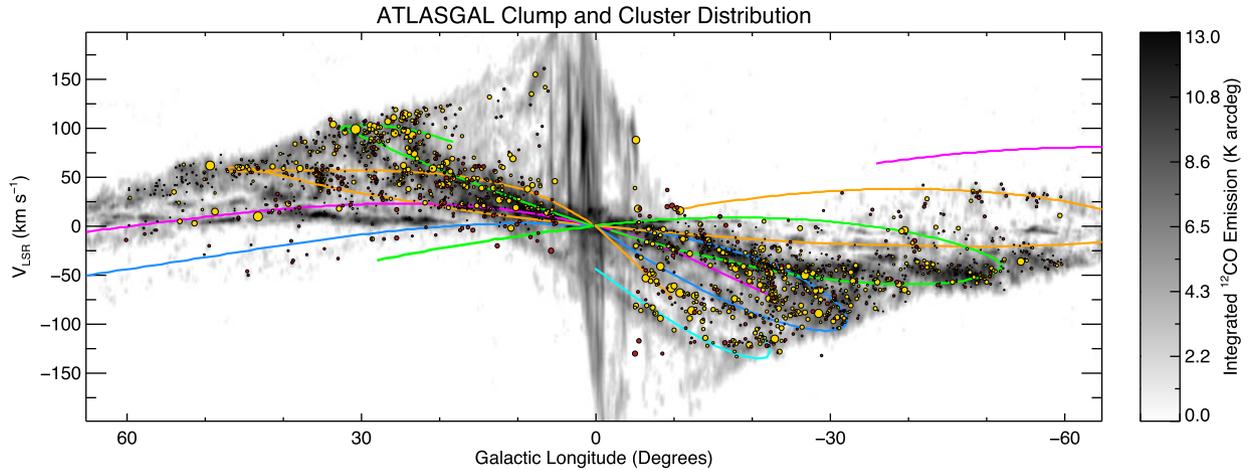
We show the bolometric luminosity distribution of the whole sample and the MSF clumps in the upper panel of Fig. 16. This reveals that the MSF clumps dominate the upper end of the luminosity distribution confirming that our previous efforts of identifying all of the most active sites of massive star 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.  $350^\circ < \ell < 355^\circ$  or  $5^\circ < \ell < 10^\circ$ ); (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.  $186^\circ < \ell < 20^\circ$ ).

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 ( $24 \pm 0.16$  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 ( $\ell$ - $v$ ) 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  $10^{23.5} \text{ cm}^{-2}$  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.

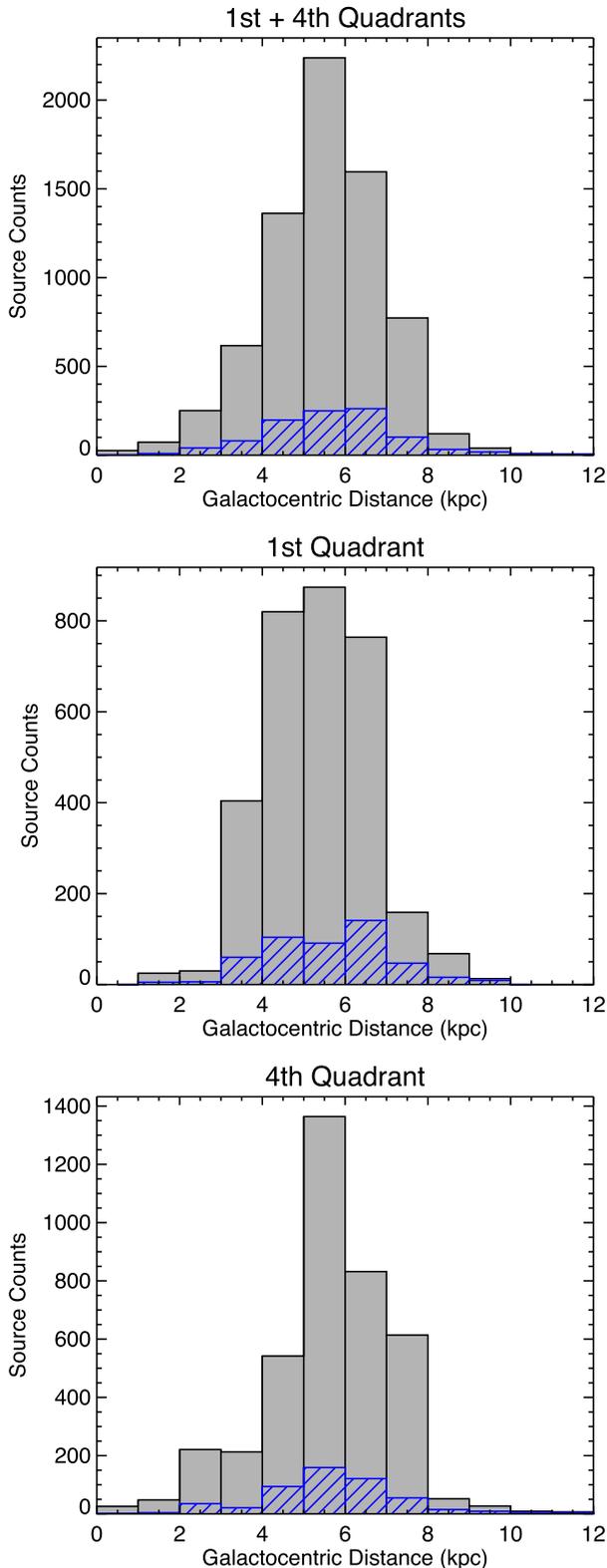
In Fig. 17, we show the longitude–velocity distribution of isolated clumps and clusters. The background grey-scale image is the integrated  $^{12}\text{CO}$  (1–0) emission (Dame et al. 2001), which traces the large-scale distribution of molecular gas. The loci of the spiral arms are overlaid (Taylor & 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  $^{12}\text{CO}$  (1–0) emission. Furthermore, we find that the vast majority of the dense gas is correlated with the spiral-arm loci located within  $30^\circ$  of the GC. Detailed analysis of this correlation is difficult in the inner regions (i.e.  $|\ell| < 40^\circ$ ) due to the high density of sources and the overlapping of the spiral arms in  $\ell v$ -space. Beyond the inner  $40^\circ$  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).

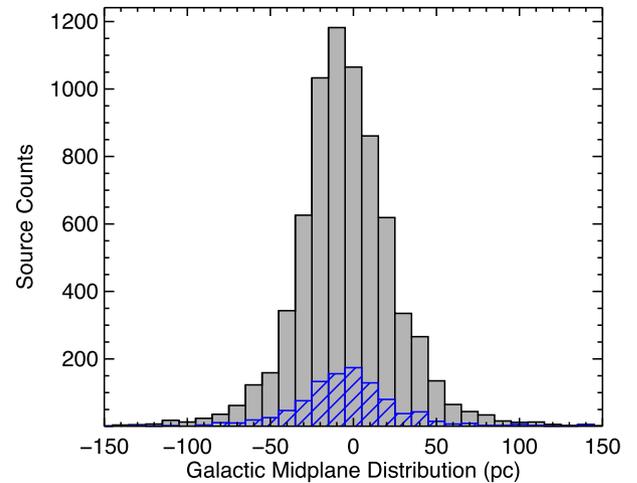
Although the vast majority of ATLASGAL sources are seen to be tightly clustered around the four-arm Milky Way model spiral-arm loci, there are a few small clusters located between the spiral-arm tangents (e.g. between the Scutum-Centaurus and Sagittarius tangents in the first quadrant and between the 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  $^{13}\text{CO}$  (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  $\sim 5.5$  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  $\sim 2.5$  kpc which is likely to be associated with the near side of the 3 kpc expanding arm.

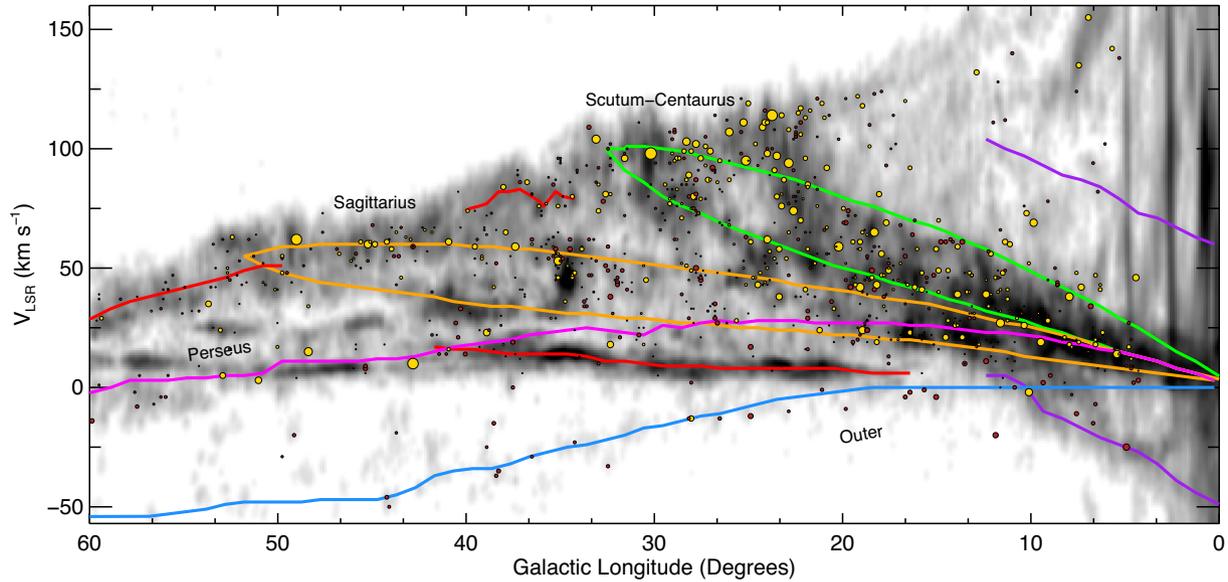
The presence of peaks in the Galactocentric distribution at locations where spiral-arm segments and the Galactic bar are predicted to be support the spiral model of the Galaxy and that there are significant differences in the structure of the first and fourth quadrants. These structural features in the 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 percent 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  $26.04 \pm 0.02$  pc for the whole sample and  $25.15 \pm 0.13$  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.

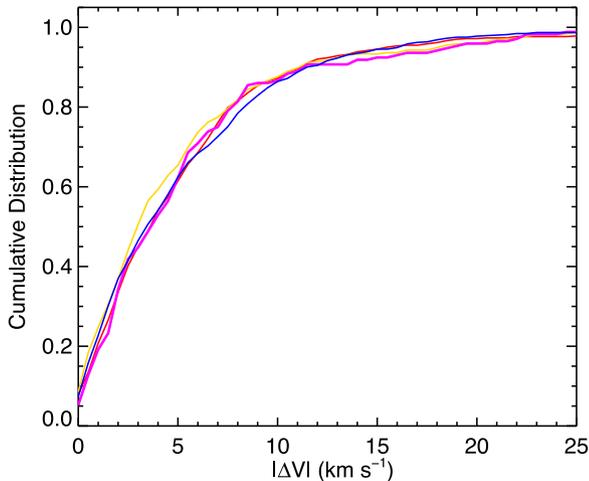
## 6.1 Correlation with spiral arms

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  $^{12}\text{CO}$  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  $10 \text{ km s}^{-1}$ ) and that there is no significant difference between the clusters and isolated clumps. We note that the loci of the spiral arms in  $l$ - $v$ -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  $\pm 7 \text{ km s}^{-1}$  (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  $\sim 90$  per cent appear to be in the process of forming stars. The clumps typically have masses of  $500\text{--}1000 M_{\odot}$  and radii of  $\sim 0.5\text{--}1$  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

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

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-

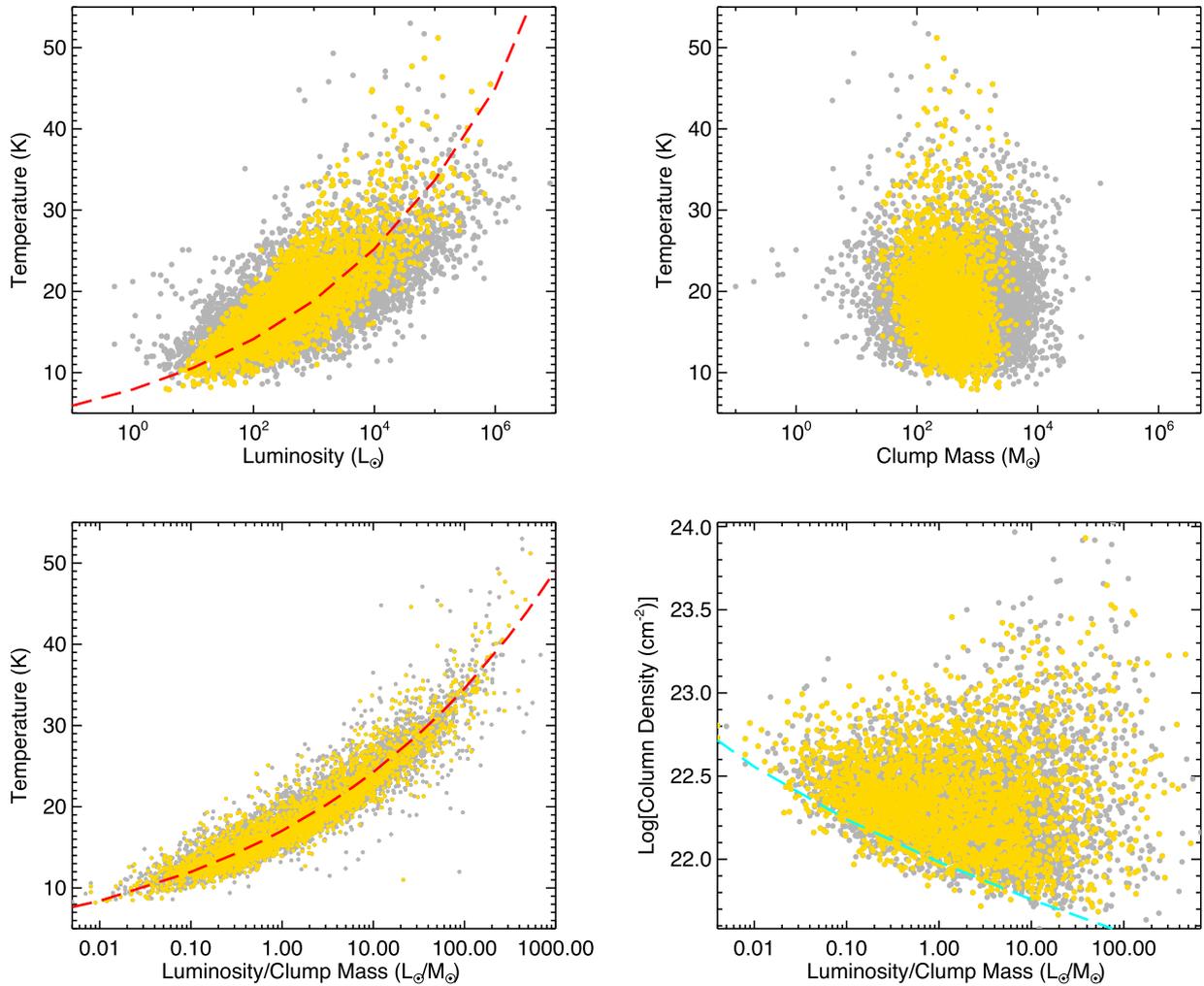
tures of their natal clumps (e.g. Urquhart et al. 2011). We are able to extend this analysis to lower luminosity and less-evolved sources using the ATLASGAL sample. In the upper left panel of Fig. 22, we show the relationship between temperature and luminosity. This plot shows a strong correlation between 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  $0.145 \pm 0.0014$  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  $L_{\text{bol}}/M_{\text{clump}}$  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  $L_{\text{bol}}/M_{\text{clump}}$  and the whole range of measured clump temperatures. The  $L_{\text{bol}}/M_{\text{clump}}$  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  $0.15235 \pm 0.00059$  and an intercept of  $1.231 \pm 0.00052$ ; this results in the power-law relation  $L_{\text{bol}}/M_{\text{clump}} = 10^{(-8.08 \pm 0.04)} \times T^{(6.564 \pm 0.035)}$ .

We have found, then, that both luminosity and  $L_{\text{bol}}/M_{\text{clump}}$  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 ( $\sim 3$  orders of magnitude). The  $L_{\text{bol}}/M_{\text{clump}}$ –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  $L_{\text{bol}}/M_{\text{clump}}$  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  $L_{\text{bol}}/M_{\text{clump}}$  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).

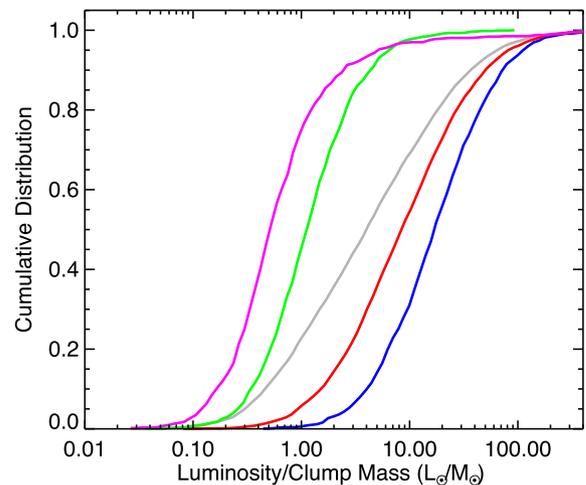
The lower right panel of Fig. 22 shows the scatter plot of column density and the  $L_{\text{bol}}/M_{\text{clump}}$  ratio. There is no obvious correlation between these two parameters, which suggests that the column density does not evolve significantly during the star formation process. The test indicates 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



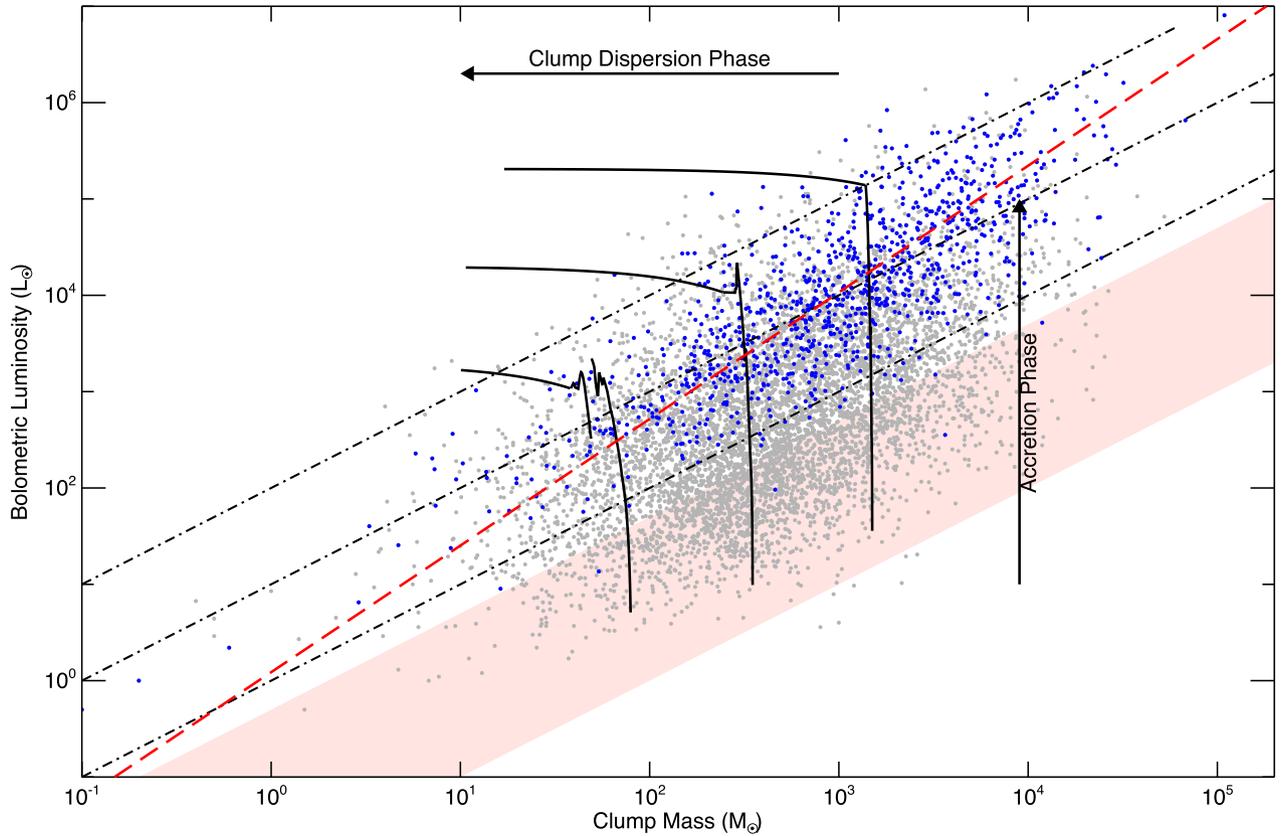
**Figure 22.** Scatter plots of temperature versus luminosity and temperature versus clump mass (upper left and right panels, respectively). The correlation between temperature and  $L_{\text{bol}}/M_{\text{clump}}$  ratio is shown in the lower left panel, while the lower right panel shows the column density and  $L_{\text{bol}}/M_{\text{clump}}$  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  $5\sigma$  sensitivity limit ( $\sim 300 \text{ mJy beam}^{-1}$ ), 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  $L_{\text{bol}}/M_{\text{clump}}$  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  $L_{\text{bol}}/M_{\text{clump}}$  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  $L_{\text{bol}}/M_{\text{clump}}$  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  $L_{\text{bol}}/M_{\text{clump}} = 1, 10$  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 shows the results of a log–log, outlier resistant, linear fit to the MSF clumps; this has a slope of  $1.314 \pm 0.0189$  and an intercept of  $0.087 \pm 0.056$ .

