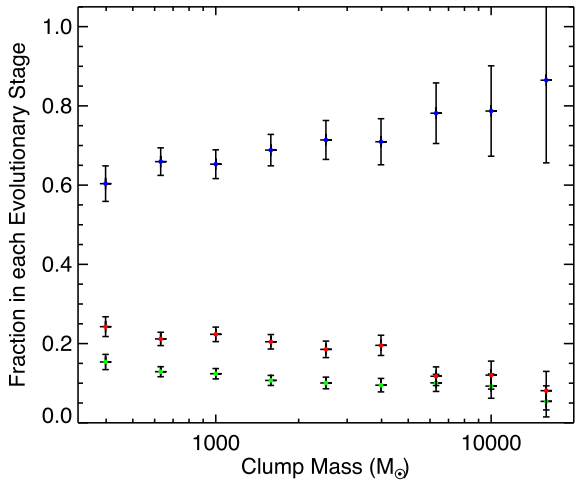

Figure 26. Fractions of clumps sorted by evolutionary stage. The green, red and blue circles indicate the fractions of quiescent, protostellar and mid-IR bright clumps (YSOs and MSF clumps), respectively, as a function of clump mass. The uncertainties are determined from Poisson counting statistics. 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 Hi-GAL 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 μm: 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 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: ($\log(t_{\rm HII}/{\rm yr} = (-0.13 \pm 0.16) \times \log(L/{\rm L}\odot) + (6.1 \pm 0.8)$). 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 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 5×10^4 yr for clumps with masses $\sim$1000 $M_\odot$, and decreases to 2×10^4 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.

$\log[M_{\rm clump}]$	log[luminosity]	Statistical lifetimes (10^5 yr)				
($M_\odot$)	($L_\odot$)	$t_{\rm HII}$	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

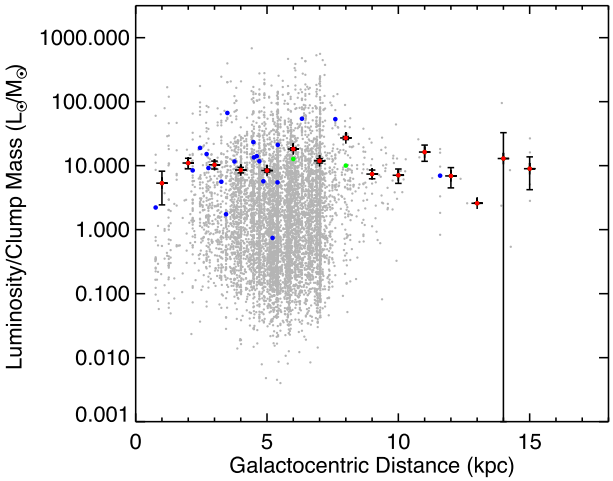

Figure 27. $L_{\rm bol}/M_{\rm clump}$ as a function of Galactocentric distance. The grey and blue circles show the distribution of all of the individual ATLASGAL clumps and the 30 most massive clusters, respectively. The red circles and the error bars show the mean values and the standard error of the mean. The two green circles show the $L_{\rm bol}/M_{\rm clump}$ 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 $L_{\rm MYSO}/M_{\rm cloud}$ was relatively flat for the inner 5 kpc, and although two significant peaks were detected at ∼6 and ∼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° (Eden et al. 2013, 2015) using GLIMPSE 8 μm, WISE 12 μm and 22 μm and Hi-GAL 70 μm 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 $L_{\rm bol}/M_{\rm clump}$ 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 $L_{\rm bol}/M_{\rm clump}$ 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 0.5–1 pc, while the clusters have typical sizes of a few tens of parsecs.

We observe two significant peaks at ∼6 and 8 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 $L_{\rm bol}/M_{\rm clump}$ 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 ∼6 and 8 kpc clearly shows that the exclusion of these two extreme star-forming regions results in a reduction of the $L_{\rm bol}/M_{\rm clump}$ ratio which brings these more in line with the $L_{\rm bol}/M_{\rm clump}$ ratios on either side. We therefore find that the $L_{\rm bol}/M_{\rm clump}$ 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 ($M_{\rm clump}$–$R_{\rm eff}$) 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, Σ(gas), of 1 and 0.05 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 ∼90 per cent). We also show the minimum threshold of ∼116–129 M$_\odot$ 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 ($d \leq 500$ 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 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 ($r_{\rm AB,C} = 0.85$) and they are located within a relatively narrow range of surface densities. The correlation coefficient for the complete sample shown in Fig. 28 is $r_{\rm AB,C} = 0.73$ 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±0.012, which is in excellent agreement with the slope determined for the MSF clumps reported in Paper III (1.67±0.025). The slope found for