### 7.3 Luminosity–mass relation

We present the luminosity–mass ( $L_{\text{bol}}-M_{\text{clump}}$ ) 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  $\sim 10-15$  K.<sup>3</sup> 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  $L_{\text{bol}}/M_{\text{clump}}$  (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  $L_{\text{bol}}/M_{\text{clump}}$  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  $L_{\text{bol}}/M_{\text{clump}}$  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.

<sup>3</sup> The lower end of the temperature range is determined by heating from the interstellar radiation field and cosmic ray heating ( $\sim 10$  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).

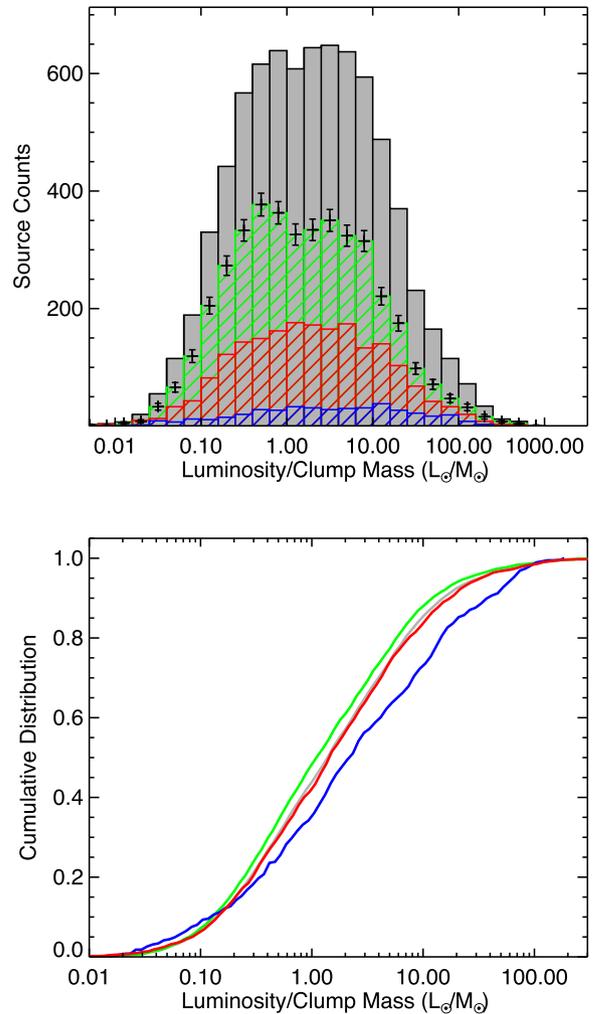
The MSF clumps form a continuous distribution that is concentrated towards an  $L_{\text{bol}}/M_{\text{clump}} \sim 10 L_{\odot}/M_{\odot}$  locus that extends over approximately 4 orders of magnitude in both axes. This clustering around  $10 L_{\odot}/M_{\odot}$  is likely to trace the transition between 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  $L_{\text{bol}}/M_{\text{clump}}$  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 ( $L_{\text{bol}}/M_{\text{clump}} \lesssim 2 L_{\odot}/M_{\odot}$ ), the YSOs accrete material and grow in mass, reaching the zero age main sequence (ZAMS) at  $L_{\text{bol}}/M_{\text{clump}} \sim 10 L_{\odot}/M_{\odot}$ , and start dispersing the parent clump. Compact H II regions become common at  $L_{\text{bol}}/M_{\text{clump}} \gtrsim 40 L_{\odot}/M_{\odot}$ , however, we also note that few regions are found with  $L_{\text{bol}}/M_{\text{clump}}$  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  $L_{\text{bol}}/M_{\text{clump}}$  of  $10 L_{\odot}/M_{\odot}$  (slope  $1.03 \pm 0.05$ ) 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) ( $\sim 1.27$ ) 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,  $1.32 \pm 0.08$ ). The luminosity and clump masses are strongly correlated, with  $r_{\text{A,B,C}} = 0.69$  and  $p \ll 0.001$ , verifying that the correlation is statistically significant.

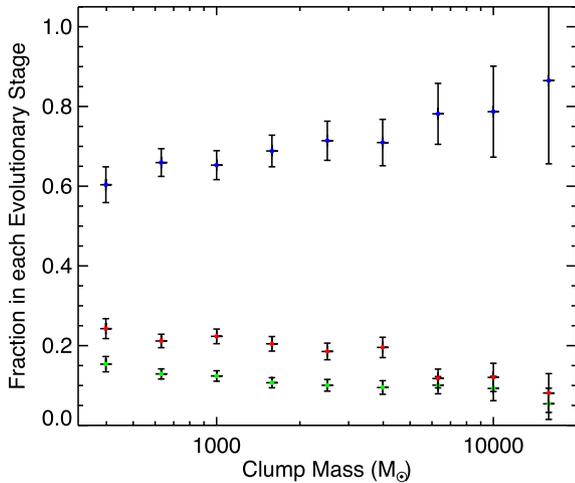
### 7.3.1 Dependence of $L/M$ on clump mass

We have previously noted that the MSF clumps are generally the most luminous and have higher  $L_{\text{bol}}/M_{\text{clump}}$  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  $L_{\text{bol}}/M_{\text{clump}}$  distribution is not the same for the whole mass range and may increase with clump mass. In Fig. 25, we compare the  $L_{\text{bol}}/M_{\text{clump}}$  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  $L_{\text{bol}}/M_{\text{clump}}$  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  $\log(L_{\text{bol}}/M_{\text{clump}})$  ratios for the increasing mass bins are  $0.60 \pm 0.012$ ,  $0.80 \pm 0.022$  and  $1.10 \pm 0.052 L_{\odot}/M_{\odot}$ , respectively. Closer inspection of the lowest mass range reveals a bimodal distribution with peaks a few tenths of a dex to either side of  $L_{\text{bol}}/M_{\text{clump}} = 1 L_{\odot}/M_{\odot}$  (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  $L_{\text{bol}}/M_{\text{clump}}$  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  $L_{\text{bol}}/M_{\text{clump}}$  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  $L_{\text{bol}}/M_{\text{clump}}$  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 ( $70 \mu\text{m}$  dark), protostellar ( $24 \mu\text{m}$  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  $\sim 15$  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  $70 \mu\text{m}$ . This supports the hypothesis that the statistical lifetime of the earliest stages



**Figure 26.** Fractions of clumps sorted by evolutionary stage. The green, red and blue circles indicate the fractions of quiescent, protostellar and mid-IR bright clumps (YSOs and MSF clumps), respectively, as a function of clump mass. The uncertainties are determined from Poisson counting statistics. This plot only includes mass bins above the nominal survey completeness ( $\sim 400 M_{\odot}$ ), and averages the source counts in bins of 0.2 dex.

decreases with the mass of the clump (e.g. Mottram et al. 2011; Davies et al. 2011).

It is also clear that the fraction of protostellar clumps diminishes with increasing clump mass: initially the decrease is quite modest but becomes noticeable for clump masses over a few times  $10^3 M_{\odot}$  where it drops from  $\sim 20$  per cent to  $\sim 10$  per cent. This suggests that once star formation begins, it evolves more rapidly in the more massive clumps than it does in the lower mass clumps.

At this point, it is worth noting that the fraction of star-forming clumps identified by Elia et al. (2017) is significantly lower than that found with our sample ( $\sim 25$  per cent; see also Ragan et al. 2016). They are, however, approximately 100 times more sensitive than ATLASGAL, and many of the 100 000 sources detected by HiGAL are likely to be lower mass clumps. The trend seen in Fig. 26 of lower SFF with decreasing clump mass would likely explain the difference in the SFF observed between the results reported here and those reported by Elia et al. (2017).

Ordinarily, the dip seen in the full-sample histogram in the upper panel of Fig. 25 might not have been viewed as notable. As already

shown, this dip is significant in the lowest mass clump distribution, and therefore suggests that this may be connected to the statistical lifetimes of the different evolutionary stages. We have already noted that the lifetimes of the quiescent and protostellar stages decrease with clump mass, and the bimodal distribution seen in the distribution of lower mass clumps might be related to the pre-stellar phases. Conversely, the smoothness of the distributions of the higher mass clumps may indicate that there is effectively no observable pre-stellar stages for these larger clumps. This may seem to contradict the pre-stellar lifetime determined from Fig. 26, but this quantity is estimated from the proportion of clumps that are dark at  $70 \mu\text{m}$ : as mentioned earlier, not all of these are necessarily pre-stellar as some have been associated with outflows.

The lifetimes of massive stars has been investigated by Davies et al. (2011) and Mottram et al. (2011), both of which determined the lifetimes for  $\text{H II}$  regions to be several  $10^5$  yr. Mottram fitted the empirical data and derived the following relationship for the lifetime as a function of the source luminosity: ( $\log(t_{\text{HII}}/\text{yr}) = (-0.13 \pm 0.16) \times \log(L/L_{\odot}) + (6.1 \pm 0.8)$ ). We can obtain an estimate of the clump lifetimes by combining this with the relationship we derived for the clump mass and luminosity discussed in Section 7.3. If we assume that the  $\text{H II}$  region lifetime dominates the time-scales of all of the embedded stages (Motte et al. 2007), we can use the fraction of sources in each stage to estimate the statistical lifetime of each stage.

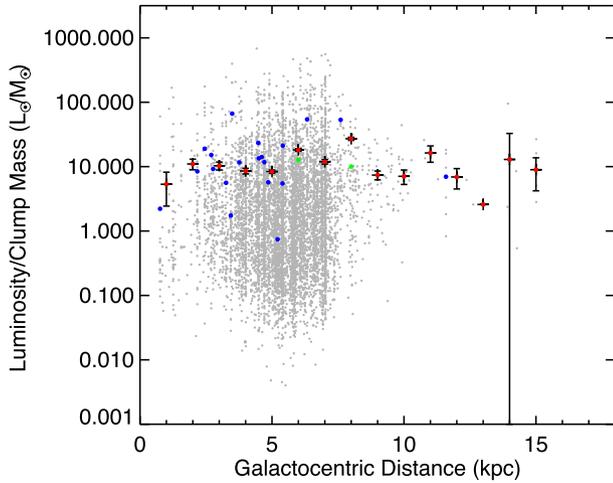
In Table 6, we present the statistical lifetimes for the various evolutionary stages as a function of clump mass. The lifetime of the quiescent stage is approximately  $5 \times 10^4$  yr for clumps with masses  $\sim 1000 M_{\odot}$ , and decreases to  $2 \times 10^4$  yr for clumps with masses of  $\sim 1 \times 10^4 M_{\odot}$ . These values are broadly consistent with the statistical lifetimes of massive starless clumps as determined by a number of studies reported in the literature (e.g. Motte, Bontemps & Louvet 2017; Csengeri et al. 2014; Tackenberg et al. 2012; Motte et al. 2007). We also note that the lifetime of the quiescent stage for clump masses above a few times  $10^4 M_{\odot}$  decreases to negligible values.

### 7.3.2 Dependence on L/M on Galactocentric distance

In Section 6, we investigated the Galactocentric distribution of the ATLASGAL clumps, and this revealed a number of peaks that coincided with the expected positions of various spiral arms. These

**Table 6.** Statistical lifetime of clumps as a function of clump mass.

$\log[M_{\text{clump}}]$ ( $M_{\odot}$ )	$\log[\text{luminosity}]$ ( $L_{\odot}$ )	Statistical lifetimes ( $10^5$ yr)				
		$t_{\text{HII}}$	Quiescent	Protostellar	YSO	MSF
2.60	3.50	4.41	0.68	1.07	2.25	0.41
2.80	3.77	4.08	0.53	0.86	2.08	0.61
3.00	4.03	3.77	0.47	0.84	1.80	0.66
3.20	4.29	3.48	0.37	0.71	1.68	0.71
3.40	4.56	3.22	0.32	0.60	1.58	0.72
3.60	4.82	2.98	0.28	0.58	1.20	0.91
3.80	5.08	2.75	0.28	0.32	1.19	0.96
4.00	5.34	2.54	0.24	0.31	1.01	0.99
4.20	5.61	2.35	0.13	0.19	1.14	0.89
4.40	5.87	2.17	0.08	0.50	0.42	1.17
4.60	6.13	2.01	–	–	1.34	0.67
4.80	6.39	1.86	–	–	0.93	0.93
5.00	6.66	1.72	–	–	–	1.72



**Figure 27.**  $L_{\text{bol}}/M_{\text{clump}}$  as a function of Galactocentric distance. The grey and blue circles show the distribution of all of the individual ATLASGAL clumps and the 30 most massive clusters, respectively. The red circles and the error bars show the mean values and the standard error of the mean. The two green circles show the  $L_{\text{bol}}/M_{\text{clump}}$  ratios for the 6 and 8 kpc bins after the W51 and W49 have been excluded. These red and green circles are evaluated in bins of 1 kpc.

correlations suggest that the spirals play a significant role in the star formation process, although the nature of this role is still unclear.

As mentioned in Section 1, previous efforts by Moore et al. (2012) to evaluate the role of the spiral arms reported higher average luminosities and molecular cloud masses at well-defined Galactocentric distances in the first quadrant. They found that  $L_{\text{MYSO}}/M_{\text{cloud}}$  was relatively flat for the inner 5 kpc, and although two significant peaks were detected at  $\sim 6$  and  $\sim 8$  kpc, these peaks were attributed to the presence of W51 and W49 at these distances. The work of Moore et al. (2012) has been extended to investigate the SFE along lines of sight centred at  $\ell = 30^\circ$  and  $40^\circ$  (Eden et al. 2013, 2015) using GLIMPSE  $8 \mu\text{m}$ , WISE  $12 \mu\text{m}$  and  $22 \mu\text{m}$  and Hi-GAL  $70 \mu\text{m}$  to estimate the IR luminosity and the clump masses calculated with the BGPS. These studies estimated the SFE for spiral arms and interarm regions, and found that there was no significant variation either between the different arms or the interarm regions on the kiloparsec scale.

These studies suggest that the spiral arms play a role in aggregating material but that the increase in star formation is perhaps the result of source crowding within the spiral arms, and not due to the influence of the spiral arms themselves. It is worth noting that these studies have been limited to either small regions or limited sample sizes, and all have been conducted within the first quadrant. We are able to test these results and extended them to a larger fraction of the Galactic plane.

Fig. 27 presents the  $L_{\text{bol}}/M_{\text{clump}}$  distribution as a function of Galactocentric distance for clumps (grey circles), clusters (blue circles) and for all sources averaged over kiloparsec scales (red circles). This plot reveals that although there are significant variations in the clump to clump  $L_{\text{bol}}/M_{\text{clump}}$  ratios, this decreases when we move to the larger clusters and becomes almost constant when averaged over the largest scales; the clumps have typical sizes of  $0.5\text{--}1 \text{ pc}$ , while the clusters have typical sizes of a few tens of parsecs.

We observe two significant peaks at  $\sim 6$  and  $8 \text{ kpc}$  in the 1 kpc bins, coincident with those previously identified by Moore et al. (2012) as W51 and W49, the two most active star-forming regions in the Galaxy. To check that the peaks seen in our data are also

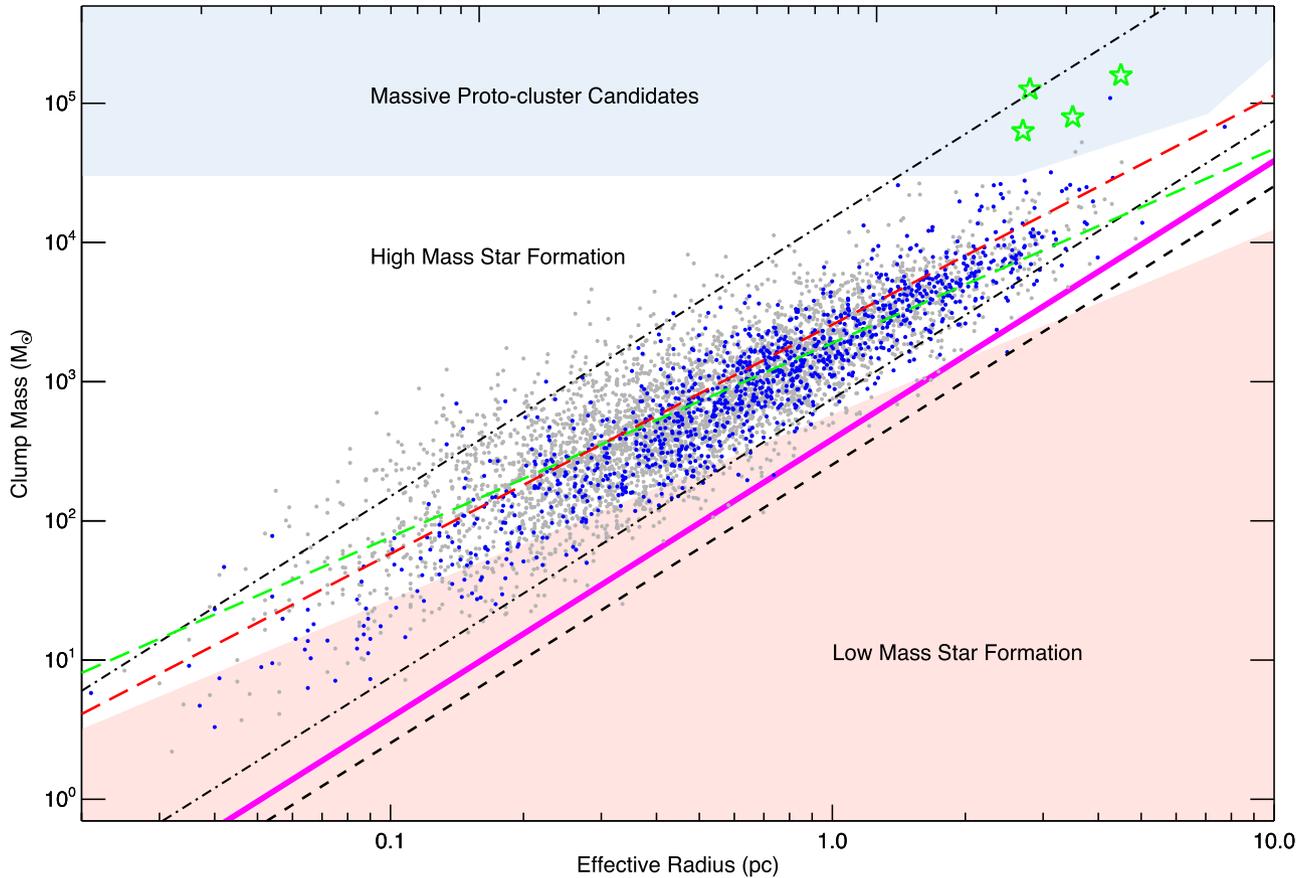
due to the presence of these extreme regions, we have recalculated the  $L_{\text{bol}}/M_{\text{clump}}$  ratio in these two distance bins (these values are shown as green circles on Fig. 27). A comparison between the red and green circles at  $\sim 6$  and  $8 \text{ kpc}$  clearly shows that the exclusion of these two extreme star-forming regions results in a reduction of the  $L_{\text{bol}}/M_{\text{clump}}$  ratio which brings these more in line with the  $L_{\text{bol}}/M_{\text{clump}}$  ratios on either side. We therefore find that the  $L_{\text{bol}}/M_{\text{clump}}$  ratio is relatively flat between 2 and 9 kpc when evaluated over kiloparsec scales. There is, however, significant variation on smaller scales, and this variation increases as the physical scales decrease. This is fully consistent with the results of the previous studies (Moore et al. 2012; Eden et al. 2013, 2015).

#### 7.4 Mass–radius distribution

We present a mass–radius ( $M_{\text{clump}}\text{--}R_{\text{eff}}$ ) diagram of the whole distance-determined sample of ATLASGAL sources in Fig. 28. We have shown similar diagrams in the previous papers in this series and found that the MSF clumps exhibit a strong correlation between these parameters. We extend our previous analysis by significantly increasing the sample size, and including lower luminosity and less-evolved clumps as well as providing more reliable measurements for the MSF clumps resulting from the improved dust temperature estimates. We find the populations of the MSF clumps and non-MSF clumps to be similarly distributed, forming a continuous spread over almost 3 orders of magnitude in radius and almost 5 orders of magnitude in clump mass. The diagonal upper and lower dashed–dotted lines shown in Fig. 28 indicate lines of constant surface density,  $\Sigma(\text{gas})$ , of 1 and  $0.05 \text{ g cm}^{-2}$ , respectively; these provide a reasonable empirical fit to the upper and lower range of the whole sample of clumps (this range encompasses  $\sim 90$  per cent). We also show the minimum threshold of  $\sim 116\text{--}129 M_{\odot} \text{ pc}^{-2}$  (thick magenta band) proposed by Lada et al. (2010) and Heiderman et al. (2010, hereafter referred to as the ‘LH threshold’) for ‘efficient’ star formation in nearby molecular clouds ( $d \leq 500 \text{ pc}$ ).

The LH threshold seems to represent a stronger constraint for the lower limit for the ATLASGAL clumps, although the vast majority of clumps have significantly higher mean surface densities. If we expect the clumps to form gradually through global accretion from larger scales, and to undergo a significant amount of evolution before star formation, we might expect to find that the population of MSF clumps dominates the higher surface-density region of the narrow range and the non-MSF clumps dominate the region near the  $0.05 \text{ g cm}^{-2}$  threshold; this does not, however, appear to be the case. We have already found that there is no correlation between clump mass and evolution, and as there is no clear separation between the different evolutionary stages we may surmise that the surface density does not change significantly either.

As seen in the previous papers, there is a strong correlation between these parameters for the MSF clumps ( $r_{\text{AB,C}} = 0.85$ ) and they are located within a relatively narrow range of surface densities. The correlation coefficient for the complete sample shown in Fig. 28 is  $r_{\text{AB,C}} = 0.73$  with  $p \ll 0.01$ ). This is slightly lower than that found for the MSF clumps alone, but there is noticeably more scatter associated with the non-MSF clumps, which may include clumps that will remain starless or are destined to form only lower mass stars. We have fitted the whole sample and a distance-limited subset of the sample (between 2 and 5 kpc) to determine the slopes: these are plotted on Fig. 28 as red and green dashed lines, respectively. The slope obtained for the whole sample is  $1.647 \pm 0.012$ , which is in excellent agreement with the slope determined for the MSF clumps reported in Paper III ( $1.67 \pm 0.025$ ). The slope found for



**Figure 28.** The mass–size relationship of all ATLASGAL clumps (grey) and those associated with MSF (blue). The green stars show the distribution of the MPC candidates found towards the GC (Longmore et al. 2012; Immer et al. 2012). The light red shaded region shows the part of the parameter space found to be devoid of massive star formation that satisfies the relationship  $m(r) \leq 580 M_{\odot} (R_{\text{eff}}/\text{pc})^{1.33}$  (cf. Kauffmann & Pillai 2010). The light blue shaded area towards the top of the diagram indicates the region of parameter space where YMC progenitors are expected to be found (i.e. Bressert et al. 2012). The long-dashed red line shows the result of a linear power-law fit to the whole sample of associated clumps, while the green line shows the fit to the distance-limited sample. The dashed black line shows the sensitivity of the ATLASGAL survey ( $N_{\text{H}_2} \sim 10^{22} \text{ cm}^{-2}$ ) and the upper and lower dot-dashed lines mark surface densities of 1 and  $0.05 \text{ g cm}^{-2}$ , respectively. The diagonal magenta band indicates the gas surface density  $[\Sigma(\text{gas})]$  parameter space between  $116$  and  $129 M_{\odot} \text{ pc}^{-2}$ , suggested by Lada, Lombardi & Alves (2010) and Heiderman et al. (2010), respectively, to be the threshold for ‘efficient’ star formation.

the distance-limited sample is shallower ( $1.397 \pm 0.014$ ) and more reliable.

The parameter space covered in Fig. 28 can be separated into three regions of interest: the region where low-mass stars are found (light red shaded region), the region where efficient high-mass star formation is expected to be found (unshaded region; Kauffmann et al. 2010) and the region where massive protocluster (MPC) candidates would be located (light blue shaded region – compact clumps with a minimum mass of  $3 \times 10^4 M_{\odot}$ ; Bressert et al. 2012). We find that the vast majority of our sources are located above the empirical threshold required for efficient massive star formation, and so our expectation that most of these clumps are capable of forming massive stars is justified. We also note that only three sources have sufficient mass and are compact enough to fulfill the criteria to be considered as viable MPC candidates (we will discuss these further in Section 7.4.1).

#### 7.4.1 Galactic population of young massive protoclusters

Young massive clusters (YMCs, e.g. Arches, Quintuplet) are the most active clusters in the Galaxy. These can be broadly defined as being younger than 100 Myr and having stellar masses  $\gtrsim 10^4 M_{\odot}$  (Portegies Zwart, McMillan & Gieles 2010). Identifying

the precursors to these rare and extreme clusters will provide some insight into their initial conditions and the impact of the environment on their formation and evolution. Bressert et al. (2012) suggested a set of threshold criteria that a clump would need to satisfy in order to form a YMC (i.e. massive and compact enough to ensure that the gravitational force was sufficiently strong to counteract the enormous feedback long enough for the YMC to form). Assuming an SFE of 30 per cent, Bressert et al. (2012) estimated the progenitors of these YMCs would need a minimum mass of  $3 \times 10^4 M_{\odot}$  and have a radius of no more than a few parsecs: clumps that satisfied these criteria are referred to as MPCs.

There have been a number of studies that have used survey data to search for MPCs both towards the GC region (Longmore et al. 2012; Immer et al. 2012) and in the disc (Ginsburg et al. 2012; Longmore et al. 2017; Contreras et al. 2017). In the previous papers in this series, we have also identified 16 MPC candidates located in the disc that are associated with MSF tracers, many of which have been independently identified in these other studies (see Table 7 for MPC candidates and references to studies that have discovered them). Our previous search for these objects used a temperature of 20 K to estimate clump masses, and was further biased to clumps associated with MSF tracers and so excluded the majority of clumps ( $\sim 80$  per cent). Here, we are able to extend and improve on our

**Table 7.** Derived parameters for MPC candidates identified here and in Paper III. Masses have been estimated assuming dust temperatures derived from the SED fits for all sources except G002.53+00.016, for which a range of temperatures between 19 and 27 K has been estimated from SED fits to individual pixels across the clump (see Longmore et al. 2012 for details).

ATLASGAL name	Complex name	Association type	Distance (kpc)	$R_{\text{eff}}$ (pc)	$T_{\text{dust}}$ (K)	$\log[M_{\text{clump}}]$ ( $M_{\odot}$ )	$\log[N(\text{H}_2)]$ ( $\text{cm}^{-2}$ )	$\log[L_{\text{bol}}]$ ( $L_{\odot}$ )	$L_{\text{bol}}/M_{\text{clump}}$ ( $L_{\odot}/M_{\odot}$ )	Reference
Galactic disc										
AGAL043.166+00.011	W49	MSF	11.1	4.255	33.3	5.04	23.892	6.91	74.0	1,2,4,5
AGAL341.932–00.174	G341.982–00.125	YSO	12.4	3.549	14.9	4.65	22.944	4.75	1.3	1
AGAL341.942–00.166	G341.982–00.125	YSO	12.4	3.670	14.4	4.72	23.214	4.82	1.2	1
AGAL010.472+00.027	G010.47+00.02	MSF	8.5	2.320	25.1	4.41	23.803	5.65	17.4	2,3,5
AGAL012.208–00.102	G012.203–00.117	MSF	13.4	2.724	24.4	4.44	23.340	5.52	11.8	2,5 <sup>a</sup>
AGAL019.609–00.234	G019.649–00.239	MSF	12.6	2.749	29.6	4.36	23.447	6.01	44.6	2,4,5 <sup>a</sup>
AGAL032.797+00.191	G032.768+00.192	MSF	13.0	2.450	34.2	4.15	23.171	6.10	87.8	2,5 <sup>a</sup>
AGAL043.148+00.014	W49	MSF	11.1	2.316	31.2	4.26	22.894	5.93	46.5	2,4
AGAL043.164–00.029	W49	MSF	11.1	3.124	31.2	4.50	23.265	6.21	50.3	2,4,5
AGAL043.178–00.011	W49	MSF	11.1	2.155	31.9	4.23	22.882	6.60	236.5	2
AGAL049.472–00.367	W51	MSF	5.3	2.124	30.3	4.34	23.420	6.38	109.7	2,4,5
AGAL049.482–00.402	W51	MSF	5.3	2.176	28.6	4.26	23.372	6.17	81.8	2
AGAL049.489–00.389	W51	MSF	5.3	1.408	31.6	4.41	24.025	6.29	76.2	2,4,5
AGAL328.236–00.547	RCW99	MSF	2.7	1.178	–	–	–	–	–	2,3,8
AGAL329.029–00.206 <sup>c</sup>	G329.111–00.278	MSF	2.9	0.807	20.6	3.45	23.358	4.18	5.5	2,3,8
AGAL330.954–00.182	G331.394–00.125	MSF	5.3	1.238	32.2	3.95	23.790	5.78	67.5	2
AGAL337.704–00.054	G337.651–00.048	MSF	12.1	2.403	23.6	4.42	23.429	5.41	9.8	2,8
AGAL350.111+00.089	G350.219+00.081	MSF	10.5	1.940	27.3	4.29	23.080	5.83	34.2	2,3,9
Galactic Centre										
G000.253+00.016	‘The Brick’	Starless?	8.4	2.8	–	5.1	23.39	–	–	6,7
AGAL000.411+00.051 (d) <sup>c</sup>	–	Starless?	8.4	3.5	–	4.9	23.23	–	–	7
AGAL000.476–00.007 (e) <sup>c</sup>	–	Starless?	8.4	4.5	–	5.2	23.53	–	–	7
AGAL000.494+00.019 (f) <sup>c</sup>	–	MM	8.4	2.7	–	4.8	23.26	–	–	7

References: (1) This work, (2) Paper III, (3) Paper II, (4) Paper I, (5) Ginsburg et al. (2012), (6) Longmore et al. (2012), (7) Immer et al. (2012), (8) Contreras et al. (2017), (9) Longmore et al. (2017).

Notes: <sup>a</sup>Identified by Ginsburg et al. (2012) as a massive clump, but the derived mass was slightly less than the  $3 \times 10^4 M_{\odot}$  required for inclusion in their sample of MPC candidates.

<sup>b</sup>The distance to this source is uncertain and should be used with caution.

<sup>c</sup>The source names have been taken from the ATLASGAL CSC (Contreras et al. 2013), while the letters in parentheses are the nomenclature used by Lis et al. (1999).

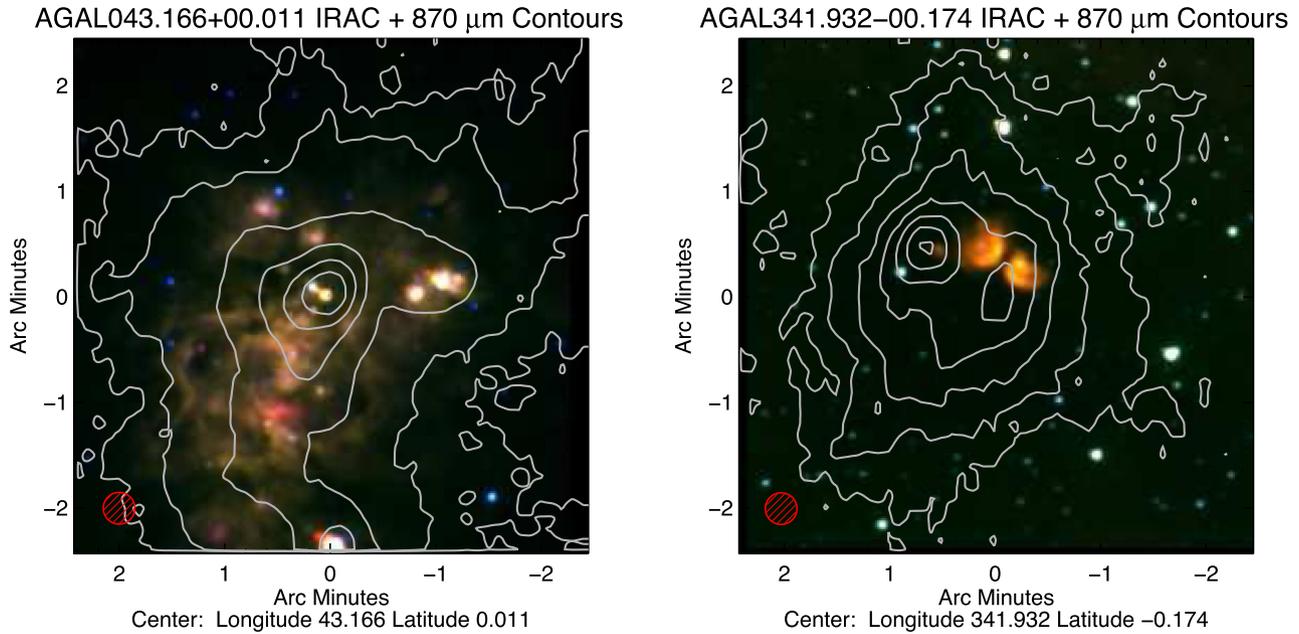
previous work by using the dust temperatures and distances determined.

As described in the introduction, the ATLASGAL CSC is likely to include  $\sim 90$  per cent of all massive clumps in the Galaxy and is sensitive to all clumps with a surface density  $> 3600 M_{\odot} \text{ beam}^{-1}$ . Our sample will therefore include all MPC candidates within the Galactic disc that are located away from the GC ( $355^{\circ} < \ell < 5^{\circ}$ ). As we have determined distances to 97 per cent of the sample (7770 of 8002 clumps), we are very likely to be able to identify all MPC candidates in the region and perhaps in the whole Galaxy. As mentioned in Section 7.4, the comparison of our data and the MPC parameter space has only resulted in the identification of 3 MPC candidates: AGAL043.166+00.011, AGAL341.932–00.174 and AGAL341.942–00.166. Only one of the previously identified clumps remains in the sample: three of the other sources have been moved to the near distance (AGAL328.236–00.547, AGAL329.029–00.206 and AGAL330.954–00.182), but the dust temperatures for the others are significantly higher than the 20 K previously assumed, and this has resulted in lower calculated clump masses which no longer satisfy the mass criterion.

The two new clumps (AGAL341.932–00.174 and AGAL341.942–00.166) are part of the same larger scale molecular structure, as can be seen in Fig. 29. Although neither of these sources has been previously associated with an MSF tracer, the spatial correlation between the extended mid-IR emission and

peaks in the dust emission seen in the image suggest that star formation is already underway. These clumps have been placed at the far distance from the analysis presented in this paper: as we have not been able to find any confirmation of this distance in the literature, their classification as MPC candidates is not considered robust. The remaining clump, AGAL043.166+00.011, is the central part of W49, which is one of the most active star formation regions in the Galaxy, and has a well-constrained maser parallax distance. All three MPCs identified in the disc appear to be actively star forming, although AGAL043.166+00.011 has an  $L_{\text{bol}}/M_{\text{clump}}$  ratio  $\sim 50$  times larger than the other two candidates and therefore is much more evolved.

We have found no starless MPC candidates in the disc and, given the completeness of the survey, it seems unlikely that there are any others to be found. This is consistent with the studies of Ginsburg et al. (2012), Tackenberg et al. (2012), Svoboda et al. (2016) and Longmore et al. (2017), who also failed to find any starless MPC candidates, although one potential starless candidate (AGAL331.029–00.431) has been identified by Contreras et al. (2017). This source is located inside the Solar circle and is therefore subject to the KDA as pointed out by Contreras et al. (2017). They place this at the far distance due to the absence of a correlation with an infrared dark cloud (IRDC), but point out that this source is placed at the near distance by Wien et al. (2015). Our independent assessment of the H I data has associated this source with



**Figure 29.** Three-colour mid-IR GLIMPSE image of the two regions associated with MPC candidates discussed in the text. The contours trace the dust emission detected by ATLASGAL and the red circle shown in the lower left corner of each map indicates the resolution of the ATLASGAL survey ( $\sim 19$  arcsec). In the left-hand panel, we show AGAL043.166+00.011, which is one of most active star formation regions in the Galaxy. In the right-hand panel, we show the AGAL341.942–00.166 and AGAL341.932–00.174 (left and right contour peaks, respectively).

the G331.104–00.413 cluster (see upper panel of Fig. 30), which consists of 53 clumps, 44 of which are placed at the near distance, three at the far distance and six are uncertain. As  $\sim 90$  per cent of the distance solutions point to a near distance solution, the near distance assignment has been adopted for this cluster. This region was also included in a survey of H II regions reported by Caswell & Haynes (1987, 331.314–0.336) who also placed it at the near distance. We are therefore relatively certain that this source is located at the near distance, indicating that this clump is not a viable MPC candidate.

Three of the other four MPC candidates identified by Contreras et al. (2017, private communication) had already been identified in our previous papers (i.e. Paper III; AGAL328.236–00.547, AGAL329.029–00.206 and AGAL337.704–00.054). The other MPC identified in the Contreras et al. (2017) study that was not already identified is AGAL348.183+00.482. This is placed at the far distance by Contreras et al. (2017) and the MALT90 team, but we find this source is associated with the RCW120 (SH 2–3) region (see middle panel of Fig. 30; this is identified as ATLASGAL cluster G348.240+00.507 in this paper). This cluster is associated with 29 clumps, all but one of which have been placed at the near distance in agreement with Caswell & Haynes (1987, 348.225+0.459). It seems likely, therefore, that AGAL348.183+00.482 is located at the near distance and so can also be discarded as a MPC candidate.

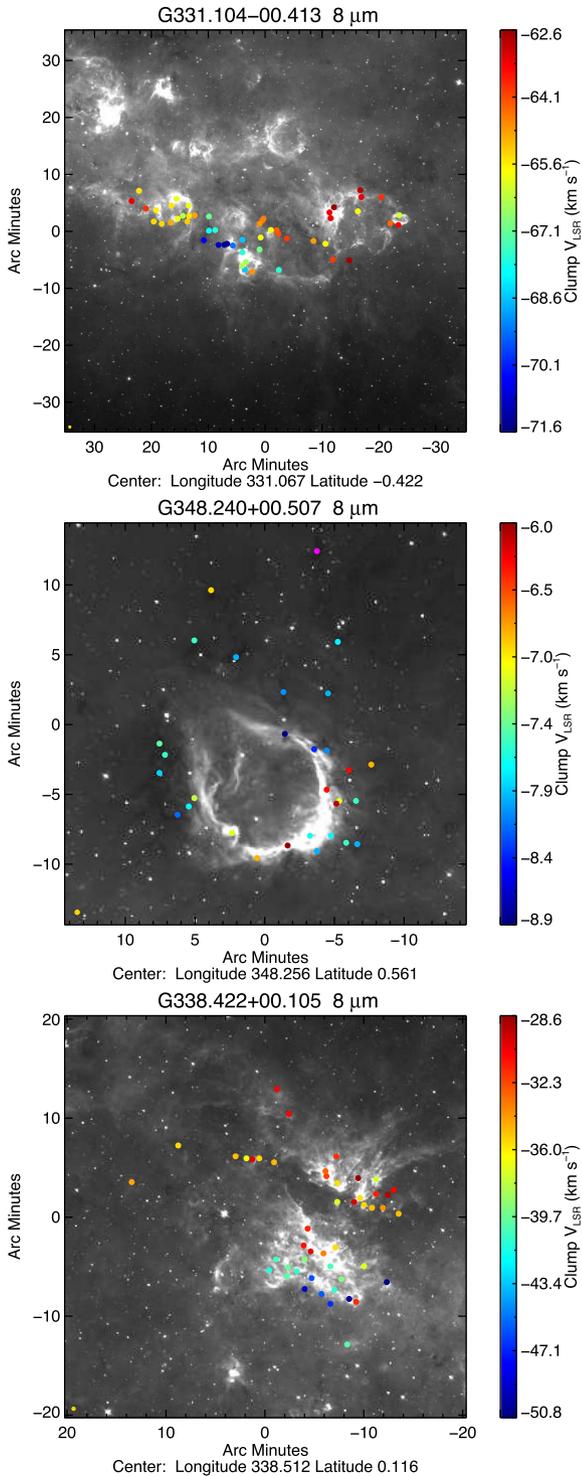
Longmore et al. (2017) has used the H<sub>2</sub>O Southern Galactic Plane Survey (HOPS) study to identify a sample of seven MPCs, but they discard four as either being located at the near distance or being too diffuse to be able to collapse to form an MPC on a realistic time-scale. This leaves three possible candidates from their sample: G330.881–0.371, G338.464+0.034 and G350.170+0.070. The first of these is associated with our cluster G331.104–00.413 (see upper panel of Fig. 30), which we have already discussed and is located at the near distance ( $\sim 4$  kpc). The second is associated with ATLASGAL cluster G338.422+00.105 (see lower panel of Fig. 30), which we have placed at a distance of 2.7 kpc; these two MPC

candidates are therefore less probable. The one remaining MPC candidate identified by Longmore et al. (2017, G350.170+0.070) was included in the sample we presented in Paper III.

We note that Elia et al. (2017) have identified 22 MPC candidates between  $\ell = 16^\circ$  and  $67^\circ$ , however, these results from a default assignment at the far distance for all sources where the KDA has not been resolved. Until reliable distances have been derived for these sources their identification as MPC candidates is somewhat speculative.

Between them, the Longmore et al. (2017) and Contreras et al. (2017) studies have identified 12 MPC candidates. For the reasons discussed in the previous paragraphs only four are likely to be viable, all of which were identified in our previous analysis. We include all of the MPC candidates that were identified in Paper III with their updated parameters in Table 7. We have updated the references to indicate those that have been confirmed by Longmore et al. (2017) and Contreras et al. (2017). Only the top three are still considered to be reliable MPC candidates (i.e.  $\log(M_{\text{clump}}) > 4.48 M_\odot$ ), but we include the others to provide updated physical parameters for them. Although they do not satisfy the mass criterion, these still represent the most massive and compact clumps in the Galaxy, and some of these are close enough to the threshold that a small change in the distance or dust emissivity solution might bring some back into play.

Our hunt for MPC candidates through a large fraction of the inner Galactic disc has only identified three MPC candidates, two of which are associated with the same cloud. We have also assumed an SFE of  $\sim 30$  per cent, which is an upper limit to the SFEs that have been determined in nearby cloud studies (Lada & Lada 2003). If a more modest SFE of 10 per cent is assumed then the number of MPC candidates in the disc reduces to one (i.e. AGAL043.166+00.011). These objects are therefore extremely rare in the disc, and given that they all are associated with mid-IR emission, it is likely that star formation is already ongoing in all of them. Assuming the



**Figure 30.** ATLASGAL clusters associated with MPC candidates identified by Contreras et al. (2017) and Longmore et al. (2017) identified by the friends-of-friends analysis. The background image is the 8 μm emission and filled circles show the positions of ATLASGAL CSC objects, the colours of which provide an indication of their velocities (see colour bars for values).

lifetime of a YMC to be of order 10 Myr and the number of clusters with masses between  $10^4$  and  $10^5 M_{\odot}$  to be  $\sim 25$ , Longmore et al. (2012) estimated the YMC formation rate to be one every  $\sim 2.5$  Myr, which agrees well with the number we have detected. Given the low number found in the disc, it is extremely unlikely that one will be

detected in a pre-stellar stage, which is also consistent with what we have found here.

The situation in the disc stands in contrast to that found in the GC region, where twice as many MPC candidates have been identified including three of which currently appear to be starless (Longmore et al. 2012; Immer et al. 2012): the parameters for these MPC candidates are included at the bottom of Table 7. We also note that the MPC candidates found in the GC also tend to be smaller and more massive than those found in the disc, but the environmental conditions found in the GC region are the most extreme in the Galaxy and finding differences between these two regions is perhaps not all that surprising.

## 8 EVOLUTION OF CLUMPS

The analysis presented thus far has revealed trends for increasing temperatures and luminosities with the evolutionary stage of the embedded stars as they move towards the main sequence. These two parameters can be attributed to the result of feedback from the forming protocluster on its natal clump. We can investigate the feedback by looking at changes in the linewidths of molecular transitions.