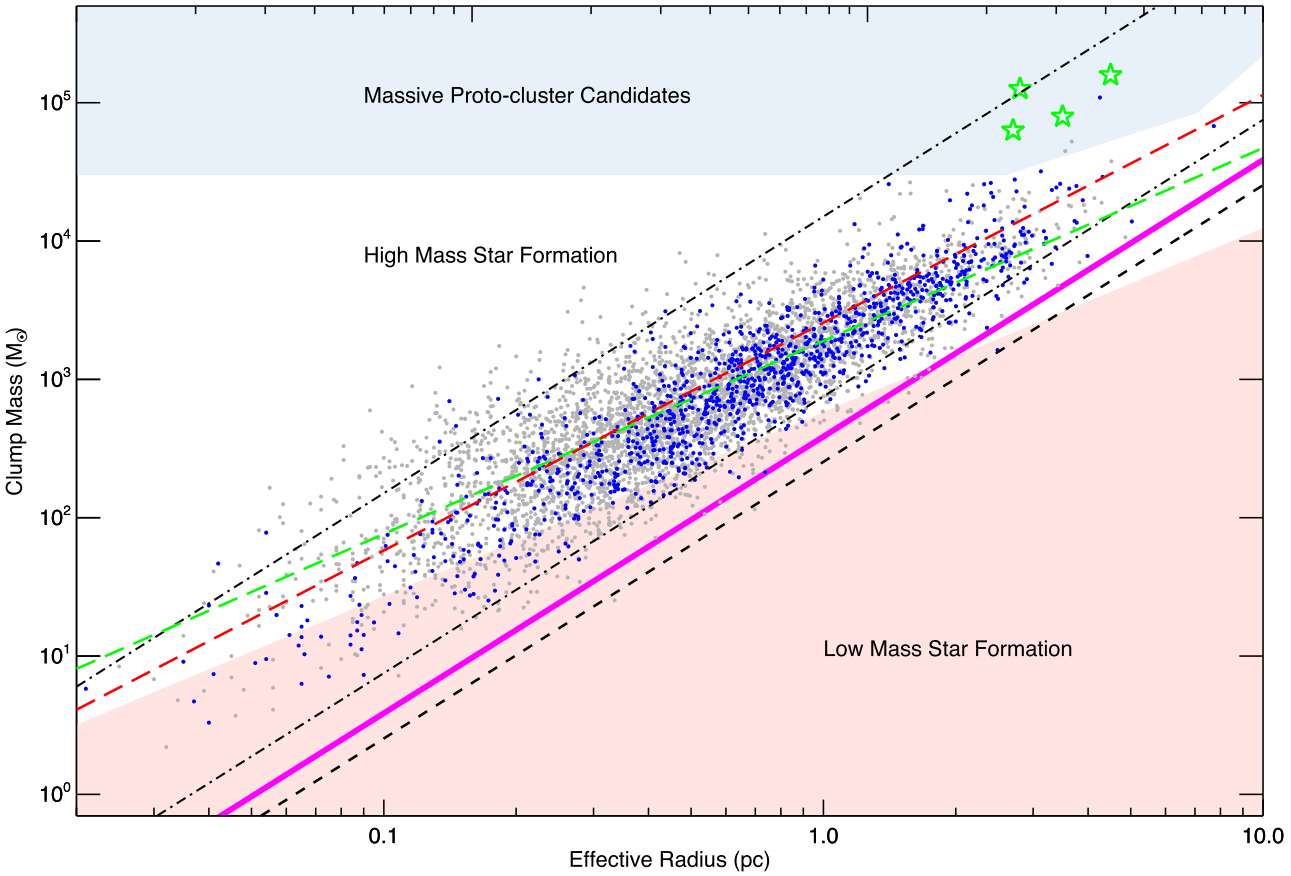

Figure 28. The mass–size relationship of all ATLASGAL clumps (grey) and those associated with MSF (blue). The green stars show the distribution of the MPC candidates found towards the GC (Longmore et al. 2012; Immer et al. 2012). The light red shaded region shows the part of the parameter space found to be devoid of massive star formation that satisfies the relationship $m(r) \leq 580\,\mathrm{M}_{\odot}\,(R_{\mathrm{eff}}/\mathrm{pc})^{1.33}$ (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 ($N_{\mathrm{H_2}} \sim 10^{22}\,\mathrm{cm}^{-2}$) and the upper and lower dot–dashed lines mark surface densities of 1 and 0.05 g cm^{-2}, respectively. The diagonal magenta band indicates the gas surface density [Σ(gas)] parameter space between 116 and 129 $\mathrm{M}_{\odot}\,\mathrm{pc}^{-2}$, 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±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 $3 \times 10^4\,\mathrm{M}_{\odot}$; 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$10 000 $\mathrm{M}_{\odot}$ (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 $3 \times 10^4\,\mathrm{M}_{\odot}$ 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 (∼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)	$R_{\rm eff}$ (pc)	$T_{\rm dust}$ (K)	$\log[M_{\rm clump}]$ ($M_\odot$)	$\log[N(H_2)]$ (cm^{-2})	$\log[L_{\rm bol}]$ ($L_\odot$)	$L_{\rm bol}/M_{\rm clump}$ ($L_\odot/M_\odot$)	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[a]
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[a]
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[a]
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[c]	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)[c]	–	Starless?	8.4	3.5	–	4.9	23.23	–	–	7
AGAL000.476−00.007 (e)[c]	–	Starless?	8.4	4.5	–	5.2	23.53	–	–	7
AGAL000.494+00.019 (f)[c]	–	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: [a]Identified by Ginsburg et al. (2012) as a massive clump, but the derived mass was slightly less than the 3×10^4 $M_\odot$ required for inclusion in their sample of MPC candidates.

[b]The distance to this source is uncertain and should be used with caution.

[c]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 ∼90 per cent of all massive clumps in the Galaxy and is sensitive to all clumps with a surface density $> 3600\,M_\odot$ 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 $L_{\rm bol}/M_{\rm clump}$ ratio ∼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 Wienen et al. (2015). Our independent assessment of the H I data has associated this source with

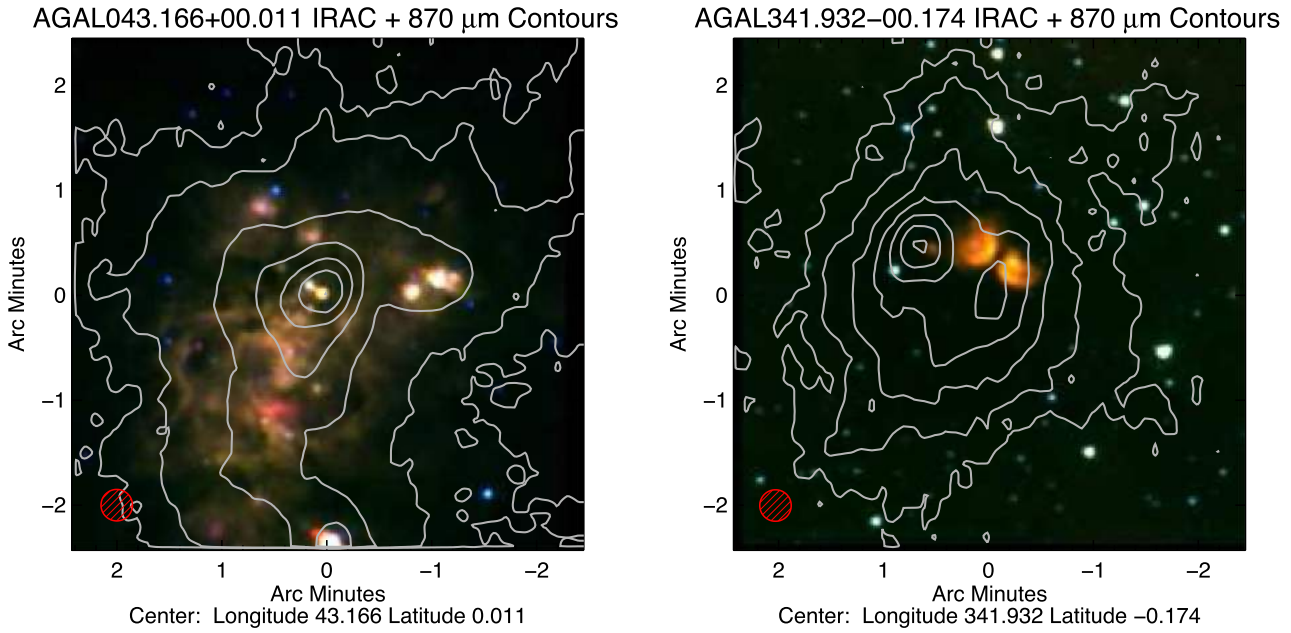

Figure 29. Three-colour mid-IR GLIMPSE image of the two regions associated with MPC candidates discussed in the text. The contours trace the dust emission detected by ATLASGAL and the red circle shown in the lower left corner of each map indicates the resolution of the ATLASGAL survey (∼19 arcsec). In 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 ∼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_2O 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 (∼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°, 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. $\log(M_{\rm clump}) > 4.48\,{\rm M}_\odot$), 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 ∼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