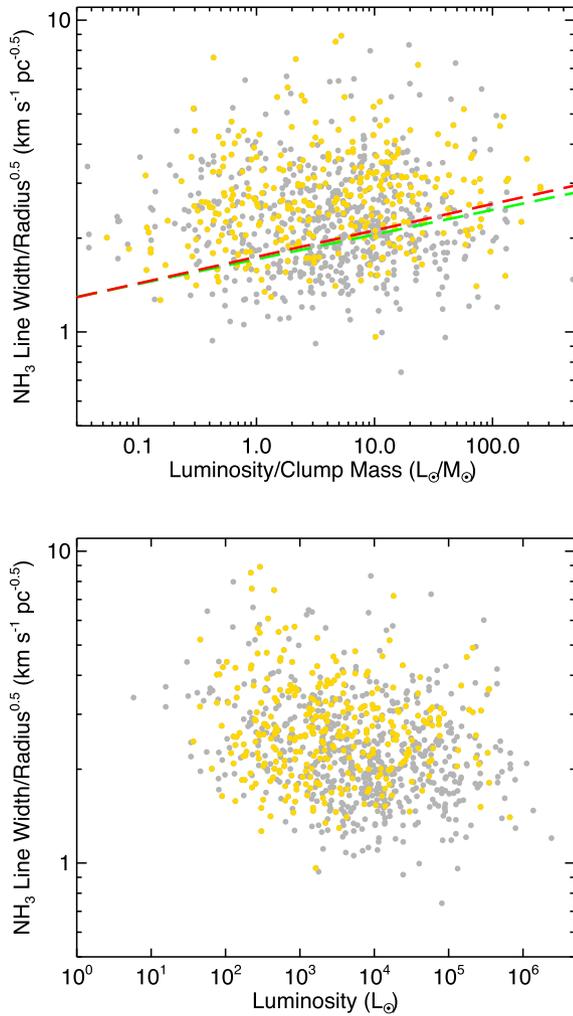
We show the correlation between the  $\text{NH}_3$  linewidth normalized by the square root of the clump radius and  $L_{\text{bol}}/M_{\text{clump}}$  ratio in the upper panel of Fig. 31. The partial-Spearman test reveals a moderate correlation between these parameters ( $r_{\text{AB,C}} = 0.12$ ,  $p$ -value = 0.0002), suggesting that the feedback from the evolving protocluster is resulting in an increase in the linewidth over time, although this increase is rather modest. The power-law fits to the full and distance-limited samples reveal slopes of 0.085 and 0.079, respectively. The increase in linewidth may not be due to feedback from the embedded protocluster as there may also be a contribution from globally infalling material from the extended envelope. We find no significant correlation between the normalized linewidth and bolometric luminosity (Fig. 31, lower panel), the lack of which, suggest that radiative feedback has little impact on the clump and that the increase in the normalized linewidth with evolution may be due to mechanical energy injected into the clump and/or changes in the global infall over the evolutionary time-scale.

The virial parameter measures the balance between gravity and the internal energy that can support the clump against gravitational collapse. The virial parameter is defined as:

$$\alpha_{\text{vir}} = \frac{5\sigma_v^2 R_{\text{eff}}}{GM_{\text{clump}}}, \quad (4)$$

where  $R_{\text{eff}}$  is the effective radius of the clumps,  $\sigma_v$  is the velocity dispersion of the ammonia (1,1) inversion transition and  $G$  is the gravitational constant.

We show the virial parameter as a function of clump mass in the upper panel of Fig. 32. A value of less than 2 is generally taken as indicating that a particular clump is unstable to gravity and is likely to be globally collapsing in the absence of significant magnetic support. It is clear from this plot that the most massive clumps are less gravitationally stable, which probably explains their rapid evolution and short pre-stellar lifetimes. For the distance-limited sample, we find  $\alpha_{\text{vir}} \propto M^{-0.4}$ ; this implies  $\sigma^2 R \propto M^{0.6}$ . In the upper panel of Fig. 31, we see a correlation between increasing linewidth and evolution and naively we might expect the virial mass to increase. It is not possible, however, to separate the turbulent component from the contribution from infall and outflow motions that are essentially random and do not contribute to the stability of the clump, and so we cannot say with any certainty how this will

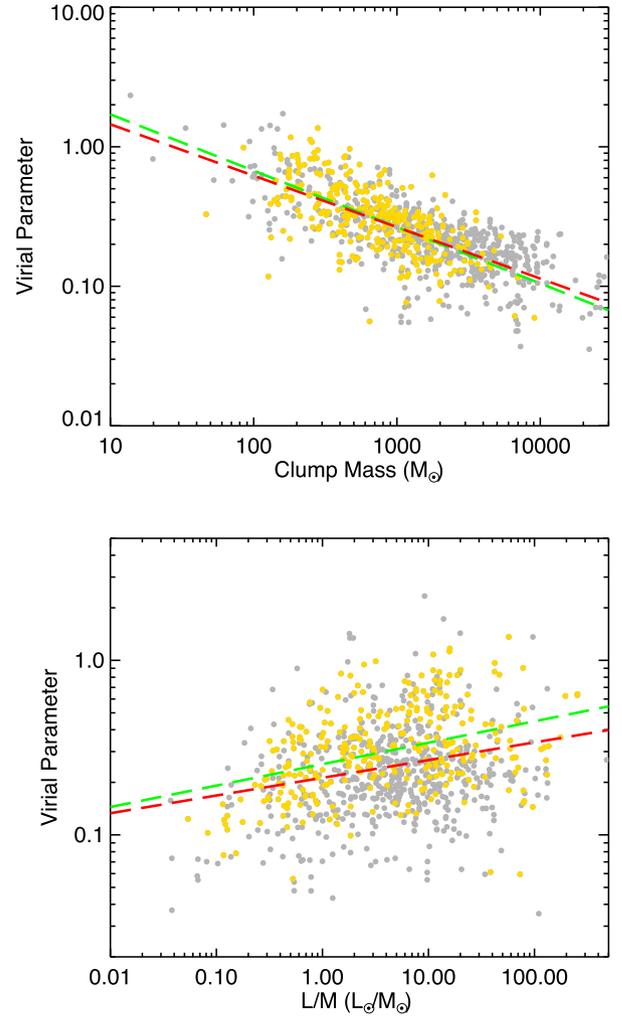


**Figure 31.**  $\text{NH}_3$  (1,1) linewidth normalized by the square root of the clump radius as a function of the  $L_{\text{bol}}/M_{\text{clump}}$  ratio (upper panel) and as a function of the bolometric luminosity (lower panel). The grey circles show the distributions of all clumps, while the yellow shows the distribution of the distance-limited sample. The distribution in the upper panel is weakly correlated, but the distribution in the lower panel is uncorrelated. The long-dashed red and green lines shown in the upper panel shows the result of a linear power-law fit to the whole sample of associated clumps and to the distance-limited sample, respectively.

affect the virial mass. We can, however, calculate how the virial parameter would be affected if all of the increase in linewidth did contribute to stabilizing the clump.

The lower panel of Fig. 32 shows a positive correlation between virial parameter and the  $L_{\text{bol}}/M_{\text{clump}}$  ratio where the  $\alpha_{\text{vir}}$  increase by a factor of 2–3. The virial mass is  $M_{\text{vir}} \propto R\Delta v^2$ , and so the increase in linewidth will result in an increase in the virial mass by a factor of 4–9, i.e. almost an order of magnitude. However, even assuming the feedback leads to a direct increase in the virial parameter, which is unlikely to be the case, the clumps will still be unstable to gravitational collapse (even the more evolved clumps have  $\alpha_{\text{vir}}$  significantly below 2) and it seems improbable that feedback could be driving an expansion of the clumps as they evolve.

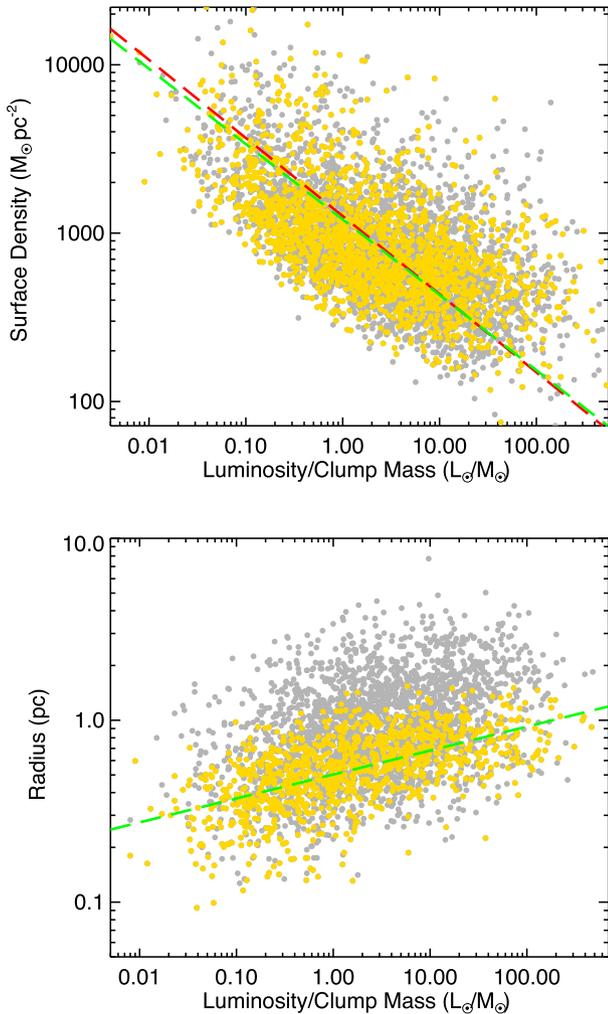
Elia et al. (2017) have also investigated how the surface density changes as a function of evolution and found no evidence to support this; however, their analysis focused on more compact objects than are considered here. In fact, their surface density versus  $L_{\text{bol}}/M_{\text{clump}}$



**Figure 32.** Virial parameter as a function of clump mass and  $L_{\text{bol}}/M_{\text{clump}}$  ratio (upper and lower panels, respectively). The long-dashed red line shows the result of a linear power-law fit to the whole sample of associated clumps, while the green line shows the fit to the distance-limited sample.

ratio plot appears very similar to our peak column density versus  $L_{\text{bol}}/M_{\text{clump}}$  ratio plot (cf. the middle left panels of their fig. 22 and the lower right panel of Fig. 22), and so when we look at similar size scales, we find similar behaviour for these parameters for both data sets.

We have also seen trends for increasing source size and decreasing mass surface density as a function of evolutionary stage (i.e. Figs 10 and 15, respectively) and so perhaps the feedback from the protocluster or accretion on to the cluster itself are having a direct impact on the structure of the clumps. We present a plot of the mass surface density as a function of the  $L_{\text{bol}}/M_{\text{clump}}$  ratio in the upper panel of Fig. 33: this plot would seem to confirm the trend for a significant decrease in the mass surface density as a function of evolution. The change in mass surface density covers 2 orders of magnitude and so the fraction of initial clump masses that would need to be depleted over the star formation time-scale is  $\sim 99$  per cent. Although this is a larger fraction than perhaps we would expect, it is at least feasible given that eventually all of a clump’s mass needs to be dissipated. However, if the clump mass was being so dramatically depleted we would expect to find it and a clump’s peak column density to be inversely correlated with the



**Figure 33.** Upper panel: the mass surface density as a function of the  $L_{\text{bol}}/M_{\text{clump}}$  ratio. Lower panel: clump radius as a function of  $L_{\text{bol}}/M_{\text{clump}}$  ratio. The grey circles show the distributions of all clumps, while the yellow shows the distribution of the distance-limited sample. The dashed red and green lines are linear fit to the log–log values of the full and distance-limited samples, respectively.

$L_{\text{bol}}/M_{\text{clump}}$  ratio, which is not the case (see lower right panel of Fig. 22). Furthermore, if clumps were in that main dispersal phase,  $T$  should be correlated with  $M$  and we would expect the range of  $T$  to be much larger and the upper left region of the  $L$ – $M$  plot would be populated, neither of which is observed.

If the clump mass is not decreasing significantly as a function of evolution then the only other way to produce a decrease in the mass surface density is to increase the radius. There are two questions that need to be addressed: (1) is the increase in radius sufficient to explain the decrease in the mass surface density and (2) what is driving the increase in clump size? To address the first, we plot the clump radius as a function of evolutionary stage in the lower panel of Fig. 33, and the increase in size is approximately a factor of 10 and so this can explain the decrease in mass surface density.

A similar set of trends have been recently reported from a study of a large sample ATLASGAL CSC sources by the MALT90 team (Contreras et al. 2017). They found that as the temperature increased, the volume density decreased and the virial parameter increased, and concluded that the clumps had started to expand, increasing their size and decreasing their volume density. However,

as previously mentioned, even if the virial parameter increases as the star formation evolves, the clumps would still be gravitationally unstable and so it is unlikely that the feedback processes could drive an expansion of the clumps (virial parameters significantly lower than 2).

An alternative explanation is that, as the temperature of the clump increases, so too does the column density sensitivity of the ATLASGAL survey, and as a result, we are simply more sensitive to the extended envelopes of the evolved clumps due to their higher temperatures than we are for the less-evolved clumps. We have tested this hypothesis using RATRAN to investigate the expected change in source size assuming a similar flux sensitivity of  $200 \text{ mJy beam}^{-1}$ , a distance of 4 kpc and a radial density profile of  $n \propto R^{-1.5}$  for the same clump over a temperature range of 12–45 K; this results in a change in the observed size of the clumps of a factor of 4.2 and so this provides an explanation for approximately half of the difference in size between the quiescent and MSF clumps, and if we were to also consider the impact of infalling material this would be larger still. Although both explanations are plausible, we think the scheme of clump expansion at the same time as the increase in star formation is significantly less likely than the observed increase in radius being the result of an observational bias resulting from a temperature increase.

We have found that the feedback from the evolving protocluster is warming the clumps and increasing the molecular linewidths. Although the material towards the centre of the clump is depleted as the embedded protocluster evolves through a combination of accretion and outflows, this is unlikely to have a significant impact on the mass of the clumps given that there is no correlation with evolution for either the clump mass or peak column density. The lack of any observed correlation between mass and evolution suggests, additionally, that global infall has little impact on the mass of the clumps (as calculated in Section 5.5): if there was an impact, we would have expected to find a positive correlation between the two parameters. It is of course possible that the amount of material gained through infall is almost exactly matched by material accreted on the protocluster and removed by outflows, although the existence of such a delicate balance over such large ranges in clump mass and evolution is difficult to imagine.

## 9 PROPERTIES OF MOST ACTIVE COMPLEXES IN THE GALAXY

We present the derived physical parameters for the 30 most massive clusters identified in Table 8. A comparison of the mass and luminosity contained in these 30 regions with those of the whole sample ( $\sim 1 \times 10^7 M_{\odot}$  and  $\sim 1.3 \times 10^8 L_{\odot}$ , respectively) shows that they collectively contribute 36 per cent of all of the mass and 52 per cent of all of the luminosity, but only 16.4 per cent of the total population of clumps. This rather limited number of regions is therefore responsible for a majority share of the current star formation rate in the Galaxy.

This sample includes all of the previously well-known star-forming complexes such as W31, W33, W43, W49, W51, G305 and G333. All clusters contribute between  $\sim 0.5$  and 3 per cent of the total mass and no cluster stands out in this regard. A comparison of luminosities is similar, although clusters W49 and W51 do stand out from the others. These two clusters together contribute  $\sim 20$  per cent of the total luminosity found in the inner part of the Galactic disc.

We previously used the  $L_{\text{bol}}/M_{\text{clump}}$  ratio as a means to evaluate the current evolutionary state of the embedded star formation, where

**Table 8.** Derived cluster parameters for all 776 clusters identified, ordered by their total dense gas mass. We include the standard deviations for the longitude, latitude, velocity and temperature.

Cluster name	Literature name	Number of members	$\ell$ (°)	$b$ (°)	$v_{\text{lsr}}$ (km s <sup>-1</sup> )	Distance (kpc)	$T$ (K)	$\log(L_{\text{bol}})$ ( $L_{\odot}$ )	$L_{\text{bol}}$ (per cent)	$\log(M_{\text{clump}})$ ( $M_{\odot}$ )	$M_{\text{clump}}$ (per cent)	$L_{\text{bol}}/M_{\text{clump}}$ ( $L_{\odot}/M_{\odot}$ )
G030.650–00.015		245	30.651 ± 0.472	−0.014 ± 0.170	97.44 ± 8.90	5.2	9.5 ± 6.09	6.608	3.12	5.530	3.34	12.0
G043.141–00.018	W49	14	43.141 ± 0.053	−0.018 ± 0.035	9.47 ± 3.96	11.1	21.2 ± 4.58	7.181	11.69	5.354	2.23	67.1
G024.319+00.191		61	24.313 ± 0.165	0.189 ± 0.091	112.77 ± 4.66	7.8	10.9 ± 5.30	5.750	0.43	5.349	2.20	2.5
G049.261–00.318	W51	88	49.256 ± 0.220	−0.316 ± 0.099	60.92 ± 6.80	5.3	12.9 ± 4.98	7.045	8.56	5.320	2.06	53.2
G350.219+00.081		31	350.222 ± 0.081	0.082 ± 0.048	−64.61 ± 4.37	10.5	13.2 ± 4.55	6.477	2.31	5.295	1.95	15.2
G337.228–00.065		30	337.228 ± 0.071	−0.065 ± 0.065	−70.13 ± 3.40	11.0	11.5 ± 5.01	6.407	1.97	5.262	1.80	14.0
G331.394–00.125		99	331.394 ± 0.243	−0.125 ± 0.124	−90.03 ± 5.65	5.3	11.5 ± 6.05	6.562	2.81	5.195	1.54	23.3
G333.125–00.348	G333	240	333.123 ± 0.425	−0.346 ± 0.233	−50.75 ± 4.11	3.6	10.4 ± 5.70	6.524	2.57	5.192	1.53	21.5
G347.650+00.210		22	347.650 ± 0.064	0.210 ± 0.049	−86.91 ± 9.73	9.8	10.3 ± 7.03	6.428	2.07	5.150	1.39	19.0
G349.168+00.118		32	349.168 ± 0.188	0.120 ± 0.032	−68.53 ± 5.38	10.5	13.6 ± 3.48	6.108	0.99	5.129	1.33	9.5
G349.130–00.003		6	349.130 ± 0.033	−0.003 ± 0.021	15.25 ± 1.34	19.7	18.6 ± 2.83	5.933	0.66	5.089	1.21	7.0
G337.017–00.001		37	337.012 ± 0.189	−0.001 ± 0.077	−116.38 ± 6.66	7.7	15.5 ± 5.10	5.889	0.60	5.043	1.09	7.0
G025.656–00.087		27	25.656 ± 0.168	−0.087 ± 0.080	93.50 ± 5.91	10.2	10.3 ± 5.59	6.166	1.13	5.036	1.07	13.5
G337.677–00.048		9	337.677 ± 0.026	−0.037 ± 0.033	−48.53 ± 1.57	12.1	15.4 ± 3.18	5.773	0.46	5.034	1.07	5.5
Bania Clump 1		12	354.759 ± 0.067	0.335 ± 0.056	87.92 ± 10.35	8.4	12.4 ± 5.98	5.376	0.18	5.030	1.06	2.2
G342.533+00.179		9	342.533 ± 0.054	0.179 ± 0.008	−41.91 ± 0.49	12.5	11.5 ± 4.06	4.881	0.06	5.010	1.01	0.7
G020.739–00.136		22	20.739 ± 0.049	−0.136 ± 0.105	57.68 ± 2.67	11.7	11.8 ± 5.37	5.784	0.47	4.982	0.95	6.3
G353.511–00.053		21	353.511 ± 0.079	−0.053 ± 0.038	−53.86 ± 6.00	10.2	18.3 ± 5.43	5.916	0.64	4.972	0.92	8.8
G023.375–00.101		52	23.374 ± 0.111	−0.098 ± 0.114	93.09 ± 10.98	5.9	12.0 ± 5.68	6.026	0.82	4.970	0.92	11.4
G351.598+00.189		18	351.598 ± 0.076	0.189 ± 0.033	−42.02 ± 2.30	11.5	14.8 ± 6.78	6.745	4.28	4.962	0.90	60.6
G305.453+00.065	G305	98	305.459 ± 0.269	0.062 ± 0.198	−37.34 ± 3.82	3.8	12.3 ± 6.40	6.296	1.52	4.916	0.81	24.0
G338.586+00.043		11	338.586 ± 0.024	0.043 ± 0.050	−23.67 ± 2.20	13.6	13.1 ± 3.63	5.614	0.32	4.870	0.73	5.6
G354.479+00.087		4	354.480 ± 0.018	0.087 ± 0.006	16.80 ± 1.49	23.3	20.2 ± 4.05	5.822	0.51	4.869	0.73	9.0
G320.403+00.131		10	320.403 ± 0.024	0.131 ± 0.027	−5.11 ± 1.52	12.5	12.1 ± 5.74	5.728	0.41	4.860	0.71	7.4
G012.203–00.117	W33	4	12.203 ± 0.009	−0.117 ± 0.010	26.12 ± 1.29	13.4	22.8 ± 3.06	6.175	1.15	4.828	0.66	22.2
G018.929–00.343		57	18.934 ± 0.121	−0.349 ± 0.205	63.94 ± 3.29	5.0	10.7 ± 4.84	5.592	0.30	4.812	0.64	6.0
G048.651+00.131		13	48.651 ± 0.084	0.131 ± 0.097	13.85 ± 4.29	10.8	12.6 ± 5.22	5.869	0.57	4.801	0.62	11.7
G019.649–00.239		6	19.649 ± 0.039	−0.239 ± 0.032	40.85 ± 2.21	12.6	18.9 ± 3.52	6.117	1.01	4.789	0.61	21.3
G035.553+00.008		11	35.553 ± 0.040	0.008 ± 0.063	52.08 ± 3.14	10.4	15.4 ± 4.52	5.653	0.35	4.789	0.61	7.3
G033.527–00.011		25	33.521 ± 0.195	−0.010 ± 0.028	103.25 ± 2.21	6.5	10.4 ± 5.58	5.175	0.12	4.778	0.59	2.5

Notes: A more complete table that includes parameters for all 776 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>.

higher values are associated with more-evolved objects. Here, we estimate the average  $L_{\text{bol}}/M_{\text{clump}}$  ratio for the clusters as a whole to obtain a measure of their star formation activity: these values are given in the last column of Table 8. The average  $L_{\text{bol}}/M_{\text{clump}}$  ratio for the sample is  $15.9 L_{\odot}/M_{\odot}$  with a standard deviation of  $16.3 L_{\odot}/M_{\odot}$ . While most clusters are within one standard deviation of the mean, W51 and G351 are both nearly  $3\sigma$  from the mean and W49 is over  $3\sigma$  away from the mean. These three clusters have  $L_{\text{bol}}/M_{\text{clump}}$  ratios at least twice as large as the next highest cluster, and are the most-evolved star formation regions (have the highest instantaneous SFE) in the Galaxy. These values put them in the wings of the  $L_{\text{bol}}/M_{\text{clump}}$  distribution: they are therefore rather extreme compared to all of the other regions, and are probably the best examples of ‘mini-starbursts’ we have in the Galaxy. We note that as pointed out by Eden et al. (2015), these sources are not outliers from the general (log)normal  $L_{\text{bol}}/M_{\text{clump}}$  distribution: while certainly extreme, they are not abnormal, and sources like these are expected to arise statistically when the sample is large enough.

## 10 SUMMARY AND CONCLUSIONS

The ATLASGAL survey has identified  $\sim 10\,000$  dense clumps located across the inner mid-plane of the Galaxy. We concentrate in this paper on a subsample of 8002 clumps located within the disc and away from the GC region. We have avoided the innermost  $5^{\circ}$  of the Galactic plane due to problems with source confusion and difficulty obtaining reliable distances. We have used a combination of archival molecular line data sets, line surveys reported in the literature, and dedicated follow-up observations of ATLASGAL sources (some of which are reported here) to determine a radial velocity ( $v_{\text{lsr}}$ ) for 7807 clumps (corresponding to  $\sim 98$  per cent of the sample).

These velocities have been used in conjunction with a Galactic rotation model to estimate kinematic distances, and analysis of H I spectra has been conducted to resolve the distance ambiguities that affect all sources located inside the Solar circle. These kinematic distance solutions have been improved using maser parallax and spectroscopic distances reported in the literature where these quantities have been available. We use a friends-of-friends algorithm to identify clusters of clumps that are coherent in position and velocity (i.e.  $\ell$ ,  $b$  and  $v$ ). This identifies groups of sources that are likely to be part of larger scale structures such as GMCs and giant molecular filaments (GMFs). This analysis has resulted in the identification of 776 clusters, and has recovered all of the most active and well-known star-forming complexes in the Galaxy.

The identifications of clumps as parts of larger complexes allows the properties of star-forming complexes to be compared to each other, and the role of the spiral arms and other environmental conditions to be explored. This can additionally reduce the inherent uncertainties associated with resolving the KDA. We have been able to assign a distance to 7770 clumps ( $\sim 97$  per cent of the whole sample), but many of the remaining sources are weak and often diffuse.

We have performed aperture photometry on mid-IR to submillimetre survey data, and used these flux measurements to fit the SED of the clumps. These provide a measure of the dust temperatures and bolometric fluxes. This analysis has also been used to separate the clumps into four evolutionary types: quiescent, protostellar, young stellar objects and MSF. The distances, temperatures and bolometric fluxes are used to derive the physical parameters for the clumps (luminosities, sizes, mass and column densities), and comparisons have been made between the different evolutionary stage subsamples to identify trends that may provide some insight to the star formation process and the evolution of the clumps. These

comparisons have revealed that dust temperatures, bolometric luminosities and  $L_{\text{bol}}/M_{\text{clump}}$  ratios increase with advanced evolutionary stage.

Our main findings are as follows:

(i) A comparison of the mass–radius distribution to empirically determined thresholds for the efficient formation of massive stars finds that the vast majority of all clumps are capable of forming massive stars. Furthermore, the vast majority of clumps appear to be unstable against gravity and can therefore be considered to be in a pre-stellar state as opposed to starless.

(ii) We find that there is a very strong positive correlation between column density and the fraction of clumps associated with massive star formation; this saturates at  $\log(N) \sim 23 \text{ cm}^{-2}$ , above which the proportion is 100 per cent. We do not find a correlation between column density and evolution, and so we can exclude the possibility that the column density increases over time and conclude that the reason we do not find any quiescent, protostellar or YSO-forming clumps with similar column densities is because the evolution time-scale is inversely correlated to the column density. A consequence of this is that the highest column density clumps evolve so rapidly and are so rare that the early stages are simply not observed.

(iii) We find that the clump mass and evolutionary stage are uncorrelated: the clump mass is not significantly altered by infalling mass from the larger scale environment over its lifetime. Clumps are widely thought to form over long periods of time from infalling material either funnelled along filaments or from global collapse from larger scales and although there have been some studies that have found evidence to support this hypothesis, it does not seem to apply to the majority of clumps.

(iv) The fractions of clumps in the quiescent and protostellar stages are  $\sim 10$  and  $\sim 20$  per cent for clumps up to  $4 \times 10^3 M_{\odot}$  but drop to  $\sim 5$  per cent and 10 per cent, respectively, for more massive clumps. The statistical lifetime of the early quiescent and protostellar stages are therefore relatively brief and are negatively correlated with the mass of the clump. As the most massive clumps are associated with the most massive stars (which have the shortest star formation time-scales), the actual lifetimes for the most massive clumps will be very short indeed.

(v) The short quiescent and protostellar statistical lifetimes are consistent with the hypothesis that the clumps form rapidly with most of their mass in place, and are initially very unstable to gravity so that star formation begins almost immediately. The mass infall rate is relatively modest, as is the mass accreted on to the protocluster over the star formation time-scales, resulting in the clump mass being independent of evolution. The key conclusion here is that the formation of the clump and ensuing star formation must be very rapid, and once it has begun, is little influenced by the larger scale environment.

(vi) We have compared the distribution of clumps and clusters to the large-scale features of the Galaxy and find a good correlation between the sources and the spiral arms ( $\sim 90$  per cent found within  $10 \text{ km s}^{-1}$  of a spiral-arm loci). There is also a good correlation between the expected peaks in the Galactocentric distance distribution of the spiral arms and the MSF clumps. We find that all of the clumps are tightly correlated with the mid-plane of the Galaxy with a scaleheight of  $\sim 26 \text{ pc}$ : we find no difference between the different subsamples.

(vii) We find little evidence for variations in the  $L_{\text{bol}}/M_{\text{clump}}$  ratio or mean clump temperature as a function of Galactocentric distance ( $\sim 10 L_{\odot}/M_{\odot}$  and  $\sim 20 \text{ K}$ , respectively) within the Solar circle when averaged over kiloparsec scales. The mean temperature

increases slightly outside of the Solar circle, possibly due to the lower density of material and decreased shielding from the interstellar radiation field, but the  $L_{\text{bol}}/M_{\text{clump}}$  ratio appears to be relatively unchanged. There are more significant variations on smaller size scales of clusters and clumps (several tens of parsecs and parsec scales, respectively).

(viii) We do not find any evidence for any enhancements in the  $L_{\text{bol}}/M_{\text{clump}}$  ratio at the expected galactocentric distances of the spiral arms. From this, we conclude that although the spiral arms play a role in concentrating the molecular material into large-scale structures they do not play a significant role in the star formation process. This finding is consistent with other recent studies.

(ix) We have combined the properties of the clumps grouped into clusters to estimate the properties of these larger scale structures. We find that the 30 most massive of these structures consists of only 16 per cent of all of the clumps, but contain 36 per cent of the mass and 52 per cent of the total bolometric luminosity. These clusters therefore are responsible for a significant fraction of the Galactic star formation rate and energy budget.

(x) The majority of clusters identified have  $L_{\text{bol}}/M_{\text{clump}}$  ratios  $< 25 L_{\odot}/M_{\odot}$ , however, three clusters have significantly higher values ( $> 50 L_{\odot}/M_{\odot}$ ) and so are much more active and evolved regions (these are W49, W51 and G351.598+00.189). Combined, these three clusters produce almost 25 per cent of the total luminosity produced by the ATLASGAL sources (11.7 per cent for W49, 8.56 per cent for W51 and 4.3 per cent for G351.598+00.189). The most extreme of these three regions is W49, the most luminous complex in the Galaxy. W49 has the highest  $L_{\text{bol}}/M_{\text{clump}}$  ratio and is one of only three objects found in the disc that satisfies the criteria to be considered a MPC candidate.

This paper is the fourth in a series that is focusing on exploiting the ATLASGAL survey to provide a catalogue of properties for all dense clumps located within the inner Galaxy. This paper completes this work and in subsequent papers we will use this catalogue to further investigate the role of the spiral arms and Galactic environment and investigate the difference between clump properties and star formation between the inner and outer Galaxy.

## ACKNOWLEDGEMENTS

We thank the referee for a timely report and helpful comments and suggestions that have improved the clarity of this paper. The ATLASGAL project is a collaboration between the Max–Planck–Gesellschaft, the European Southern Observatory and the Universidad de Chile. This research has made use of the SIMBAD data base operated at CDS, Strasbourg, France. This work was partially funded by the Collaborative Research Council 956, subproject A6, funded by the Deutsche Forschungsgemeinschaft. This paper made use of information from the Red MSX Source survey data base at [http://rms.leeds.ac.uk/cgi-bin/public/RMS\\_DATABASE.cgi](http://rms.leeds.ac.uk/cgi-bin/public/RMS_DATABASE.cgi) which was constructed with support from the Science and Technology Facilities Council of the UK.

## REFERENCES

- Aguirre J. E. et al., 2011, *ApJS*, 192, 4  
 Anderson L. D., Bania T. M., 2009, *ApJ*, 690, 706  
 Anderson L. D., Bania T. M., Jackson J. M., Clemens D. P., Heyer M., Simon R., Shah R. Y., Rathborne J. M., 2009, *ApJS*, 181, 255  
 Araya E., Hofner P., Churchwell E., Kurtz S., 2002, *ApJS*, 138, 63

- Barnes P. J., Muller E., Indermuehle B., O’Dougherty S. N., Lowe V., Cunningham M., Hernandez A. K., Fuller G. A., 2015, *ApJ*, 812, 6
- Battisti A. J., Heyer M. H., 2014, *ApJ*, 780, 173
- Benjamin R. A. et al., 2003, *PASP*, 115, 953
- Beuther H., Henning T., Linz H., Krause O., Nielbock M., Steinacker J., 2010, *A&A*, 518, L78
- Beuther H. et al., 2012, *ApJ*, 747, 43
- Braiding C. et al., 2015, *Publ. Astron. Soc. Aust.*, 32, e020
- Brand J., Blitz L., 1993, *A&A*, 275, 67
- Bressert E., Ginsburg A., Bally J., Battersby C., Longmore S., Testi L., 2012, *ApJ*, 758, L28
- Bronfman L., Nyman L.-A., May J., 1996, *A&AS*, 115, 81
- Bronfman L., Casassus S., May J., Nyman L.-A., 2000, *A&A*, 358, 521
- Burton M. G. et al., 2013, *Publ. Astron. Soc. Aust.*, 30, 44
- Busfield A. L., Purcell C. R., Hoare M. G., Lumsden S. L., Moore T. J. T., Oudmaijer R. D., 2006, *MNRAS*, 366, 1096
- Caswell J. L., Haynes R. F., 1987, *A&A*, 171, 261
- Clemens D. P., 1985, *ApJ*, 295, 422
- Collins C. A., Mann R. G., 1998, *MNRAS*, 297, 128
- Colombo D. et al., 2014, *ApJ*, 784, 4
- Compiègne M., 2010, in Kothes R., Landecker T. L., Willis A. G., eds, *ASP Conf. Ser.*, Vol. 438, *The Dynamic Interstellar Medium: A Celebration of the Canadian Galactic Plane Survey*. Astron. Soc. Pac., San Francisco, p. 55
- Compiègne M., Flagey N., Noriega-Crespo A., Martin P. G., Bernard J.-P., Paladini R., Molinari S., 2010, *ApJ*, 724, L44
- Contreras Y. et al., 2013, *A&A*, 549, A45
- Contreras Y., Rathborne J. M., Guzman A., Jackson J., Whitaker S., Sanhueza P., Foster J., 2017, *MNRAS*, 466, 340
- Cordes J. M., 2004, in Clemens D., Shah R., Brainerd T., eds, *ASP Conf. Ser.*, Vol. 317, *Milky Way Surveys: The Structure and Evolution of our Galaxy*. Astron. Soc. Pac., San Francisco, p. 211
- Csengeri T. et al., 2014, *A&A*, 565, A75
- Csengeri T. et al., 2016a, *A&A*, 586, A149
- Csengeri T. et al., 2016b, *A&A*, 585, A104
- Csengeri T. et al., 2017, *A&A*, 600, L10
- Dame T. M., Hartmann D., Thaddeus P., 2001, *ApJ*, 547, 792
- Davies B., Hoare M. G., Lumsden S. L., Hosokawa T., Oudmaijer R. D., Urquhart J. S., Mottram J. C., Stead J., 2011, *MNRAS*, 416, 972
- Dempsey J. T., Thomas H. S., Currie M. J., 2013, *ApJS*, 209, 8
- Dobbs C. L., Bonnell I. A., Pringle J. E., 2006, *MNRAS*, 371, 1663
- Downes D., Wilson T. L., Bieging J., Wink J., 1980, *A&AS*, 40, 379
- Dunham M. K., Robitaille T. P., Evans N. J., II, Schlingman W. M., Cyganowski C. J., Urquhart J., 2011a, *ApJ*, 731, 90
- Dunham M. K., Rosolowsky E., Evans N. J., II, Cyganowski C., Urquhart J. S., 2011b, *ApJ*, 741, 110
- Eden D. J., Moore T. J. T., Plume R., Morgan L. K., 2012, *MNRAS*, 422, 3178
- Eden D. J., Moore T. J. T., Morgan L. K., Thompson M. A., Urquhart J. S., 2013, *MNRAS*, 431, 1587
- Eden D. J., Moore T. J. T., Urquhart J. S., Elia D., Plume R., Rigby A. J., Thompson M. A., 2015, *MNRAS*, 452, 289
- Egan M. P. et al., 2003, *VizieR Online Data Catalog*, 5114, 0
- Elia D. et al., 2017, *MNRAS*, 471, 100
- Fish V. L., Reid M. J., Wilner D. J., Churchwell E., 2003, *ApJ*, 587, 701
- Fukui Y. et al., 2009, *ApJ*, 705, 144
- Giannetti A. et al., 2013, *A&A*, 556, A16
- Giannetti A., Wyrowski F., Leurini S., Urquhart J., Csengeri T., Menten K. M., Bronfman L., van der Tak F. F. S., 2015, *A&A*, 580, L7
- Giannetti A., Leurini S., Wyrowski F., Urquhart J., Csengeri T., Menten K. M., König C., Güsten R., 2017, *A&A*, 603, A33
- Ginsburg A., Bressert E., Bally J., Battersby C., 2012, *ApJ*, 758, L29
- Green J. A., McClure-Griffiths N. M., 2011, *MNRAS*, 417, 2500
- Griffin M. J. et al., 2010, *A&A*, 518, L3
- Güsten R., Nyman L. Å., Schilke P., Menten K., Cesarsky C., Booth R., 2006, *A&A*, 454, L13
- Guzmán A. E., Sanhueza P., Contreras Y., Smith H. A., Jackson J. M., Hoq S., Rathborne J. M., 2015, *ApJ*, 815, 130
- Heiderman A., Evans N. J., II, Allen L. E., Huard T., Heyer M., 2010, *ApJ*, 723, 1019
- Heyer M., Dame T. M., 2015, *ARA&A*, 53, 583
- Hildebrand R. H., 1983, *QJRAS*, 24, 267
- Hill T., Longmore S. N., Pinte C., Cunningham M. R., Burton M. G., Minier V., 2010, *MNRAS*, 402, 2682
- Hoare M. G. et al., 2012, *PASP*, 124, 939
- Hughes A. et al., 2010, *MNRAS*, 406, 2065
- Immer K., Schuller F., Omont A., Menten K. M., 2012, *A&A*, 537, A121
- Jackson J. M. et al., 2006, *ApJS*, 163, 145
- Jackson J. M., Finn S. C., Rathborne J. M., Chambers E. T., Simon R., 2008, *ApJ*, 680, 349
- Jackson J. M. et al., 2013, *Publ. Astron. Soc. Aust.*, 30, 57
- Kauffmann J., Pillai T., 2010, *ApJ*, 723, L7
- Kauffmann J., Pillai T., Shetty R., Myers P. C., Goodman A. A., 2010, *ApJ*, 716, 433
- Kim W.-J., Wyrowski F., Urquhart J. S., Menten K. M., Csengeri T., 2017, *A&A*, 602, A37
- Klein B., Hochgürtel S., Krämer I., Bell A., Meyer K., Güsten R., 2012, *A&A*, 542, L3
- Kolpak M. A., Jackson J. M., Bania T. M., Clemens D. P., Dickey J. M., 2003, *ApJ*, 582, 756
- König C. et al., 2017, *A&A*, 599, A139
- Krishnan V., Ellingsen S. P., Reid M. J., Bignall H. E., McCallum J., Phillips C. J., Reynolds C., Stevens J., 2017, *MNRAS*, 465, 1095
- Lada C. J., Lada E. A., 2003, *ARA&A*, 41, 57
- Lada C. J., Lombardi M., Alves J. F., 2010, *ApJ*, 724, 687
- Lee Y., Stark A. A., Kim H.-G., Moon D.-S., 2001, *ApJS*, 136, 137
- Li G.-X., Wyrowski F., Menten K., Belloche A., 2013, *A&A*, 559, A34
- Li G.-X., Urquhart J. S., Leurini S., Csengeri T., Wyrowski F., Menten K. M., Schuller F., 2016, *A&A*, 591, A5
- Lis D. C., Li Y., Dowell C. D., Menten K. M., 1999, in Cox P., Kessler M., eds, *ESA SP-427, The Universe as Seen by ISO*. ESA, Noordwijk, p. 627
- Longmore S. N. et al., 2012, *ApJ*, 746, 117
- Longmore S. N. et al., 2017, *MNRAS*, 470, 1462
- Loup C., Forveille T., Omont A., Paul J. F., 1993, *A&AS*, 99, 291
- Lumsden S. L., Hoare M. G., Urquhart J. S., Oudmaijer R. D., Davies B., Mottram J. C., Cooper H. D. B., Moore T. J. T., 2013, *ApJS*, 208, 11
- McClure-Griffiths N. M., Dickey J. M., 2007, *ApJ*, 671, 427
- McClure-Griffiths N. M., Dickey J. M., Gaensler B. M., Green A. J., Haverkorn M., Strasser S., 2005, *ApJS*, 158, 178
- Miville-Deschênes M.-A., Murray N., Lee E. J., 2017, *ApJ*, 834, 57
- Moisés A. P., Damineli A., Figuerêdo E., Blum R. D., Conti P. S., Barbosa C. L., 2011, *MNRAS*, 411, 705
- Molinari S., Pezzuto S., Cesaroni R., Brand J., Faustini F., Testi L., 2008, *A&A*, 481, 345
- Molinari S. et al., 2010, *PASP*, 122, 314
- Molinari S. et al., 2016, *A&A*, 591, A149
- Moore T. J. T., Urquhart J. S., Morgan L. K., Thompson M. A., 2012, *MNRAS*, 426, 701
- Moore T. J. T. et al., 2015, *MNRAS*, 453, 4264
- Motte F., Bontemps S., Schilke P., Schneider N., Menten K. M., Brogière D., 2007, *A&A*, 476, 1243
- Motte F., Bontemps S., Louvet F., 2017, preprint ([arXiv:1706.00118](https://arxiv.org/abs/1706.00118))
- Mottram J. C. et al., 2011, *ApJ*, 730, L33+
- Murray N., Rahman M., 2010, *ApJ*, 709, 424
- Nagayama T., Omodaka T., Handa T., Honma M., Kobayashi H., Kawaguchi N., Ueno Y., 2011, *PASJ*, 63, 719
- Nguyen Luong Q. et al., 2011, *A&A*, 535, A76
- Ossenkopf V., Henning T., 1994, *A&A*, 291, 943
- Pandian J. D., Momjian E., Goldsmith P. F., 2008, *A&A*, 486, 191
- Pandian J. D., Menten K. M., Goldsmith P. F., 2009, *ApJ*, 706, 1609
- Peretto N., Fuller G. A., 2009, *A&A*, 505, 405
- Poglitsch A. et al., 2010, *A&A*, 518, L2
- Portegies Zwart S. F., McMillan S. L. W., Gieles M., 2010, *ARA&A*, 48, 431