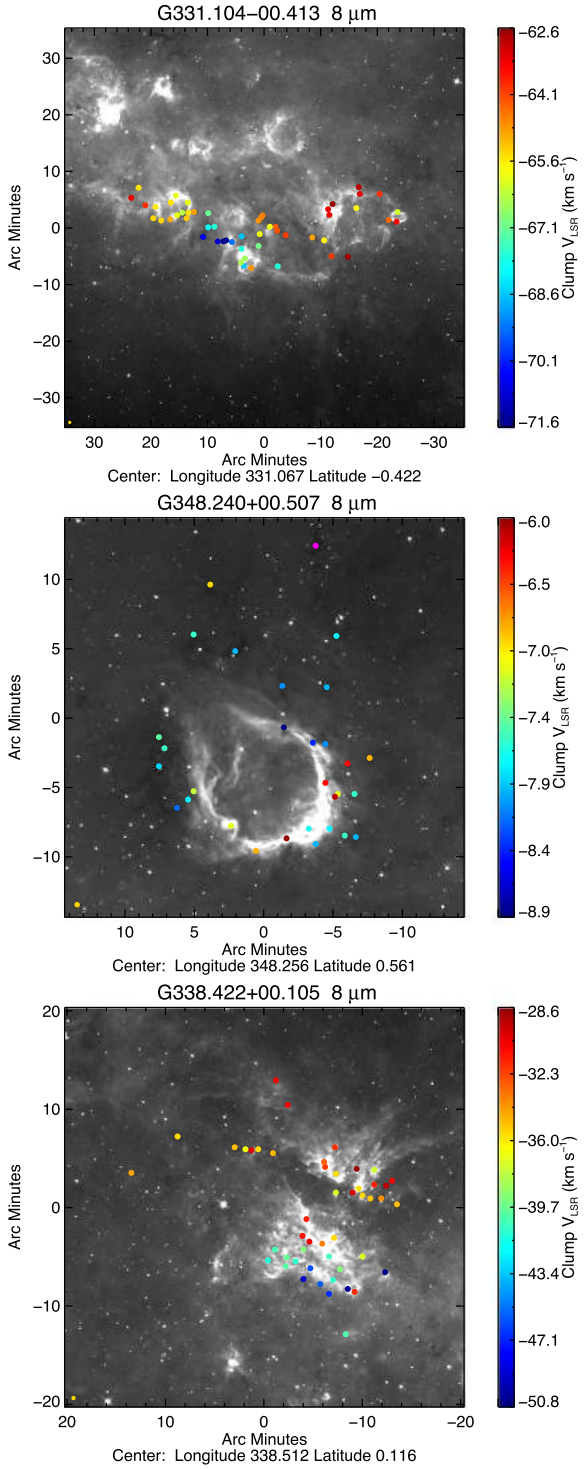

Figure 30. ATLASGAL clusters associated with MPC candidates identified by Contreras et al. (2017) and Longmore et al. (2017) identified by the friends-of-friends analysis. The background image is the 8 μm emission and filled circles show the positions of ATLASGAL CSC objects, the colours of which provide 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 ∼25, Longmore et al. (2012) estimated the YMC formation rate to be one every ∼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 NH_3 linewidth normalized by the square root of the clump radius and L_{bol}/M_{clump} ratio in the upper panel of Fig. 31. The partial-Spearman test reveals a moderate correlation between these parameters ($r_{AB,C} = 0.12$, 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:

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

where R_{eff} is the effective radius of the clumps, σ_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 $\alpha_{vir} \propto M^{-0.4}$: 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

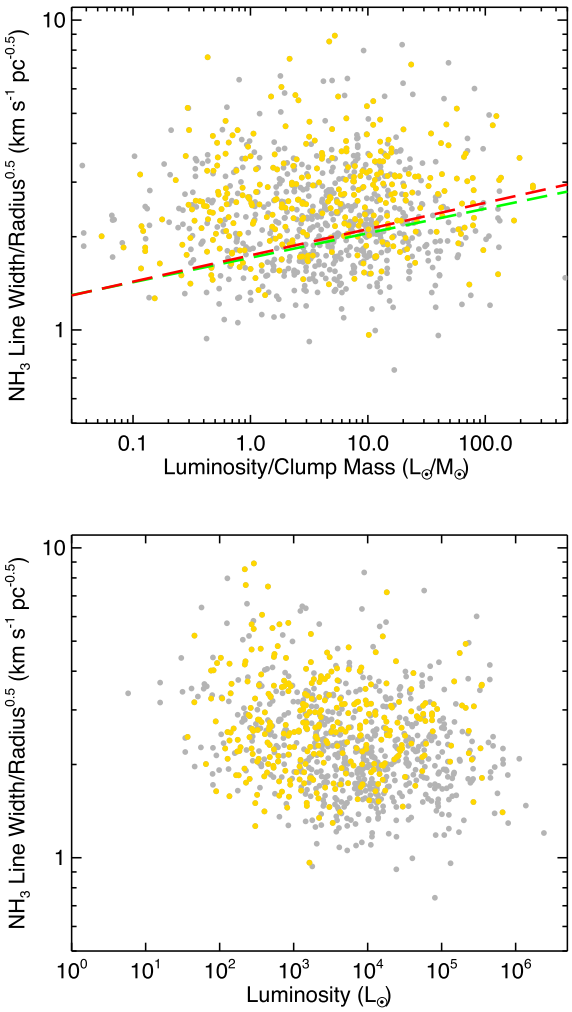

Figure 31. NH_3 (1,1) linewidth normalized by the square root of the clump radius as a function of the L_{bol}/M_{clump} ratio (upper panel) and as a function of the bolometric luminosity (lower panel). The grey circles show the distributions of all clumps, while the yellow shows the distribution of 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.