- Price S. D., Egan M. P., Carey S. J., Mizuno D. R., Kuchar T. A., 2001, *AJ*, 121, 2819
- Purcell C. R. et al., 2012, *MNRAS*, 426, 1972
- Ragan S. E., Moore T. J. T., Eden D. J., Hoare M. G., Elia D., Molinari S., 2016, *MNRAS*, 462, 3123
- Rathborne J. M., Jackson J. M., Simon R., 2006, *ApJ*, 641, 389
- Rathborne J. M., Johnson A. M., Jackson J. M., Shah R. Y., Simon R., 2009, *ApJS*, 182, 131
- Reed B. C., 2000, *AJ*, 120, 314
- Reid M. J. et al., 2014, *ApJ*, 783, 130
- Reid M. J., Dame T. M., Menten K. M., Brunthaler A., 2016, *ApJ*, 823, 77
- Rigby A. J. et al., 2016, *MNRAS*, 456, 2885
- Robitaille T. P., Whitney B. A., Indebetouw R., Wood K., 2007, *ApJS*, 169, 328
- Roman-Duval J., Jackson J. M., Heyer M., Johnson A., Rathborne J., Shah R., Simon R., 2009, *ApJ*, 699, 1153
- Sanna A., Reid M. J., Moscadelli L., Dame T. M., Menten K. M., Brunthaler A., Zheng X. W., Xu Y., 2009, *ApJ*, 706, 464
- Sanna A. et al., 2014, *ApJ*, 781, 108
- Saraceno P., Andre P., Ceccarelli C., Griffin M., Molinari S., 1996, *A&A*, 309, 827
- Sato M., Reid M. J., Brunthaler A., Menten K. M., 2010, *ApJ*, 720, 1055
- Schlingman W. M. et al., 2011, *ApJS*, 195, 14
- Schneider N., Bontemps S., Simon R., Jakob H., Motte F., Miller M., Kramer C., Stutzki J., 2006, *A&A*, 458, 855
- Schuller F., Menten K. M., Contreras Y., Wyrowski F., Schilke e. a., 2009, *A&A*, 504, 415
- Schuller F. et al., 2017, *A&A*, 601, A124
- Sewilo M., Watson C., Araya E., Churchwell E., Hofner P., Kurtz S., 2004, *ApJ*, 154, 553
- Shirley Y. L. et al., 2013, *ApJS*, 209, 2
- Stark A. A., Lee Y., 2006, *ApJ*, 641, L113
- Stead J. J., Hoare M. G., 2010, *MNRAS*, 407, 923
- Stil J. M. et al., 2006, *AJ*, 132, 1158
- Svoboda B. E. et al., 2016, *ApJ*, 822, 59
- Tackenberg J. et al., 2012, *A&A*, 540, A113
- Taylor J. H., Cordes J. M., 1993, *ApJ*, 411, 674
- Thompson M. A., White G. J., Morgan L. K., Miao J., Fridlund C. V. M., Hultgren-White M., 2004, *A&A*, 414, 1017
- Tohill N. F. H., White G. J., Matthews H. E., McCutcheon W. H., McCaughrean M. J., Kenworthy M. A., 2002, *ApJ*, 580, 285
- Traficante A., Fuller G. A., Billot N., Duarte-Cabral A., Merello M., Molinari S., Peretto N., Schisano E., 2017, *MNRAS*, 470, 3882
- Urquhart J. S., White G. J., Pilbratt G. L., Fridlund C. V. M., 2003, *A&A*, 409, 193
- Urquhart J. S. et al., 2007, *A&A*, 474, 891
- Urquhart J. S. et al., 2008, *A&A*, 487, 253
- Urquhart J. S. et al., 2011, *MNRAS*, 418, 1689
- Urquhart J. S. et al., 2012, *MNRAS*, 420, 1656
- Urquhart J. S. et al., 2013a, *MNRAS*, 431, 1752
- Urquhart J. S. et al., 2013b, *MNRAS*, 435, 400
- Urquhart J. S. et al., 2014a, *A&A*, 568, A41
- Urquhart J. S., Figura C. C., Moore T. J. T., Hoare M. G., Lumsden S. L., Mottram J. C., Thompson M. A., Oudmaijer R. D., 2014b, *MNRAS*, 437, 1791
- Urquhart J. S. et al., 2014c, *MNRAS*, 443, 1555
- Urquhart J. S. et al., 2015, *MNRAS*, 452, 4029
- Vassilev V. et al., 2008, *A&A*, 490, 1157
- Wang K., Testi L., Ginsburg A., Walmsley C. M., Molinari S., Schisano E., 2015, *MNRAS*, 450, 4043
- Watson C., Araya E., Sewilo M., Churchwell E., Hofner P., Kurtz S., 2003, *ApJ*, 587, 714
- Wienen M., Wyrowski F., Schuller F., Menten K. M., Walmsley C. M., Bronfman L., Motte F., 2012, *A&A*, 544, A146
- Wienen M. et al., 2015, *A&A*, 579, A91
- Wienen M., Wyrowski F., Menten K. M., Urquhart J. S., Walmsley C. M., Csengeri T., Koribalski B. S., Schuller F., 2017, preprint (arXiv:1708.07839)
- Wright E. L. et al., 2010, *AJ*, 140, 1868
- Wu Y. W. et al., 2014, *A&A*, 566, A17
- Wyrowski F. et al., 2016, *A&A*, 585, A149
- Xu Y., Reid M. J., Menten K. M., Brunthaler A., Zheng X. W., Moscadelli L., 2009, *ApJ*, 693, 413
- Xu Y., Moscadelli L., Reid M. J., Menten K. M., Zhang B., Zheng X. W., Brunthaler A., 2011, *ApJ*, 733, 25
- Yates M. G., Miller L., Peacock J. A., 1986, *MNRAS*, 221, 311
- Zhang B., Reid M. J., Menten K. M., Zheng X. W., Brunthaler A., Dame T. M., Xu Y., 2013, *ApJ*, 775, 79

## SUPPORTING INFORMATION

Supplementary data are available at the CDS via anonymous ftp to [cdsarc.u-strasbg.fr](http://cdsarc.u-strasbg.fr) (130.79.125.5) or via <http://cdsweb.u-strasbg.fr/cgi-bin/qcat?J/MNRAS/>.

**Table 2.** Summary of the kinematic distance analysis.

**Table 5.** Derived clump parameters.

**Table 8.** Derived cluster parameters for all 776 clusters identified, ordered by their total dense gas mass. We include the standard deviations for the longitude, latitude, velocity and temperature.

**Table A2.** CO components detected by APEX.

**Fig. 6.** Examples of clusters identified by the friends-of-friends analysis.

**Fig. 8.** Example SEDs showing a source fitted with single-component greybody model (upper panel) and with a two-component model (lower panel).

**Fig. A1.** Example APEX CO (2–1) spectra towards three sources.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

## APPENDIX A: APEX $^{13}\text{CO}$ AND $\text{C}^{18}\text{O}$ OBSERVATIONS

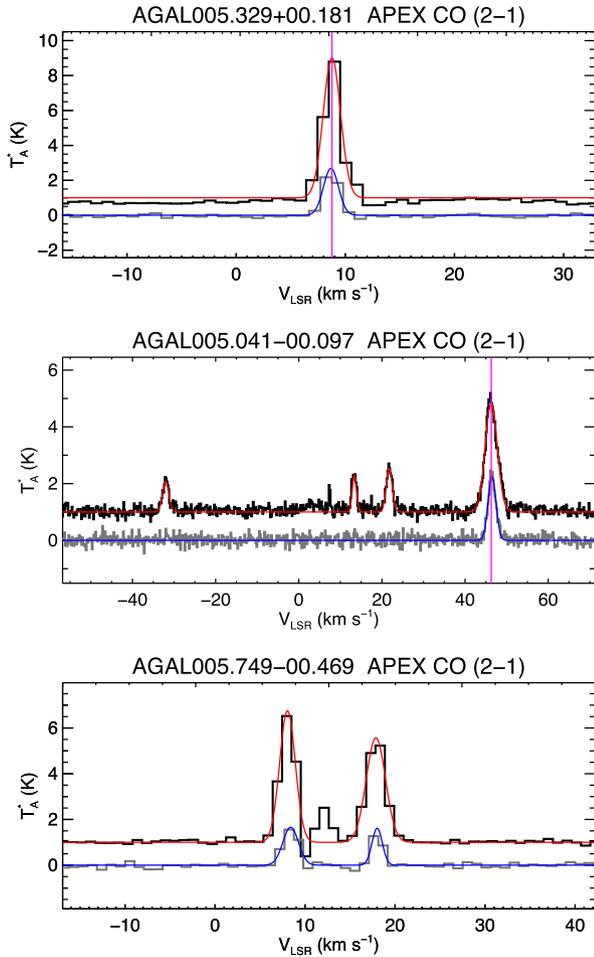
### A1 CO observations and data reduction

We targeted the  $^{13}\text{CO}$  and  $\text{C}^{18}\text{O}$  (2–1) transitions at  $\sim 220$  GHz to measure the source LSR velocity. The data were obtained between 2016 June and November at the 12 m diameter APEX telescope (Güsten et al. 2006), using the low-frequency module of the Swedish Heterodyne Facility Instrument (Vassilev et al. 2008) receiver. Backends consisted of two wide-band fast Fourier transform spectrometers (Klein et al. 2012), each covering 2.5 GHz instantaneous bandwidth with 32 768 spectral channels. We observed the transitions with a velocity resolution of  $\sim 0.1$  km s $^{-1}$  in position-switching mode with an on-source integration time of 1–4 min depending on the line strength.

We obtained spectra towards 1177 sources at an average precipitable water vapour column of  $\text{PWV} = 2.9 \pm 1.2$  mm. No emission was detected towards four sources (AGAL354.814+00.121, AGAL005.476–00.391, AGAL006.461–00.389 and AGAL338.306–00.522) and the spectra towards another eight sources were discarded due to contamination in the off-source position. These observations therefore provided useful velocity information for the remaining 1165 sources.

#### A1.1 Data reduction and line fitting

The data are analysed for peaks within a velocity range between  $\pm 200$  km s $^{-1}$  of the rest frequency of the transition to obtain



**Figure A1.** Example APEX CO (2–1) spectra towards three sources. The  $^{13}\text{CO}$  spectra is offset from the  $\text{C}^{18}\text{O}$  spectra by 1 K and the fits are shown by the red and blue lines, respectively. The vertical dashed lines shown in the upper two panels indicates the preferred source velocity. The two components shown in the lower panel have comparable intensities and optical depths and so no velocity has been assigned to the source.

the major velocity components for each position. We first added all scans for each source using the Continuum and Line Analysis Single-dish Software<sup>4</sup> from within our PYTHON code through the PYCLASS module of the PYGILDAS<sup>5</sup> package. The resulting spectra are then smoothed to a velocity resolution of  $\sim 1 \text{ km s}^{-1}$ , resulting in a mean noise level of  $\sigma = 95 \pm 50 \text{ mK}$ . The spectra are subsequently flattened using a first-order polynomial and the velocities of the emission components are determined using a PYTHON code. Here, the peaks are identified iteratively by first de-spiking the spectrum and then searching a contiguous window for emission above the  $3\sigma$

**Table A1.** Summary of the APEX observational parameters.

Parameter	Value
Galactic longitude range	$300^\circ < \ell < 60^\circ$
Galactic latitude range	$-1.54^\circ < b < 1.58^\circ$
Number of observations	1177
Number of usable spectra	1165
Frequency	220 GHz
Angular resolution	30 arcsec
Spectral resolution	$0.1 \text{ km s}^{-1}$
Smoothed spectral resolution	$1 \text{ km s}^{-1}$
Medium noise ( $T_A^*$ )	$\sim 95 \text{ mK channel}^{-1}$
Medium system temperatures ( $T_{\text{sys}}$ )	$\sim 220 \text{ K}$
Integration time (on-source)	1–4 min

noise level. The number of maxima within a window is determined and these peaks are simultaneously fitted with Gaussian profiles. The resulting sum of these profiles is subsequently subtracted from the spectrum, and the resulting spectrum is then used in the next iteration to find the next strongest emission window. Only peaks separated by at least twice the width of the fitted Gaussian are considered as major velocity components in order to avoid cluttering with multiple components per window that might be all associated with the same cloud. Negative features likely to originate from a contaminated off position were automatically identified in a similar way.

Fig. A1 presents a few examples of the spectra obtained. We have not converted these to the main beam temperature scale ( $T_{\text{mb}}$ ) as we are only concerned with the velocity information these provide and so the spectra are plotted using the antenna temperature scale ( $T_A^*$ ). In many cases, a single component is detected and the source velocity can be determined without any ambiguity (upper panel). However, in approximately 40 per cent of cases multiple components are detected (middle and lower panels); in these cases, we applied the criteria described in Section 2.1. In addition, we also used the peaks of the  $^{13}\text{CO}$  and  $\text{C}^{18}\text{O}$  components to estimate their optical depths and used this as an additional constraint – if the  $^{13}\text{CO}$  components have similar integrated intensities but one has a significantly higher optical depth (effectively selecting the component with the lowest peak  $^{13}\text{CO}/\text{C}^{18}\text{O}$  ratio) as this is also likely to have the highest column density and therefore more likely to be associated with the high column clumps identified by ATLASGAL.

In total, 3142  $^{13}\text{CO}$  and 1271  $\text{C}^{18}\text{O}$  components are detected towards 1165 clumps. As discussed in Section 2.1, in case of two velocity components, we associated the velocity of the brightest component to the ATLASGAL peak. We have been able to assign a velocity to 1115 clumps from these data. The fitted line parameters are given in Table A2. For the other 50 sources multiple emission components are detected with similar intensities and we are unable to assign a velocity with any confidence.

<sup>4</sup> <http://www.iram.fr/IRAMFR/GILDAS>

<sup>5</sup> <http://cdsarc.u-strasbg.fr/doc/man/gildas/html/gildas-python-html/>

**Table A2.** CO components detected by APEX. In this table, we indicate the velocity assigned to the source by appending a  $\star$  to the source name. We give the noise per channel of the  $^{13}\text{CO}$  in Column 2, however, the  $^{13}\text{CO}$  and  $\text{C}^{18}\text{O}$  noise agree to within a few per cent.

ATLASGAL CSC name	RMS (mK)	$^{13}\text{CO}$ (2–1)				$\text{C}^{18}\text{O}$ (2–1)				Optical depth
		$v_{\text{lsr}}$ ( $\text{km s}^{-1}$ )	Peak (K)	Width ( $\text{km s}^{-1}$ )	Intensity ( $\text{K km s}^{-1}$ )	$v_{\text{lsr}}$ ( $\text{km s}^{-1}$ )	Peak (K)	Width ( $\text{km s}^{-1}$ )	Intensity ( $\text{K km s}^{-1}$ )	
AGAL005.001+00.086 $\star$	61	2.1	5.8	1.38	20.2	2.0	1.8	0.93	4.2	0.35
AGAL005.041–00.097 $\star$	118	46.3	3.8	1.53	14.4	46.5	2.1	1.04	5.5	0.82
AGAL005.049–00.192 $\star$	72	6.1	3.2	1.53	12.4	6.1	0.9	0.97	2.2	0.32
AGAL005.139–00.097 $\star$	76	44.1	5.0	1.50	18.9	44.2	1.9	1.07	5.1	0.47
AGAL005.184+00.159	168	182.1	1.0	1.26	3.1	–	–	–	–	–
AGAL005.192–00.284 $\star$	70	8.0	3.4	2.21	18.6	8.6	0.6	1.09	1.7	0.16
AGAL005.202–00.036 $\star$	68	0.6	4.8	1.99	23.7	0.6	1.4	1.78	6.1	0.32
AGAL005.329+00.181 $\star$	69	8.8	8.0	0.79	15.9	8.7	2.7	0.69	4.6	0.40
AGAL005.371+00.319 $\star$	61	18.7	5.7	1.51	21.4	18.7	1.7	1.31	5.6	0.35
AGAL005.387–00.551 $\star$	92	8.4	1.1	1.61	4.5	–	–	–	–	–
AGAL005.389–00.384	82	21.4	1.1	9.66	26.0	–	–	–	–	–
AGAL005.392–00.409 $\star$	164	140.8	1.1	7.93	22.8	–	–	–	–	–
AGAL005.399–00.237 $\star$	53	4.8	1.9	1.34	6.3	5.0	0.4	1.01	1.1	0.24
AGAL005.409–00.302 $\star$	60	14.2	4.6	0.61	7.1	–	–	–	–	–
AGAL005.437–00.314	74	22.5	1.1	1.35	3.7	22.2	0.4	1.00	1.0	0.43

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/>.

## APPENDIX B: DISTANCE DETERMINATION

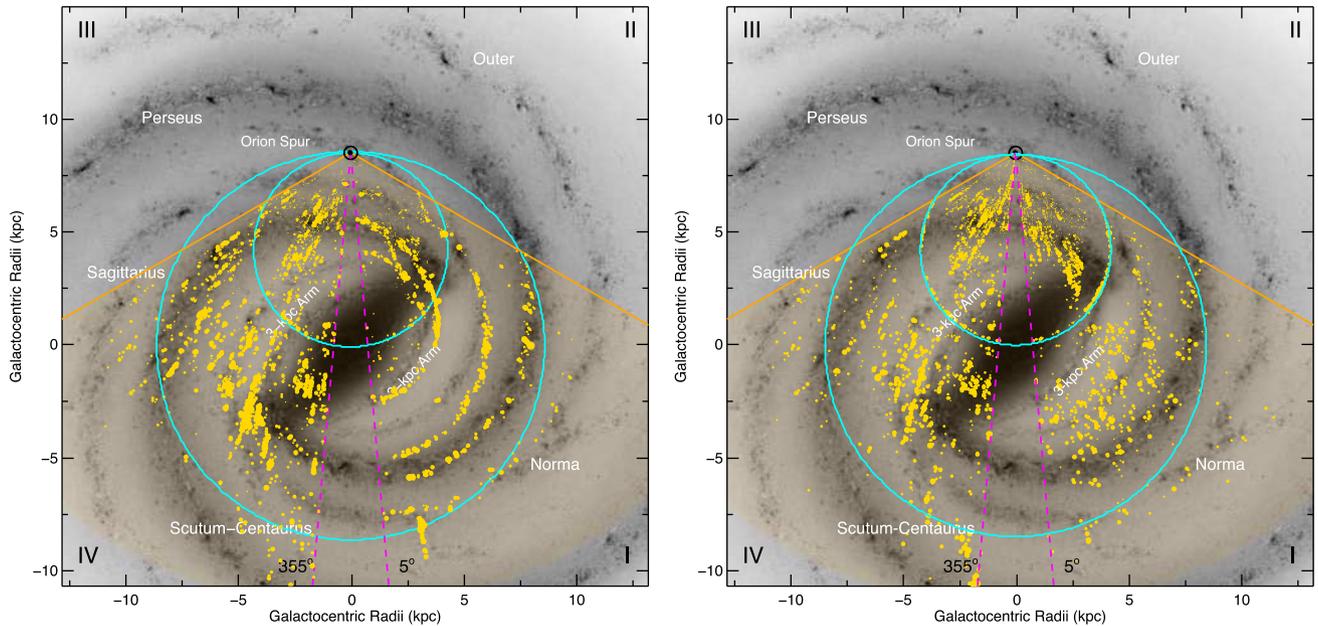
A key element required for determining the physical properties of dense clumps identified by ATLASGAL is their heliocentric distance. The most reliable way to do this is to measure the parallax of bright, compact objects such as methanol and water masers (e.g. Reid et al. 2014). These are time-consuming and complex measurements, and since they require the presence of a bright source (which is not always available, particularly for the very earliest evolutionary stages), this method is not universally applicable to all sources. The difficulties involved with these measurements and the poor frequency coverage of telescopes in the Southern hemisphere has limited the number of maser parallax measurements reported in the literature ( $\sim 150$  measurements; Reid et al. 2014) with nearly all them located in the first and second quadrants of the Galaxy. A consequence of this is that the structure of half of the Milky Way is poorly constrained. This situation is improving with the number of maser parallax measurements steadily increasing, and there has been significant progress in developing the Australia Long Baseline Array (e.g. Krishnan et al. 2017) to provide distance measurements in the third and fourth quadrants. However, it will be a long time before this method can provide distances for the majority of the ATLASGAL sample and we therefore need another method to determine distances.

Kinematic distances, although less reliable than maser parallax distance, can be determined from a source’s radial velocity with respect to the LSR ( $v_{\text{lsr}}$ ) and a model of the rotation of the Milky Way. The radial velocity can be obtained from molecular line observations towards the clumps as has been described in the previous section. There are a number of descriptions of the rotation curve of the Galaxy (e.g. Clemens 1985; Brand & Blitz 1993; Reid et al. 2014). In most cases, the kinematic distances obtained from all of them agree within their associated uncertainties (typically  $\pm 0.3$ – $1$  kpc; the largest variations are associated with sources located close to the Solar circle) and so the choice of rotation curve is not particularly critical.

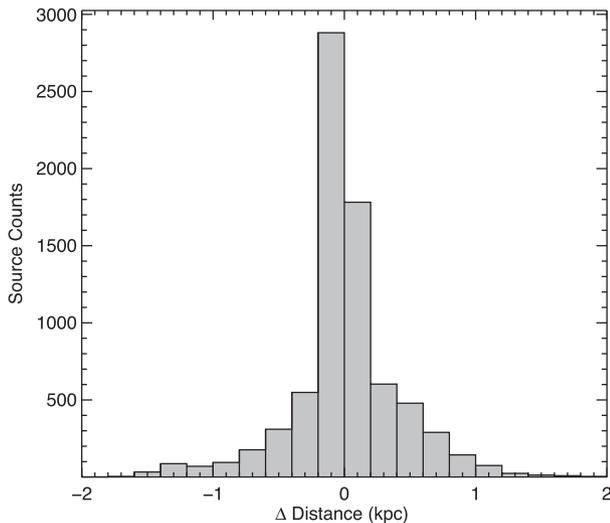
Wienen et al. (2015) have performed a detailed kinematic distance study of several thousand ATLASGAL sources using the

rotation curve of Brand & Blitz (1993). Here, we extend that work to an almost complete sample of ATLASGAL sources and refine some of their distances with an improved rotation-curve model. We have used the rotation curve determined by Reid et al. (2014) to estimate the kinematic distances as it takes advantage of all of the maser parallax measurements to constrain the model, and has been shown to provide kinematic distances that are comparable with the maser distances. An inherent problem with kinematic distances is that for sources located within the Solar circle, there are two possible distance solutions for each velocity; these are evenly spaced on either side of the tangent position and are commonly referred to as the near and far distances. These KDA need to be resolved before a unique distance can be assigned for a particular source.