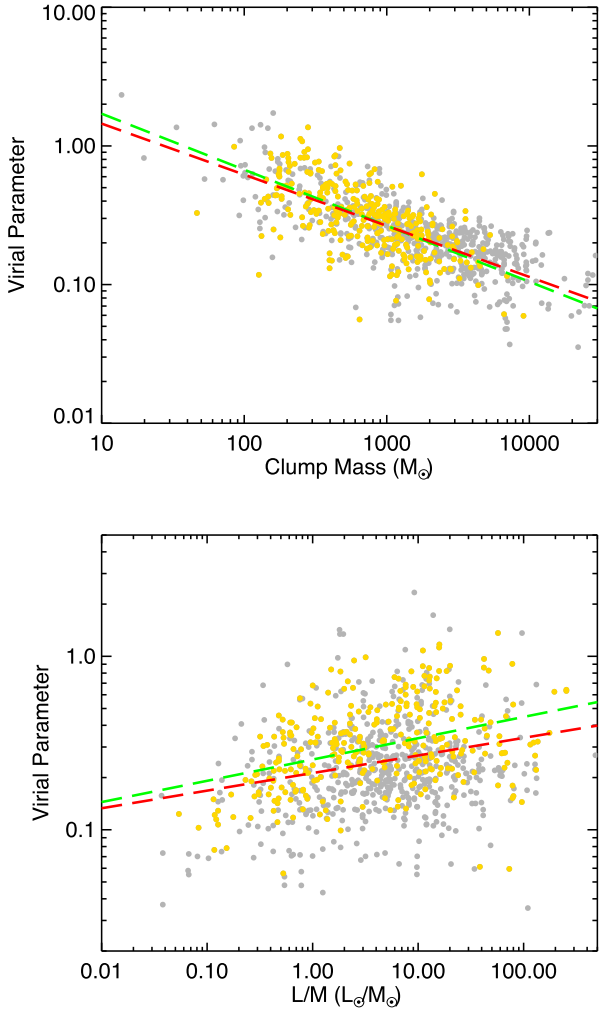

Figure 32. Virial parameter as a function of clump mass and L_{bol}/M_{clump} ratio (upper and lower panels, respectively). The long-dashed red line shows the result of a linear power-law fit to the whole sample of associated clumps, while the green line shows the fit to the distance-limited sample.

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_{bol}/M_{clump} ratio where the α_{vir} increase by a factor of 2–3. The virial mass is $M_{vir} \propto R\Delta v^2$, 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 α_{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_{bol}/M_{clump} ratio plot appears very similar to our peak column density versus L_{bol}/M_{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_{bol}/M_{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 ∼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

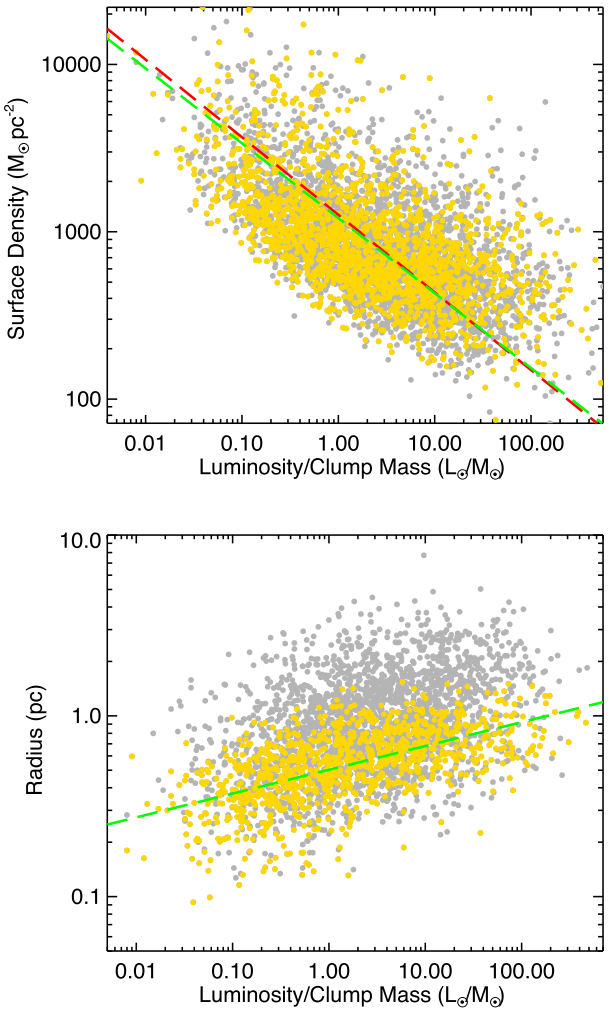

Figure 33. Upper panel: the mass surface density as a function of the $L_{\rm bol}/M_{\rm clump}$ ratio. Lower panel: clump radius as a function of $L_{\rm bol}/M_{\rm clump}$ ratio. The grey circles show the distributions of all clumps, while the yellow shows the distribution of the distance-limited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

$L_{\rm bol}/M_{\rm clump}$ 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 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 ∼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 ∼20 per cent of the total luminosity found in the inner part of the Galactic disc.

We previously used the $L_{\rm bol}/M_{\rm clump}$ 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	ℓ (°)	b (°)	$v_{\rm lsr}$ (km s^{-1})	Distance (kpc)	T (K)	$\log(L_{\rm bol})$ ($L_\odot$)	$L_{\rm bol}$ (per cent)	$\log(M_{\rm clump})$ ($M_\odot$)	$M_{\rm clump}$ (per cent)	$L_{\rm bol}/M_{\rm clump}$ ($L_\odot/M_\odot$)
G030.650−00.015	W43	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
G337.651−00.048		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 $L_{\rm bol}/M_{\rm clump}$ 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 $L_{\rm bol}/M_{\rm clump}$ 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 3σ from the mean and W49 is over 3σ away from the mean. These three clusters have $L_{\rm bol}/M_{\rm clump}$ 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 $L_{\rm bol}/M_{\rm clump}$ 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 $L_{\rm bol}/M_{\rm clump}$ 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 ∼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° 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 ($v_{\rm lsr}$) for 7807 clumps (corresponding to ∼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. ℓ, 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 (∼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 $L_{\rm bol}/M_{\rm clump}$ 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 $\log(N) \sim 23\,{\rm cm}^{-2}$, 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 ∼10 and ∼20 per cent for clumps up to $4 \times 10^3\,{\rm M}_\odot$ but drop to ∼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 (∼90 per cent found within $10\,{\rm km\,s}^{-1}$ 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 ∼26 pc: we find no difference between the different subsamples.

(vii) We find little evidence for variations in the $L_{\rm bol}/M_{\rm clump}$ ratio or mean clump temperature as a function of Galactocentric distance (∼10 $L_\odot/M_\odot$ and ∼20 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 $L_{\rm bol}/M_{\rm clump}$ 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 $L_{\rm bol}/M_{\rm clump}$ 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 $L_{\rm bol}/M_{\rm clump}$ ratios $<25\,{\rm L}_\odot/{\rm M}_\odot$, however, three clusters have significantly higher values ($>50\,{\rm L}_\odot/{\rm M}_\odot$) 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 $L_{\rm bol}/M_{\rm clump}$ 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 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.-Å., 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
Compiégne 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., Broguière D., 2007, A&A, 476, 1243
Motte F., Bontemps S., Louvet F., 2017, preprint (arXiv: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., Huldtgren-White M., 2004, A&A, 414, 1017
Tothill 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 (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}CO AND C^{18}O OBSERVATIONS

A1 CO observations and data reduction

We targeted the ^{13}CO and C^{18}O (2–1) transitions at ∼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 ∼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 PWV = 2.9 ± 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 ± 200 km s^{-1} of the rest frequency of the transition to obtain