Reid et al. (2016) have also developed a Bayesian maximum likelihood method that takes into account the relative position of the spiral arms along the line of sight and the latitude of the source to resolve the KDA and determine the distance (the application of this method to all of the ATLASGAL clumps can be seen in the left-hand panel of Fig. B1). This method assumes that all sources are likely to be associated with a spiral arm, which may be a reasonable assumption for star-forming regions; however, it is to be avoided, as one of the main purposes of studying Galactic structure is to determine the effect of spiral arms on the star formation process and this requires distinguishing arm from interarm sources, as much as possible. In the right-hand panel of Fig. B1, we show the distribution of clumps using distances calculated using only the rotation curve. We find that in the vast majority of cases ( $\sim 95$  per cent), the difference between the rotation-curve distance and the Bayesian maximum likelihood distance is smaller than 1 kpc (see Fig. B2). This difference is relatively modest and is almost negligible for sources located in the fourth quadrant as the position of the spiral arms are poorly constrained and parallax distances virtually non-existent. The Bayesian method does provide more reliable distances for source located near the Solar circle, where the kinematic distance uncertainties are significantly larger. As a result of these concerns, we decided to resolve the KDA via alternate methods (these will be discussed in the following subsection), and then adopt the Bayesian value if it is in



**Figure B1.** As Fig. 7. Left-hand panel: distribution of all ATLASGAL sources given by the Bayesian maximum likelihood tool developed by Reid et al. (2016). Right-hand panel: distribution of all ATLASGAL sources determined using the Reid et al. (2014) Galactic rotation curve.



**Figure B2.** Histogram showing the difference between the distance determined from using the Bayesian method presented by Reid et al. (2016) and the near/far distances determined using the Reid et al. (2014) rotation curve (note that the  $x$ -axis has been truncated). 98.4 per cent of the sample have velocity differences within the plotted range, and  $\sim 95$  per cent are within 1 kpc.

agreement with the kinematically determined solution. Otherwise, we adopt the kinematic value.

We present the source names, radial velocities, near and far distances determined from the rotation curve (Reid et al. 2014), the Bayesian distance (Reid et al. 2016), the kinematic solution and the assigned distance to the sources in Table 2.

### B1 Resolving the kinematic distance ambiguities

KDAs affect all sources located within the Solar circle (i.e. Galactocentric radius ( $R_{GC}$ )  $< 8.35$  kpc; Reid et al. 2014) and these need to be resolved before a unique distance can be determined for a

particular source. We present a flow chart in Fig. 5 to illustrate the various steps we have used to resolve these distance ambiguities.

(i) We match clumps to reliable distances reported in the literature (maser parallax, e.g. Reid et al. 2014 and spectroscopic measurements, e.g. Moisés et al. 2011). This is done by comparing the longitude, latitude and velocities of the sources with positions of the given these studies and where a correlation is found the distance is adopted.

(ii) Sources located outside the Solar circle are not affected by this twofold distance ambiguity and so the kinematic distance is adopted; however, this only applies to a relatively small fraction of the sources ( $\sim 300$ ).

(iii) We can exclude sources located close to the tangent velocity ( $|v_{\text{clump}} - v_{\text{tangent}}| < 10 \text{ km s}^{-1}$ ) from the KDA tests, as the difference between the near and far distances becomes smaller than their associated uncertainties: this eliminates  $\sim 10$  per cent of the sample.

(iv) The latitude distribution of molecular material and star formation is tightly correlated with the Galactic mid-plane. High-mass stars found within the Solar circle have a scale height of  $\sim 30$  pc (e.g. Reid 2000; Green & McClure-Griffiths 2011; Urquhart et al. 2014b), and since most high-mass star-forming clumps have masses on the order of  $10^3 M_{\odot}$ , we would expect them to have a similar scaleheight distribution. We can therefore use this as an additional constraint on the possible distance for clumps located inside the Solar circle. We have calculated the distance of all sources from the Galactic mid-plane assuming they are located at the far distance; if this distance is greater than 120 pc (i.e. four times the scaleheight), then the far distance is considered significantly less likely and the source is placed at the near distance.

(v) Distances to many H II regions have been determined by looking for H I absorption features against the strong radio continuum emission associated with the ionized gas (e.g. Wienen et al. 2015; Urquhart et al. 2012; Anderson & Bania 2009; Kolpak et al. 2003). The envelope of cooler H I gas surrounding a molecular cloud located between the H II region and the observer will produce an absorption feature in the H II continuum emission at the velocity of

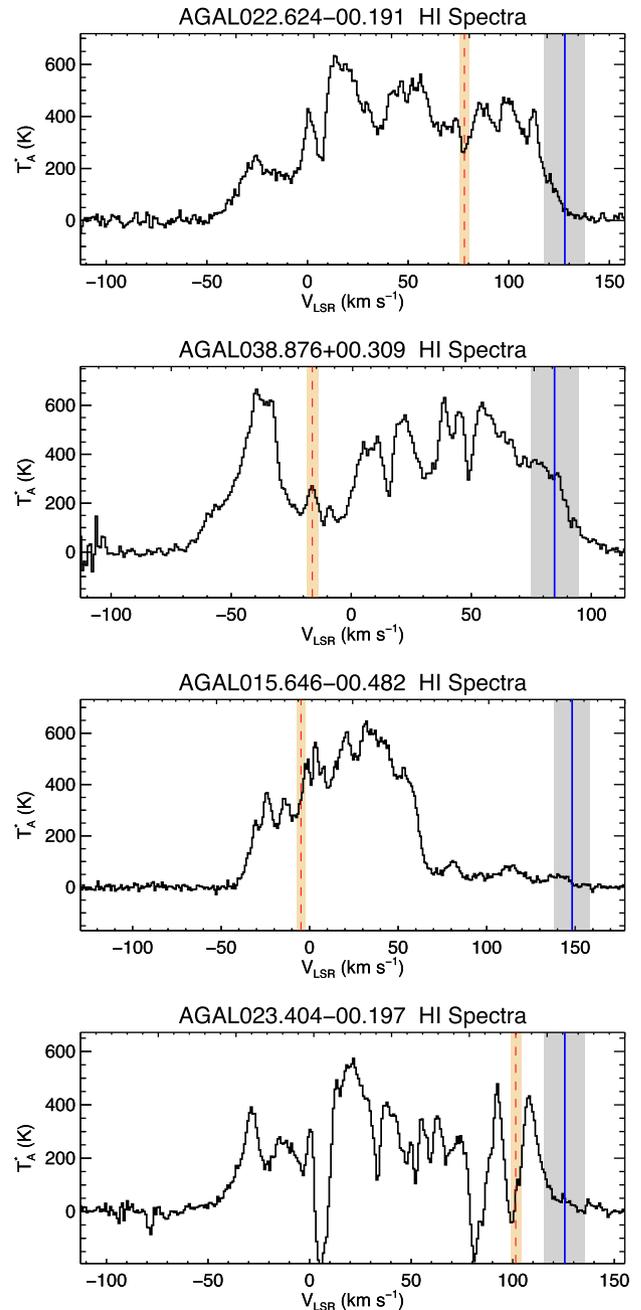
the intervening clump. If the H II region is located at the far distance, we would expect to see absorption at higher velocities than the source, and these should extend all the way up to the velocity of the tangent position. However, if the H II region is located at the near distance we would expect to see absorption at the same velocity as the H II region but not at higher velocities. We have matched our clumps with H II region studies reported in the literature and if a reliable distance has been found then this has been adopted (e.g. Kolpak et al. 2003; Urquhart et al. 2012, 2013b). If no distance is available, then the source has been returned to the rest of the sample for further analysis.

(vi) H I SA is then employed to resolve the distance ambiguity for any remaining sources. Clumps located at the near distance are likely to be associated with an absorption feature in the H I spectra at the same velocity as the source (due to the cold clump absorbing emission from the warmer diffuse H I gas located behind it). Conversely, if the clump is located at the far distance and the warm H I gas is distributed throughout the Galactic plane, there would be no corresponding dip in the H I spectra at the same velocity of the source. This method for resolving the distance ambiguities has been widely used in a number of studies (e.g. Roman-Duval et al. 2009; Anderson & Bania 2009; Jackson et al. 2006).

We have extracted H I spectra from the Southern Galactic Plane Survey (McClure-Griffiths et al. 2005) and the VLA Galactic Plane Survey (Stil et al. 2006) archives towards all ATLASGAL clump. The resolution of these H I surveys is comparable to the typical sizes of the clumps ( $\sim 1$  arcmin; Contreras et al. 2013) and so we sum emission over one resolution element of the respective surveys (typically 9 pixels). We make no attempt to remove background emission as these clumps are likely to be embedded in larger more diffuse GMC structures which will also absorb H I emission at a similar velocity as the sources, and the boundaries of these larger structures are unknown.

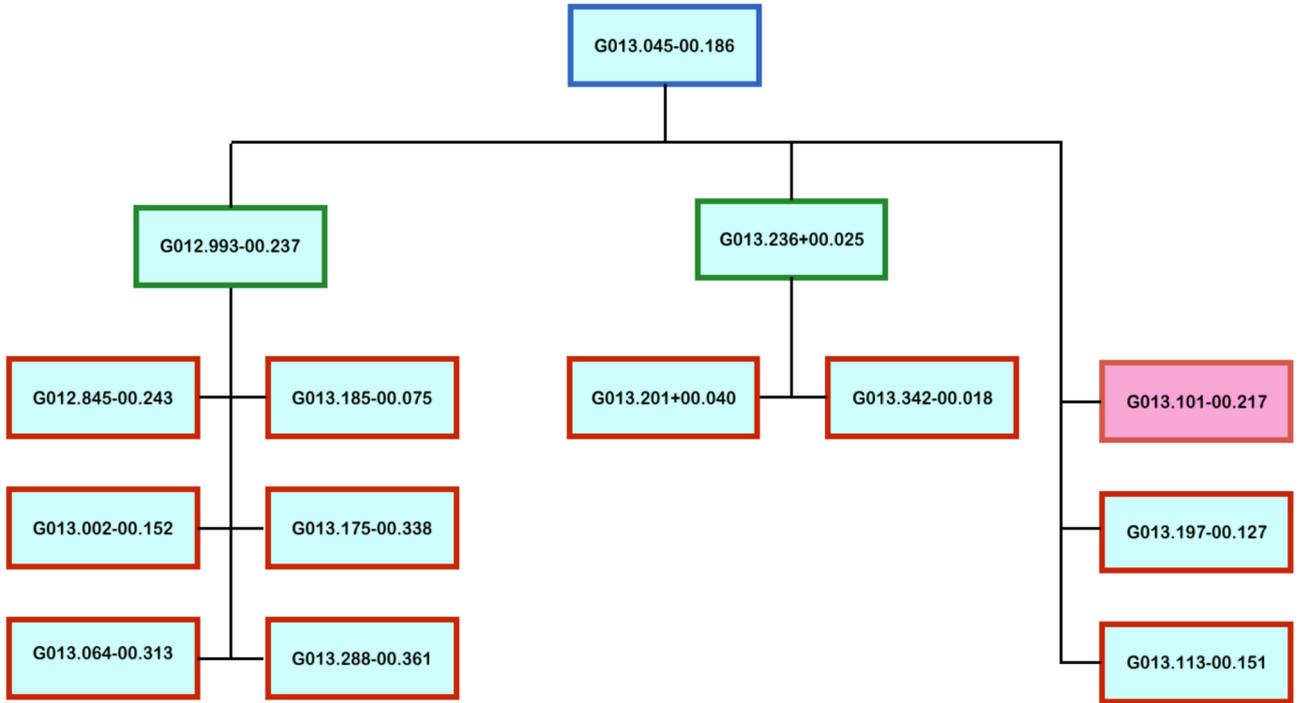
The H I spectra and the source velocity were combined in the same plot to facilitate comparison of the H I profile around the velocity of the source (see Fig. B3 for some examples). A review of a representative sample of the spectra identified four assignments (near, far, ambiguous and problem) that were found to cover all of the possible outcomes. The first three options are self-explanatory, while the last one was used to flag H I spectra that are contaminated by strong continuum emission from evolved H II regions located nearby. All plots were visually inspected by two of the lead authors (JSU and AG) independently to resolve the distance ambiguities. Once the analysis of the H I data was completed, the two sets of results were compared. If the solutions agreed, then the distance solution was adopted: these were considered to be high-reliability distances. If one of the solutions was near or far and the other was classed as ambiguous then the source was placed at the appropriate near or far distance but given a lower reliability flag. All other sources where the solutions disagreed were considered ambiguous and no distance was concluded from the H I analysis. We present a key showing the distance solution adopted depending on the possible combinations of assigned solutions in Table B1.

(vii) If the H I data were inconclusive, we searched for evidence of extinction towards the clumps. If a clump is associated with an IRDC (e.g. Rathborne, Jackson & Simon 2006), it suggests the source is in the foreground with respect to the relatively bright IR emission that fills the inner part of the Galaxy. A recent study of a small sample of IRDCs by Giannetti et al. (2015) found approximately 11 per cent were actually located at the far distance and so placing these at the near distance is reasonable for the majority of clumps. We have searched for associations using the IRDC catalogue compiled by



**Figure B3.** Example H I spectra extracted towards dense clumps identified from the ATLASGAL survey. The velocity of the tangent position is determined from a fit to H I data (e.g. McClure-Griffiths & Dickey 2007) and is indicated by the blue line; the grey shaded region covers a velocity range of  $\pm 10$  km s $^{-1}$  centred on the tangent velocity. Sources in this region are placed at the distance of the tangent point. The dashed red vertical line shows the velocity of the source with the yellow shaded region showing the typical FWHM linewidth of the molecular lines. From top to bottom, panels show examples of clumps located at the near distance, the far distance, or identified as ambiguous distance and problem sources, respectively.

Peretto & Fuller (2009) to identify sources associated with IRDCs. This catalogue does not include the inner  $10^\circ$  of the Galactic mid-plane and so we have visually inspected the mid-IR images to identify possible IRDC associations for sources in this region. Any clumps found to be associated with an IRDC are placed at the near distance (the justification for this is discussed in Appendix B3).



**Figure B4.** Schematic diagram showing the hierarchical structure of a cluster identified from the friends-of-friends analysis. At the top of the diagram (outlined in blue), we have the structure identified by using the largest set of search parameters (i.e.  $R_{\text{fof}} < 8$  arcmin and  $\Delta v_{\text{fof}} < 8 \text{ km s}^{-1}$ ), the second row (outlined in green) shows the subregions identified by using the intermediate search parameters (i.e.  $R_{\text{fof}} < 6$  arcmin and  $\Delta v_{\text{fof}} < 6 \text{ km s}^{-1}$ ), while the remaining rows show the smallest subregions identified using the smallest search parameters (i.e.  $R_{\text{fof}} < 4$  arcmin and  $\Delta v_{\text{fof}} < 4 \text{ km s}^{-1}$ ); these are outlined in red. Regions located at the near distance are shown with a cyan background, while those found at the far distance are shown in with pink background. The distances associated with the smallest subregions are considered to be the most reliable.

**Table B1.** Possible kinematic distance options applied by the two authors and how these have been used to obtain a final solution. The question marks (?) indicate that although a solution has been adopted it is considered to be at a lower level of confidence.

Author 1	Author 2	Adopted solution
Near	Near	Near
Far	Far	Far
Far	Ambiguous	Far?
Near	Ambiguous	Near?
Near	Problem	No solution
Far	Problem	No solution
Ambiguous	Ambiguous	No solution

(viii) For any remaining sources, we searched the literature to see if a distance had been previously assigned and whether it was considered to be reliable (i.e. an evidence-based distance and not simply assumed to be at the near distance). If so, the corresponding kinematic distance solution was adopted (e.g. Svoboda et al. 2016; Battisti & Heyer 2014). Spectroscopic distances (e.g. Moisés et al. 2011; Stead & Hoare 2010) and parallax distances (e.g. Reid et al. 2014, and references therein) from the literature were simply adopted.

## B2 Sources located near the Solar circle

Assigning distances to sources located near the Solar circle (i.e.  $|v_{\text{lsr}}| < 10 \text{ km s}^{-1}$ ) is a little more complicated because we also need to consider the additional uncertainty due to streaming motions in the Galaxy ( $\pm 10 \text{ km s}^{-1}$ ). These are already taken into account when estimating the errors in the derived distance. However, for

sources located near the Solar circle, this can result in a source actually located within the Solar circle (associated with a near/far distance ambiguity) appearing to be in the outer Galaxy (with no distance ambiguity) or vice versa. For sources located within the Solar circle, we can simply try to resolve the distance ambiguity in the normal way using the H I SA method; however, for sources with velocities that place them outside the Solar circle we only obtain a single distance although due to streaming motions, they may actually be located inside the Solar circle. We therefore apply the following additional criteria:

- (i) If the kinematic solution suggests that a near distance is likely, then no distance is assigned. These sources are likely to be located very close by ( $\sim 1$  kpc) and are therefore probably relatively low-mass clumps. Excluding these clumps is unlikely to impact our statistical analysis of the properties of the sample and its Galactic distribution.
- (ii) By similar rationale, if the kinematic solution suggests a far distance, then this source is likely to be at the far distance and we assign the source to the far distance.

## B3 Summary of the kinematic distance analysis and comparison with the literature

In total, we resolved the distance ambiguity for 7091 sources, of which we rate 6292 as ‘reliable’; these values correspond to 88.6 per cent and 78.6 per cent of the ATLASGAL CSC, respectively. We were unable to determine a distance for 911 sources; however, velocities were not available for 240 of these (as discussed in Section 2.3). The distribution of all of these clumps determined using the Reid et al. (2014) model is shown in the right-hand panel of

**Table B2.** Summary of the kinematic distance solutions. The roman numerals (i)–(viii) given in Column 1 refer to the various steps described in Appendix B1, while the (ix) and (x) are used to identify sources for which the H I SA is ambiguous and clumps where no velocity is available.

Step	Description of method	Total number distances assigned
(i)	Parallax/Spectroscopic	26
(ii)	Outer Galaxy	84
(iii)	Tangent	760
(iv)	Z distance	1295
(v)	H I EA	140
(vi)	H I SA Near (?)	3126 (559)
(vi)	H I SA Far (?)	1280 (612)
(vii)	IRDC associations	269
(viii)	Literature	111
(ix)	Ambiguous (Solar circle)	671 (43)
(x)	No $v_{lsr}$ available	240

Fig. B1. In Table B2, we present a breakdown of how the distance have been assigned following the application of the various steps outlined in Appendix B1.

We placed 4817 clumps at the near distance, 1430 at the far distance and 760 sources at the tangent position. The proportion of sources placed at the near and far distances is therefore  $\sim 77$  per cent and 23 per cent, respectively. This is similar to the findings of other studies (e.g. Eden et al. 2012 assigned  $\sim 75$  per cent of sources to the near distance).

We cross-correlated the positions of our sample of dense clumps with the position of IRDCs (e.g. Peretto & Fuller 2009), thereby identifying 1912 IRDCs in our sample that can be used to check the consistency of our distance assignments. IRDCs are thought to be located in the foreground between us and the bright diffuse IR emission that fills much of the inner Galaxy. They are so dense that the background IR emission is totally absorbed towards these objects resulting in these objects appearing in extinction with respect to the background field stars. If our distance solutions are reliable, we should expect to find the vast majority of IRDCs will have been placed at either the near or tangent distances.

Excluding step (vii), we have resolved the distance ambiguity towards 1606 of the matched IRDCs, of which 1457 have been placed at the near or tangent distances. This corresponds to 90.7 per cent of the IRDC sources. Detailed studies of the distances of large and representative samples of IRDCs have not yet been performed and so the actual ratio of near/far sources is not well constrained. However, a recent study of  $\sim 40$  IRDCs associated with the ATLASGAL Top100 sample (König et al. 2017) found 11 per cent to be located at the far distance (Giannetti et al. 2015). Our results are therefore consistent with what has been previously found and thereby

increases confidence in the reliability of our results. This also provides strong support for automatically placing all remaining IRDCs for which we were unable to resolve the distance ambiguity at the near distance [i.e. step (vii)].

We have also compared our distance solutions with a number of other recent studies. We present a summary of this analysis in Table B3; this reveals a high level of agreement between our distances and nearly all of the comparison samples. Given that the reliability of the H I analysis methods used here is only  $\sim 80$  per cent (e.g. Busfield et al. 2006; Roman-Duval et al. 2009), the level of agreement with most studies considered here and provides further confidence in the reliability of our distance assignments. We have not performed a direct comparison with the recent work present by Elia et al. (2017) as only  $\sim 40$  per cent of their distances were the result of distance ambiguity resolution: the remainder were arbitrarily placed at the far distance and so the samples are not comparable.

### APPENDIX C: IDENTIFYING LARGER SCALE STRUCTURES: GMCS AND COMPLEXES

In the previous section, we described how we have determined kinematic distance solutions to  $\sim 7000$  ATLASGAL clumps; however, there are still a significant number ( $\sim 10$  per cent) of sources for which we have not been able to assign a distance. Furthermore, the reliability of many of the criteria we applied to resolve the distance ambiguities have an inherent uncertainty of  $\sim 20$  per cent. We know that much of the star formation taking place in the Galaxy lies within the Solar circle ( $R_{GC} < 8.35$  kpc) and is tightly concentrated to a relatively narrow range in the mid-plane (scaleheight  $\sim 30$  pc) – this has already been used to help resolve the distance ambiguities to many sources. We also know that a large fraction of the star formation is also further concentrated in large star-forming complexes that can be associated with numerous GMCs (Murray & Rahman 2010; Urquhart et al. 2014b).

We have thus far treated the ATLASGAL clumps as isolated/individual sources and determined their velocities and distances on an individual basis. Given that these clumps are simply the highest column density regions of GMCs, we can group them together in  $lbv$ -space to identify the large-scale structures they are part of. This has many potential advantages, e.g.:

(i) It allows the distance solutions of clumps associated with the same GMCs and complexes to be assessed in a statistical way and for potential erroneous solutions to be identified and excluded, which will improve the overall reliability of the distances.

(ii) It allows a distance to be applied to sources that are found to be associated with a particular GMC but for which we were unable to resolve the ambiguity using the H I analysis.

**Table B3.** Comparison of derived distance solutions with recent reports reported in the literature.

Survey	Number of matches	Agree	Disagree per cent	Agreement	Reference
BGPS	902	698	204	77.4	1
Reid Bayesian	7130	5160	1970	72.4	2
ATLASGAL	1882	1410	472	74.9	3
RMS	695	557	138	80.1	4
BGPS	91	61	30	67.0	5
MMB	376	226	150	60.1	6

References: (1) Svoboda et al. (2016); (2) Reid et al. (2016); (3) Wienen et al. (2015); (4) Urquhart et al. (2014b); (5) Battisti & Heyer (2014) and (6) Green & McClure-Griffiths (2011).