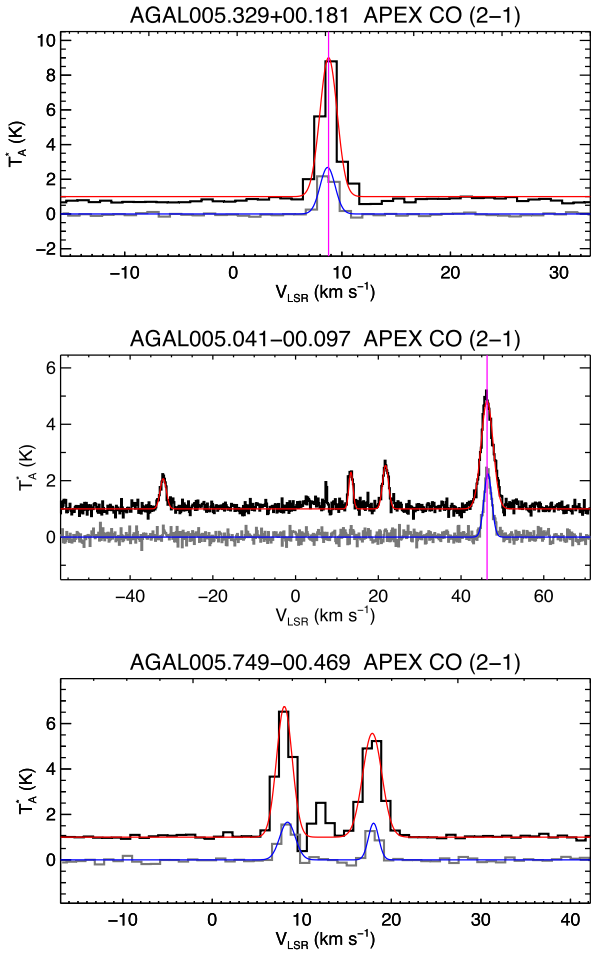

Figure A1. Example APEX CO (2–1) spectra towards three sources. The ^{13}CO spectra is offset from the C^{18}O spectra by 1 K and the fits are shown by the red and blue lines, respectively. The vertical dashed lines shown in the upper two panels indicates the preferred source velocity. The two components 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[4] from within our PYTHON code through the PYCLASS module of the PYGILDAS[5] package. The resulting spectra are then smoothed to a velocity resolution of $\sim 1\,\mathrm{km\,s^{-1}}$, resulting in a mean noise level of $\sigma = 95 \pm 50$ 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 3σ 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.

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	$0.1\,\mathrm{km\,s^{-1}}$
Smoothed spectral resolution	$1\,\mathrm{km\,s^{-1}}$
Medium noise (T_A^*)	$\sim$95 mK channel^{-1}
Medium system temperatures (T_sys)	$\sim$220 K
Integration time (on-source)	1–4 min

Fig. A1 presents a few examples of the spectra obtained. We have not converted these to the main beam temperature scale (T_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}CO and C^{18}O components to estimate their optical depths and used this as an additional constraint – if the ^{13}CO components have similar integrated intensities but one has a significantly higher optical depth (effectively selecting the component with the lowest peak ^{13}CO/C^{18}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}CO and 1271 C^{18}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.

[4] http://www.iram.fr/IRAMFR/GILDAS
[5] 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 ⋆ to the source name. We give the noise per channel of the ^{13}CO in Column 2, however, the ^{13}CO and C^{18}O noise agree to within a few per cent.

ATLASGAL		^{13}CO (2–1)				C^{18}O (2–1)				Optical
CSC name	RMS (mK)	v_{lsr} (km s^{-1})	Peak (K)	Width (km s^{-1})	Intensity (K km s^{-1})	v_{lsr} (km s^{-1})	Peak (K)	Width (km s^{-1})	Intensity (K km s^{-1})	depth
AGAL005.001+00.086⋆	61	2.1	5.8	1.38	20.2	2.0	1.8	0.93	4.2	0.35
AGAL005.041−00.097*	118	46.3	3.8	1.53	14.4	46.5	2.1	1.04	5.5	0.82
AGAL005.049−00.192*	72	6.1	3.2	1.53	12.4	6.1	0.9	0.97	2.2	0.32
AGAL005.139−00.097*	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*	70	8.0	3.4	2.21	18.6	8.6	0.6	1.09	1.7	0.16
AGAL005.202−00.036*	68	0.6	4.8	1.99	23.7	0.6	1.4	1.78	6.1	0.32
AGAL005.329+00.181*	69	8.8	8.0	0.79	15.9	8.7	2.7	0.69	4.6	0.40
AGAL005.371+00.319*	61	18.7	5.7	1.51	21.4	18.7	1.7	1.31	5.6	0.35
AGAL005.387−00.551*	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*	164	140.8	1.1	7.93	22.8	–	–	–	–	–
AGAL005.399−00.237*	53	4.8	1.9	1.34	6.3	5.0	0.4	1.01	1.1	0.24
AGAL005.409−00.302*	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 (∼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_{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 ± 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 (∼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