(iii) Many GMCs are associated with strong velocity gradients and so measuring this for the individual clumps can result in significant differences between the component clumps. Although the differences in velocity between individual clumps might actually be quite modest ( $<5 \text{ km s}^{-1}$ ), this can result in large differences in their kinematic distances ( $\sim 0.5 \text{ kpc}$ ) which may impact source properties and introduce significantly more scatter in their Galactic distribution.

(iv) Many of the most prominent complexes in the Galaxy are already well studied (e.g. W31, W43 and G305) and there are very reliable distances available: identifying sources that are likely to be associated with each other allows us to adopt these distances.

### C1 Friends-of-friends analysis

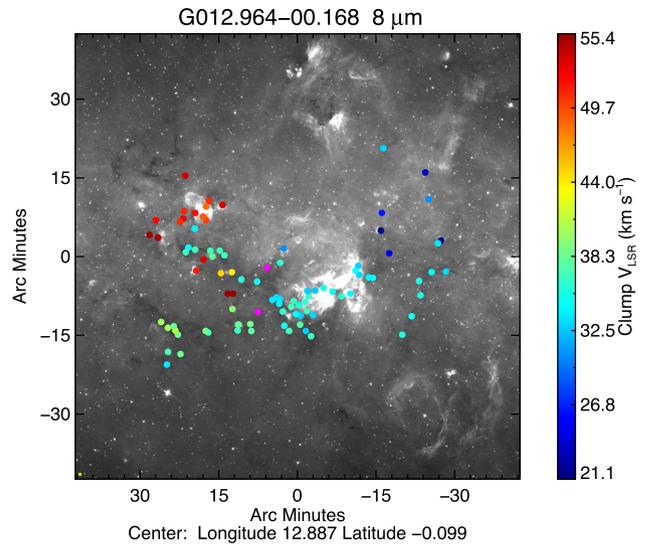
We have implemented the friends-of-friends method as a first step to identify possible larger scale associations of clumps. We employed three different sets of parameters beginning with a matching angular radius ( $R_{\text{fof}}$ ) of 4 arcmin and velocity dispersion of ( $\Delta v_{\text{fof}}$ )  $4 \text{ km s}^{-1}$  and successively increasing the search parameters by 2 arcmin and  $2 \text{ km s}^{-1}$ . The selection of the first set of search parameters is based on the analysis by Svoboda et al. (2016) on the reliability of assigning KDA solutions to groups of sources that are spatially and kinematically correlated sources (i.e.  $\ell bv$ ); they refer to this as *Distance Resolution Broadcasting*. These parameters were determined by comparing the KDA solutions in groups of sources as the search parameters were increased and evaluating the number of disagreements within the group.

The broadest set of search parameters (i.e.  $R_{\text{fof}} < 8 \text{ arcmin}$  and  $\Delta v_{\text{fof}} < 8 \text{ km s}^{-1}$ ) were selected as to include sources or groups of sources, identified using the smaller search parameters, that are associated with the same GMCs (these parameters correspond to size scales of  $\sim 30 \text{ pc}$  at a distance of 15 kpc, which is the typical cloud radius (Miville-Deschênes, Murray & Lee 2017), and FWHM velocity dispersion of  $10 \text{ km s}^{-1}$ , typical of GMCs found in the Large Magellanic Cloud (Hughes et al. 2010). The intermediate set of search parameters roughly corresponds to the properties of smaller molecular clouds.

Applying these search parameters produces a hierarchy of correlated sources that correspond to different scale structures in a GMC. At the smallest scale, these identify tightly grouped clusters of clumps associated with one of perhaps many dense subregions within a cloud but where a common distance solution can be considered reliable. The second set of search parameters associates the coherent subgroups with their large-scale molecular clouds, while the last set of search parameters links the molecular clouds with their host GMCs and star-forming complexes.

We are primarily interested in the largest scale structures identified with the largest search parameters. However, the distances for all of the subregions within these larger scale structures have been determined independently and so considering the distances of substructures provides a strong consistency check on the viability of the GMCs. We therefore begin at the top of the hierarchical tree (at the largest scale) and apply the following criteria to determine the reliability of a particular structure:

(i) The agreement in distance solutions for all constituent clumps within a structure must be better than 70 per cent. If this is not the case, then we look at the distances of the individual subregions. If a particular subregion that has a distance resolution that is out of step with that of the larger scale structure and the subregion can be removed without impacting the integrity of the larger structure then



**Figure C1.** Example of clustering analysis towards W33 complex. For description of image refer to Fig. 6. The hierarchical structure of W33 is used as an example of the association method discussed in Appendix C1 and shown in the flow chart presented in Fig. B4.

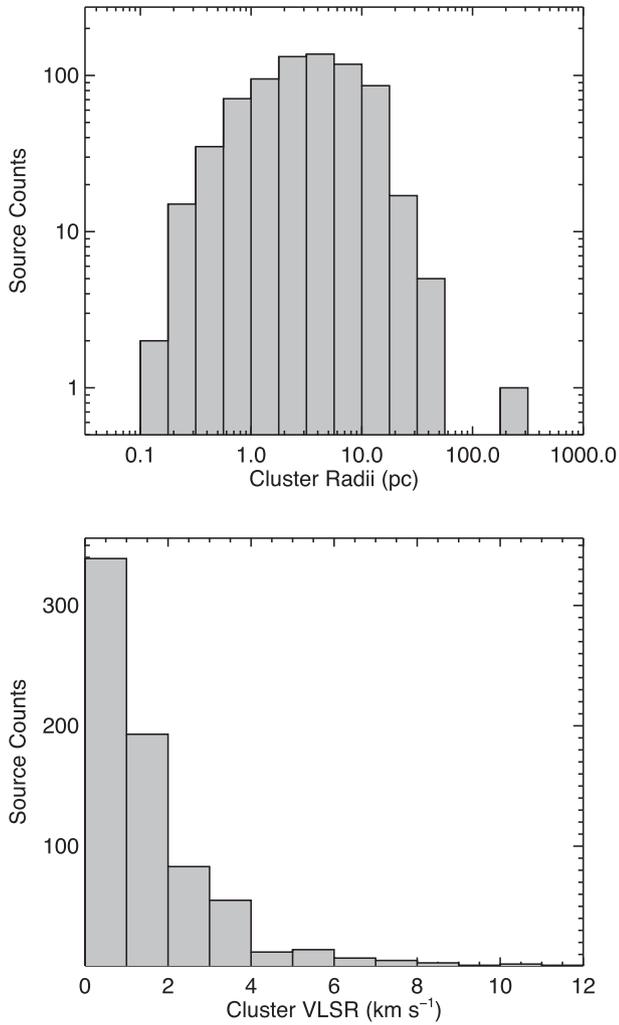
this is done. If this cannot be done, then the top level structure is discarded and we perform the same analysis on the next level down in the hierarchical structure.

(ii) A visual inspection of the distribution of the structure with respect to the large-scale structure traced by the mid-IR emission images and maps of the combined ATLASGAL+PLANCK emission (Csengeri et al. 2016b) reveal good correlation. In some cases inspection of the mid-IR images revealed that two or more clusters identified are actually part of larger scale structures and in such cases these have been combined into a single entity.

(iii) As a final step, we identify additional possible members by correlating the spatial extent of the clusters with the positions of clumps for which we were unable to determine a velocity. If a clump was found to fall within the footprint of a single cluster on the sky, then it has been assigned to that cluster. If a clump fell within the footprint of 2 or more clusters along the line of sight, then cluster membership is ambiguous and no cluster was assigned.

In Fig. B4, we present a schematic diagram showing the results of this analysis for one such cluster identified. The largest structure (G013.045–00.186) consists of two subregions (G012.993–00.237 and G013.236+00.025), each of which themselves consist of between 2 and 6 smaller subregions, and three smaller subregions (G013.101–00.217, G013.197–00.127 and G013.113–00.151). The distances for the smallest subregions are determined independently, and as can be seen in Figs B4, 10 of the 11 of these have been placed at the near distance (cyan backgrounds); this includes all of the subregions associated with the two larger regions identified by the intermediate set of search parameters and these are therefore considered to be reliable structures. In this case, we are able to exclude the one subregion without affecting the integrity of the top level structure. In Fig. C1, we show an image of the cluster towards W33 identified by this analysis.

In Fig. C2, we present plots of the physical size and velocity dispersion of the clusters as determined from the standard deviation of their associated clumps. The size of the clusters are from a fraction of a parsec to several tens of parsecs, but sizes of a few parsecs are more



**Figure C2.** Distribution of cluster sizes and velocity dispersions. The bin size used in the upper panel is 0.25 dex and in the lower panel, it is  $1 \text{ km s}^{-1}$ .

typical. One cluster stands out as being exceptionally large with a size of several 100 parsecs; this source is the *Wisp* structure identified by Li et al. (2013), which is one of the largest velocity coherent molecular structures detected so far. The velocity dispersion of the clumps associated with each cluster are typically lower than a few  $\text{km s}^{-1}$  and in line with the velocity dispersions reported for Galactic and extragalactic GMCs ( $5\text{--}20 \text{ km s}^{-1}$ , Schneider et al. 2006; Fukui et al. 2009).

## APPENDIX D: SED: PHOTOMETRY AND FITTING PROCEDURE

### D1 Photometry

We use an aperture-and-annulus geometry to measure flux densities about each source. We initially set a circular aperture to a spatial position that is optimal across all observed bands for the given source. This aperture is then re-centred to within one ATLASGAL beam radius (i.e.  $r_{\text{search}} = 19.2 \text{ arcsec}$ ) of the peak emission found in the 250, 350 or 870  $\mu\text{m}$  band, using the first band where no saturation

is found<sup>6</sup> within the source aperture. We set the aperture radius to twice the major axis reported in Contreras et al. (2013) and Urquhart et al. (2014a) in order to ensure the aperture to be large enough that most of the source emission lies within the aperture, while still being small enough to avoid cutting into the background for more complex sources. Saturated pixels present within the aperture are set to the maximum pixel value of the  $5 \times 5 \text{ arcmin}$  image and the flux is thereafter regarded only as a lower limit.

The background is determined over an annulus about the aperture with inner and outer radii of 1.5 and 2.5 times the aperture size, respectively. In case that the annulus width is smaller than the ATLASGAL beam size, the outer annulus radius is increased to be at least three times the ATLASGAL pixel size (i.e. 18 arcsec) larger than the inner radius, to allow for a statistical analysis of the background annulus flux. Any pixels within the background annulus with a flux above the source aperture’s peak-pixel flux are omitted, as we assume these pixels represent sources within the background aperture. We take the median background flux (rather than the mean) to reduce any influence of a fainter source within the annulus. The background flux is then normalized to the area of the source aperture and subtracted from the latter, yielding the background-corrected source flux.

A negative background-corrected flux may be indicative of either the background flux being overestimated as the background annulus is cutting into a nearby source or of absorption in the source aperture. We take care of the former issue as described in the last paragraph by ignoring pixels above the maximum aperture flux and taking the median value of the background pixels, but the latter issue may result in negative flux measurements even when a point source is clearly visible within the aperture. To automatically identify a point source within the source aperture, we have empirically determined that a point source in an arbitrary group of pixels has at least 80 per cent, 50 per cent and 20 per cent of the pixels above the  $1\sigma$ ,  $2\sigma$  and  $3\sigma$  noise levels, respectively. In case a point source is identified in this way, we restrict the pixels taken into account for photometry to those with a value above the median background pixel level. Any pixels below this limit are likely to be caused by absorption within the aperture and are therefore neglected, resulting in a positive background-corrected flux. Applying this method allowed us to obtain 2727 fluxes over all bands, affecting a total of 1814 sources, successfully recovering the flux for a visible point source that otherwise would have been missed.

The flux uncertainties are estimated from the pixel noise level of the image and a general flux measurement uncertainty added in quadrature. We assume a rather conservative flux measurement uncertainty, as we not only consider the absolute calibration error but also take into account the uncertainty involved in determining the source aperture (which is applied to all bands and might not always be optimal). We therefore assume a flux measurement uncertainty of 15 per cent for ATLASGAL, 20 per cent for the 70, 160, 250 and 350  $\mu\text{m}$  *Herschel* bands, and an uncertainty of 50 per cent for the 500  $\mu\text{m}$  *Herschel* band due to the large pixel size. We assume a general measurement uncertainty of 30 per cent for the mid-IR bands, although the 8  $\mu\text{m}$  MSX band earns an increased measurement uncertainty of 50 per cent due to the possible influence of PAHs within the band. The pixel noise level is determined over the full  $5 \text{ arcmin} \times 5 \text{ arcmin}$  image, where pixels within a beam around local maxima are blanked. The noise level is then determined over

<sup>6</sup> Pixels where the detector suffered from saturation are found as NaN values in the reduced maps.

the filtered image, where the  $1\sigma$  level is determined from a Gaussian fit to the histogram of the remaining pixel fluxes.

## D2 Overview of the fitting procedure

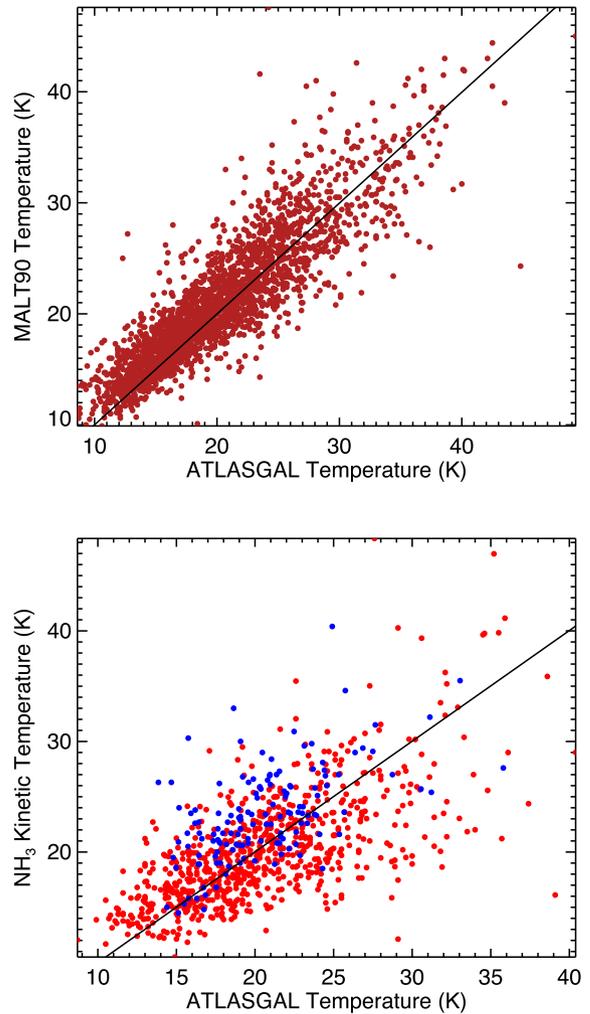
As in König et al. (2017), the measured SED is fit simultaneously by a combination of a greybody and a blackbody, where the greybody is fit to wavelengths upwards of  $20\ \mu\text{m}$ , fitting the emission of the cold dust. The blackbody component is added to reflect the presence of a more evolved, optically thick embedded hot component and is only added if at least two flux measurements are present downward of  $70\ \mu\text{m}$ . The fitting is performed using a Levenberg–Marquardt least-squares minimization, allowing us to estimate the fitting parameter uncertainties from the covariance matrix as calculated by the algorithm. To reduce the number of free parameters, we keep the dust emissivity spectral index of the greybody fixed to 1.75, which is the mean value as calculated over all dust models from Ossenkopf & Henning (1994), allowing a fit with fewer available data points, as well as allowing better comparison with the literature (e.g. Thompson et al. 2004; Nguyen Luong et al. 2011).

We take a different approach for emission in the  $70\ \mu\text{m}$  band in contrast to König et al. (2017). The emission in this band is generally interpreted as being contaminated with the emission of small grains (Compiègne 2010; Compiègne et al. 2010) in addition to having the previously addressed contributions from both the cold dust and a more evolved hot component within the clump (e.g. Beuther et al. 2010). For these reasons, we take the emission at  $70\ \mu\text{m}$  as an upper limit in the fitting process. When no emission is found at the  $70\ \mu\text{m}$  band, we use a conservative upper limit of five times the noise determined over an aperture including the background annulus or the point source sensitivity level as determined for the instrument (i.e.  $21\ \text{mJy beam}^{-1}$ ),<sup>7</sup> whichever is higher. Fitting the SEDs provides us with reliable measurements of the dust temperature of the cold component that characterizes the emission from dust and the bolometric flux; these will be used in the following section together with source distances to derive the physical properties of the clumps.

## D3 Consistency tests

In the upper panel of Fig. D1, we compare the temperatures derived from the greybody fits with those derived by a recent study by Guzmán et al. (2015) who performed a similar method of aperture photometry on a sample of  $\sim 2500$  ATLASGAL clumps observed as part of the MALT90 survey (Jackson et al. 2013). This plot illustrates the excellent agreement between the greybody temperatures determined from our work and the work reported by Guzmán et al. (2015) to estimate the fluxes and to fit the SEDs. The mean difference between the two sets of temperature measurements is  $0.8 \pm 0.1\ \text{K}$  with a standard deviation of  $2.8\ \text{K}$ . In the lower panel of Fig. D1, we show the correlation between the kinetic dust temperatures derived from the ammonia ( $\text{NH}_3$ ) (1,1) and (2,2) transitions (Urquhart et al. 2011; Wienen et al. 2012). This plot also shows a strong correlation between the temperature of the gas and the dust (the Spearman rank coefficient  $r$  is 0.64 with a  $p$ -value  $\ll 0.01$ ), although we note the scatter is clearly more significant (the mean difference is  $0.7 \pm 0.14\ \text{K}$  with a standard deviation is  $4.8\ \text{K}$ ).

The bolometric fits derived here are consistent with the results of a similar analysis of dust emission reported by Guzmán et al. (2015)



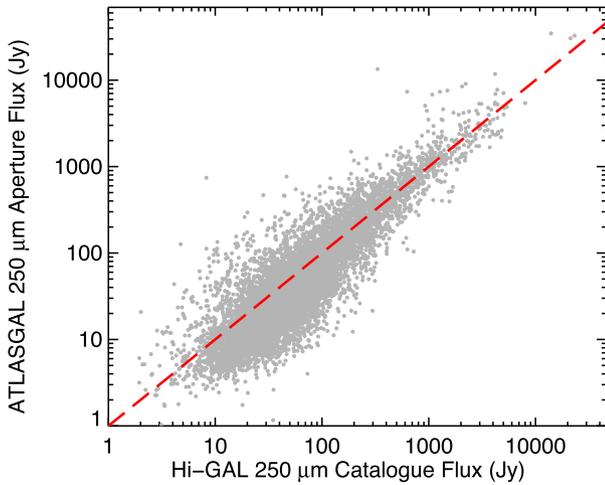
**Figure D1.** Upper panel: comparison between dust temperatures derived here and those determined by the MALT90 team (Guzmán et al. 2015). This subsample of  $\sim 2500$  includes a significant fraction of all of the sources observed as part of the MALT90 survey (a total of  $\sim 3000$  ATLASGAL clumps were observed as part of this programme of follow-up observations). The solid line indicates the line of equality. Lower panel: comparison between the dust temperatures and kinematic gas temperatures determined for  $\sim 1500$  clumps from  $\text{NH}_3$  (1,1) and (2,2) inversion transitions (Wienen et al. 2012; Urquhart et al. 2011, red and blue circles, respectively). Least-squares fits to the data presented in both plots result in slopes that are within  $1\sigma$  of being linear.

and with independently estimated kinetic temperatures of the dense molecular gas. The strong agreement with these two previously published studies provides confidence in our fitting method and the derived results.

Finally, we also compared the fluxes derived in this work with those found in the Hi-GAL CSC for the inner Galaxy (Molinari et al. 2016). To obtain the fluxes from the Hi-GAL CSC, we summed up the fluxes of all compact sources found within our source aperture. The result can be seen in Fig. D2, for the  $250\text{-}\mu\text{m}$  band. Although there is a significant amount of scatter for individual sources ( $\sim 10$ ), a strong correlation between the fluxes obtained with the two significantly different methods is found ( $r = 0.86$  with a  $p$ -value  $< 0.0013$ ), showing the general consistency between the two photometric approaches.

In total, we were able to successfully fit the SEDs of 7861 sources ( $\sim 98$  per cent of the full CSC sample). 2586 of these sources

<sup>7</sup> [herschel.esac.esa.int/Docs/PMODE/html/ch02s03.html](https://herschel.esac.esa.int/Docs/PMODE/html/ch02s03.html)



**Figure D2.** Comparison of flux densities obtained from the Hi-GAL CSC with the fluxes obtained through aperture photometry in this work. The dashed red line indicates equality between the two sets of measurements.

(~33 per cent) are fitted with a single-component greybody and 5275 sources (~67 per cent) with the two-component model.

<sup>1</sup>*Centre for Astrophysics and Planetary Science, University of Kent, Canterbury CT2 7NH, UK*

<sup>2</sup>*Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-53121 Bonn, Germany*

<sup>3</sup>*INAF – Istituto di Radioastronomia and Italian ALMA Regional Centre, Via P. Gobetti 101, I-40129 Bologna, Italy*

<sup>4</sup>*INAF – Osservatorio Astronomico di Cagliari, Via della Scienza 5, I-09047 Selargius (CA), Italy*

<sup>5</sup>*Astrophysics Research Institute, Liverpool John Moores University, Liverpool Science Park, 146 Brownlow Hill, Liverpool L3 5RF, UK*

<sup>6</sup>*Science and Technology Research Institute, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK*

<sup>7</sup>*School of Physics, University of New South Wales, Sydney, NSW 2052, Australia*

<sup>8</sup>*Armagh Observatory, College Hill, Armagh BT61 9DG, Northern Ireland*

<sup>9</sup>*East Asian Observatory, 660 North A’Ohoku Pl, Hilo, Hawaii, HI 96720, USA*

<sup>10</sup>*Wartburg College, Waverly, IA 50677, USA*

<sup>11</sup>*IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France*

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