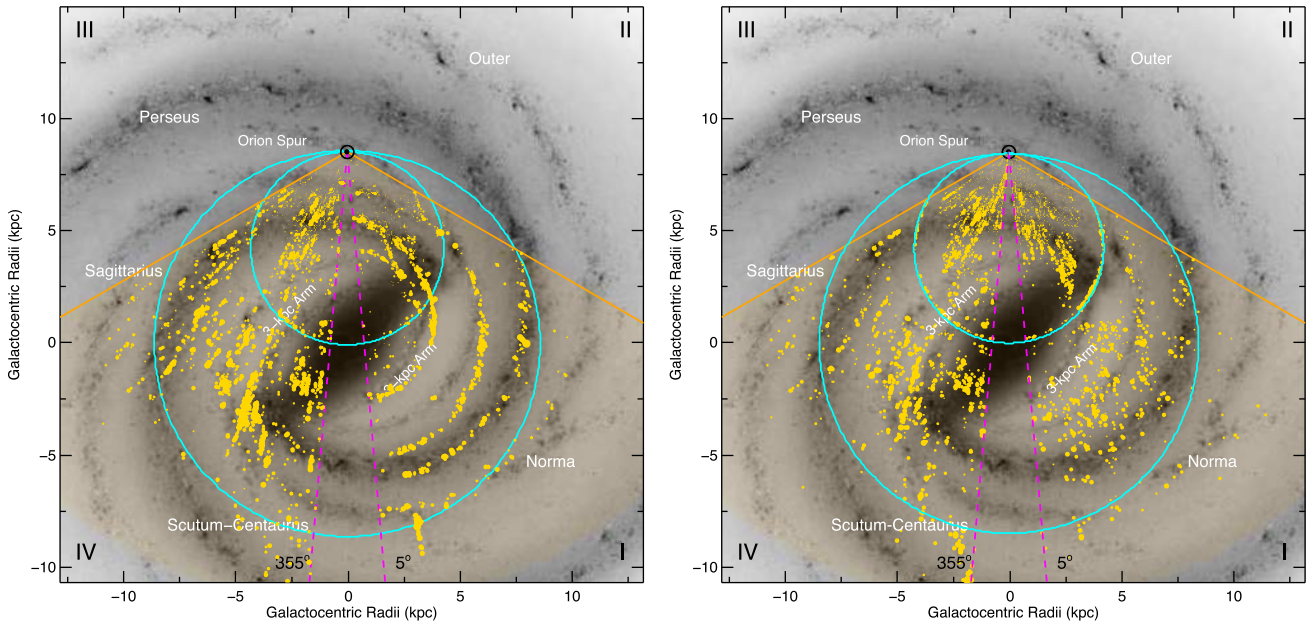

Figure B1. As Fig. 7. Left-hand panel: distribution of all ATLASGAL sources given by the Bayesian maximum likelihood tool developed by Reid et al. (2016). Right-hand panel: distribution of all ATLASGAL sources determined using the Reid et al. (2014) Galactic rotation curve.

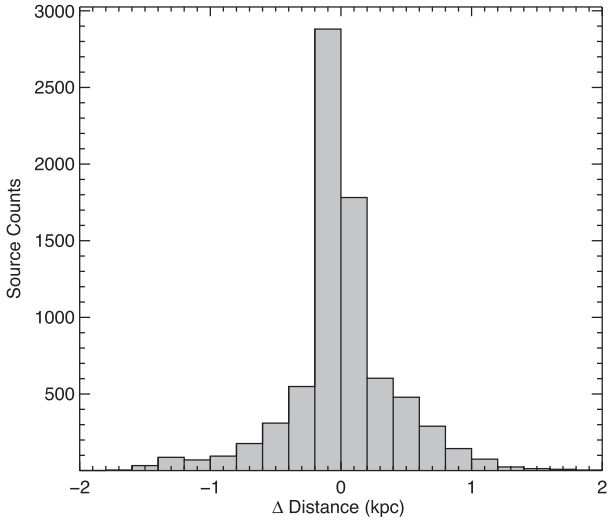

Figure B2. Histogram showing the difference between the distance determined from using the Bayesian method presented by Reid et al. (2016) and the near/far distances determined using the Reid et al. (2014) rotation curve (note that the *x*-axis has been truncated). 98.4 per cent of the sample have velocity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

KDAsaffect 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 (∼300).

(iii) We can exclude sources located close to the tangent velocity ($|v_{\rm clump} - v_{\rm tangent}| < 10\,{\rm km\,s^{-1}}$) from the KDA tests, as the difference between the near and far distances becomes smaller than their associated uncertainties: this eliminates ∼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 ∼30 pc (e.g. Reed 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 (∼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 Peretto & Fuller (2009) to identify sources associated with IRDCs. This catalogue does not include the inner 10° 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).

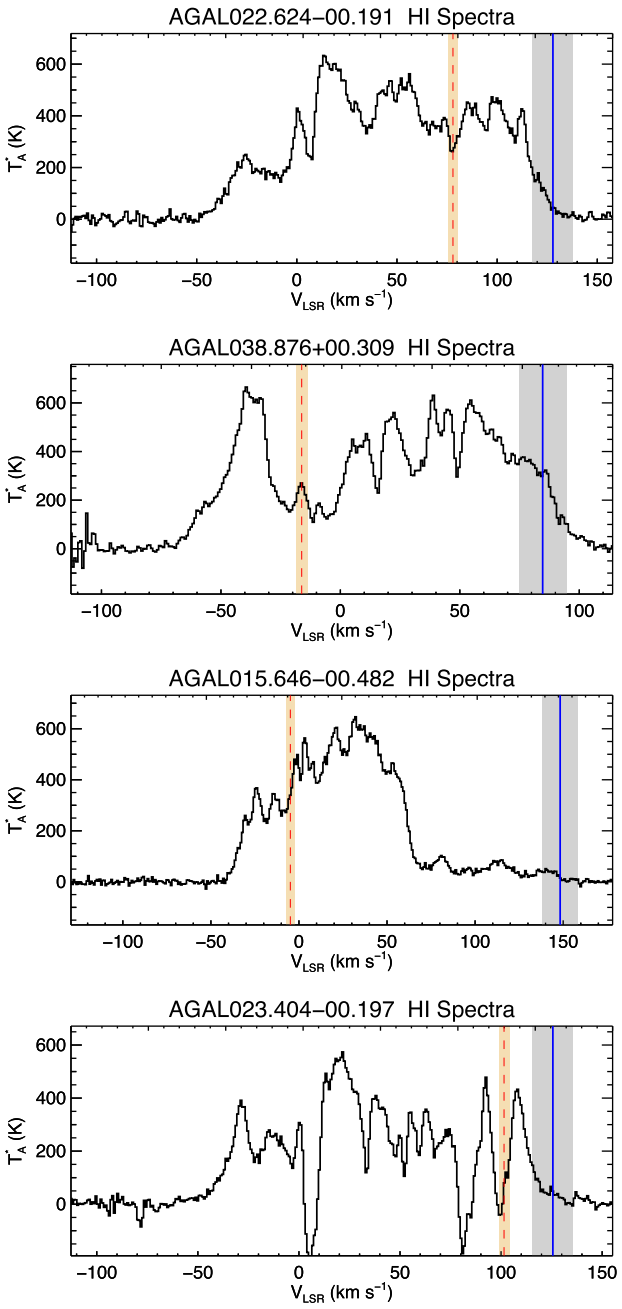

Figure B3. Example H I spectra extracted towards dense clumps identified from the ATLASGAL survey. The velocity of the tangent position is determined from a fit to H I data (e.g. McClure-Griffiths & Dickey 2007) and is indicated by the blue line; the grey shaded region covers a velocity range of ±10 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.

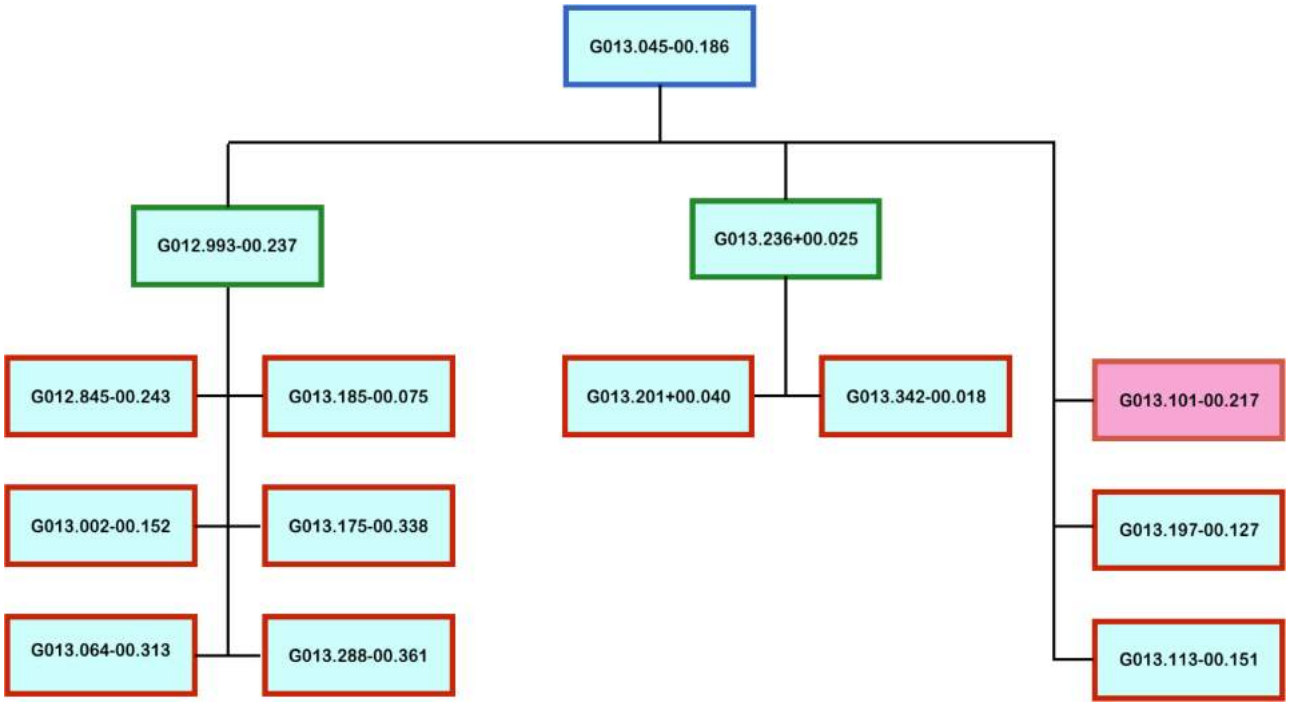

Figure B4. Schematic diagram showing the hierarchical structure of a cluster identified from the friends-of-friends analysis. At the top of the diagram (outlined in blue), we have the structure identified by using the largest set of search parameters (i.e. $R_{\rm fof} < 8$ arcmin and $\Delta v_{\rm fof} < 8\,{\rm km\,s^{-1}}$), the second row (outlined in green) shows the subregions identified by the using the intermediate search parameters (i.e. $R_{\rm fof} < 6$ arcmin and $\Delta v_{\rm fof} < 6\,{\rm km\,s^{-1}}$), while the remaining rows show the smallest subregions identified using the smallest search parameters (i.e. $R_{\rm fof} < 4$ arcmin and $\Delta v_{\rm fof} < 4\,{\rm km\,s^{-1}}$); 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. $|v_{\rm lsr}| < 10\,{\rm km\,s^{-1}}$) is a little more complicated because we also need to consider the additional uncertainty due to streaming motions in the Galaxy ($\pm$ 10 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 (∼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 ATLAGAL 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 ∼77 per cent and 23 per cent, respectively. This is similar to the findings of other studies (e.g. Eden et al. 2012 assigned ∼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 ∼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 ∼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 ∼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 ∼7000 ATLASGAL clumps; however, there are still a significant number (∼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 ∼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 ∼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 ($<5\,\mathrm{km\,s^{-1}}$), this can result in large differences in their kinematic distances ($\sim$0.5 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 ($R_{\rm fof}$) of 4 arcmin and velocity dispersion of ($\Delta v_{\rm fof}$) $4\,\mathrm{km\,s^{-1}}$ and successively increasing the search parameters by 2 arcmin and $2\,\mathrm{km\,s^{-1}}$. 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. ℓbv); 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. $R_{\rm fof} < 8$ arcmin and $\Delta v_{\rm fof} < 8\,\mathrm{km\,s^{-1}}$) 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 pc at a distance of 15 kpc, which is the typical cloud radius (Miville-Deschênes, Murray & Lee 2017), and FWHM velocity dispersion of $10\,\mathrm{km\,s^{-1}}$, 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 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.

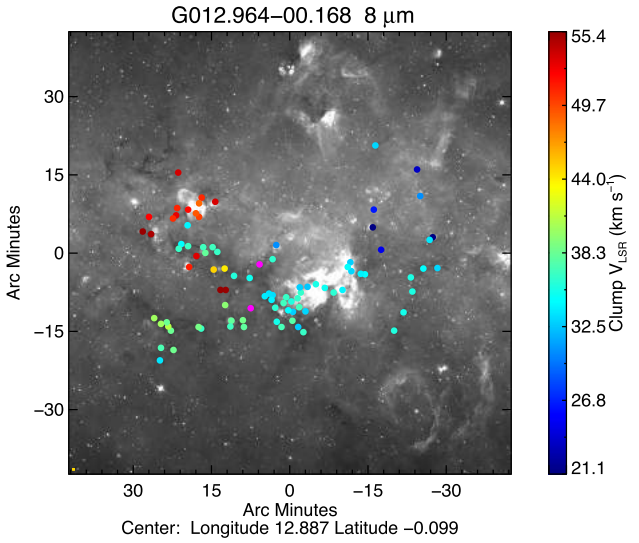

Figure C1. Example of clustering analysis towards W33 complex. For description of image refer to Fig. 6. The hierarchical structure of W33 is used as an example of the association method discussed in Appendix C1 and shown in the flow chart presented in Fig. B4.

(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

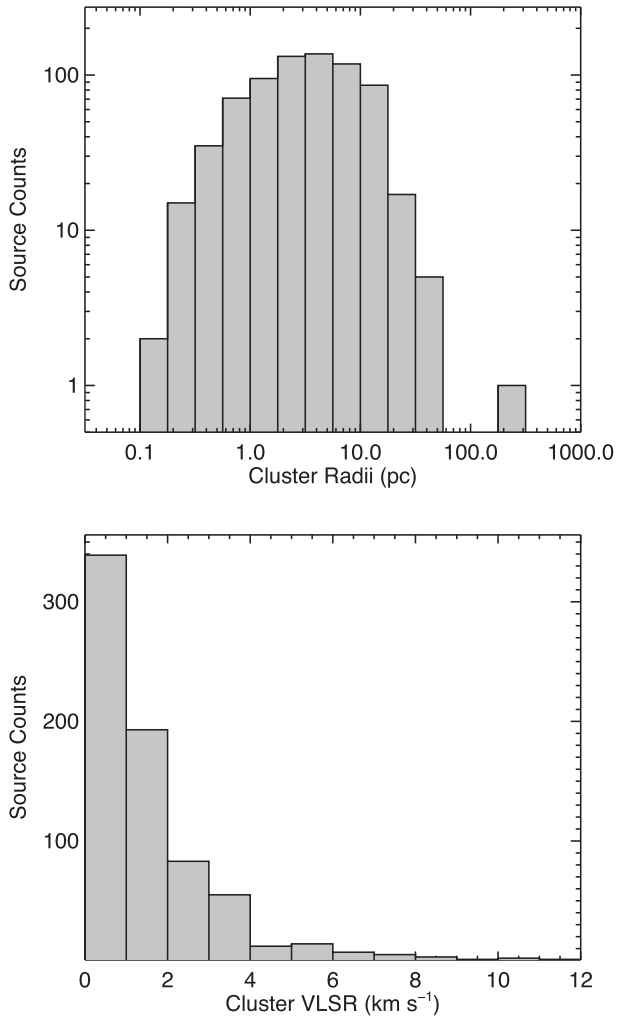

Figure C2. Distribution of cluster sizes and velocity dispersions. The bin size used in the upper panel is 0.25 dex and in the lower panel, it is 1 km s^{-1}.

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 km s^{-1} and in line with the velocity dispersions reported for Galactic and extragalactic GMCs (5–20 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. $r_{\rm search} = 19.2$ arcsec) of the peak emission found in the 250, 350 or 870 μm band, using the first band where no saturation is found[6] 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 × 5 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 1σ, 2σ and 3σ 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 μm *Herschel* bands, and an uncertainty of 50 per cent for the 500 μm *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 μm 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 arcmin × 5 arcmin image, where pixels within a beam around local maxima are blanked. The noise level is then determined over

[6] Pixels were the detector suffered from saturation are found as NaN values in the reduced maps.

the filtered image, where the 1σ 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 μm, 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 μm. 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 μm 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 (Compiégne 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 μm as an upper limit in the fitting process. When no emission is found at the 70 μm 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 mJy beam^{-1}),[7] 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 ∼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 ± 0.1 K with a standard deviation of 2.8 K. In the lower panel of Fig. D1, we show the correlation between the kinetic dust temperatures derived from the ammonia (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 ± 0.14 K with a standard deviation is 4.8 K).

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

[7] herschel.esac.esa.int/Docs/PMODE/html/ch02s03.html

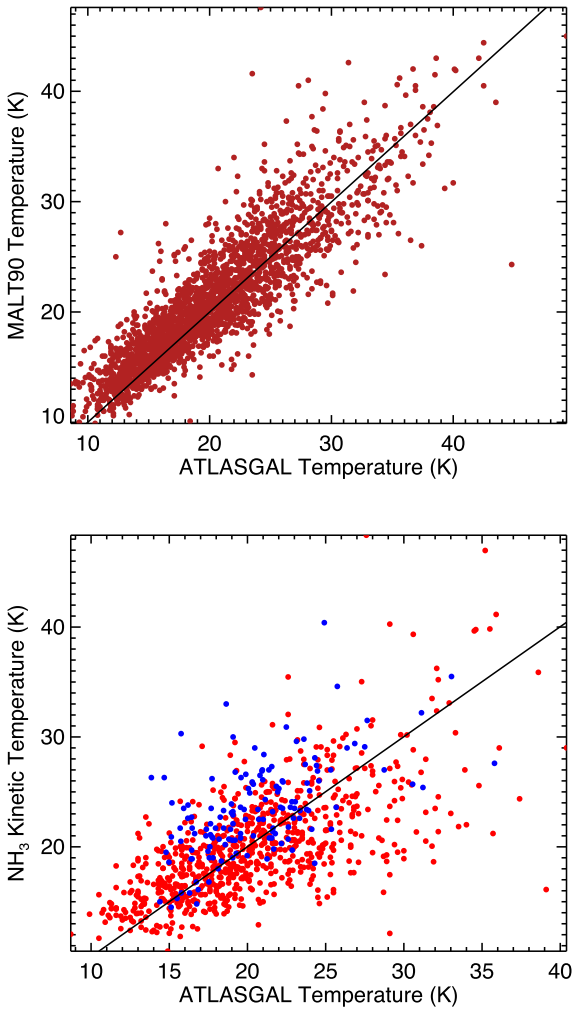

Figure D1. Upper panel: comparison between dust temperatures derived here and those determined by the MALT90 team (Guzmán et al. 2015). This subsample of ∼2500 includes a significant fraction of all of the sources observed as part of the MALT90 survey (a total of ∼3000 ATLASGAL clumps were observed as 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 ∼1500 clumps from 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 1σ 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-μm band. Although there is a significant amount of scatter for individual sources (∼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 (∼98 per cent of the full CSC sample). 2586 of these sources

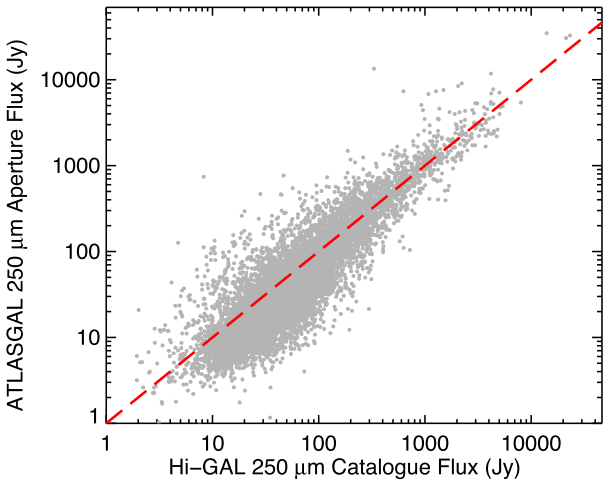

Figure D2. Comparison of flux densities obtained from the Hi-GAL CSC with the fluxes obtained through aperture photometry in this work. The dashed red line indicates equality between the two sets of measurements.

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

[1]*Centre for Astrophysics and Planetary Science, University of Kent, Canterbury CT2 7NH, UK*
[2]*Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-53121 Bonn, Germany*
[3]*INAF – Istituto di Radioastronomia and Italian ALMA Regional Centre, Via P. Gobetti 101, I-40129 Bologna, Italy*
[4]*INAF – Osservatorio Astronomico di Cagliari, Via della Scienza 5, I-09047 Selargius (CA), Italy*
[5]*Astrophysics Research Institute, Liverpool John Moores University, Liverpool Science Park, 146 Brownlow Hill, Liverpool L3 5RF, UK*
[6]*Science and Technology Research Institute, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK*
[7]*School of Physics, University of New South Wales, Sydney, NSW 2052, Australia*
[8]*Armagh Observatory, College Hill, Armagh BT61 9DG, Northern Ireland*
[9]*East Asian Observatory, 660 North A'Ohoku Pl, Hilo, Hawaii, HI 96720, USA*
[10]*Wartburg College, Waverly, IA 50677, USA*
[11]*IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France*

This paper has been typeset from a TEX/LATEX file prepared by the author.