

ATLASGAL – properties of compact H II regions and their natal clumps[★]

J. S. Urquhart,^{1†} M. A. Thompson,² T. J. T. Moore,³ C. R. Purcell,⁴ M. G. Hoare,⁴
F. Schuller,⁵ F. Wyrowski,¹ T. Csengeri,¹ K. M. Menten,¹ S. L. Lumsden,⁴ S. Kurtz,⁶
C. M. Walmsley,^{7,8} L. Bronfman,⁹ L. K. Morgan,³ D. J. Eden³ and D. Russeil¹⁰

¹Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-53121 Bonn, Germany

²Centre for Astrophysics Research, Science and Technology Research Institute, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK

³Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK

⁴School of Physics and Astrophysics, University of Leeds, Leeds LS2 9JT, UK

⁵European Southern Observatory, Alonso de Cordova 3107, Vitacura, Santiago, Chile

⁶Centro de Radioastronomía y Astrofísica, Universidad Nacional Autónoma de México, Antigua Carretera a Ptzcuaro # 8701 Morelia, 58089 Michoacán, Mexico

⁷Osservatorio Astrofisico di Arcetri, Largo E. Fermi, 5, I-50125 Firenze, Italy

⁸Dublin Institute for Advanced Studies, Burlington Road 10, Dublin 4, Ireland

⁹Departamento de Astronomía, Universidad de Chile, Casilla 36-D, Santiago, Chile

¹⁰Aix Marseille Université, CNRS, LAM (Laboratoire d'Astrophysique de Marseille) UMR 7326, F-13388 Marseille, France

Accepted 2013 July 15. Received 2013 July 15; in original form 2013 March 20

ABSTRACT

We present a complete sample of molecular clumps containing compact and ultracompact H II (UC H II) regions between $\ell = 10^\circ$ and 60° and $|b| < 1^\circ$, identified by combining the APEX Telescope Large Area Survey of the Galaxy submm and CORNISH radio continuum surveys with visual examination of archival infrared data. Our sample is complete to optically thin, compact and UC H II regions driven by a zero-age main-sequence star of spectral type B0 or earlier embedded within a $1000 M_\odot$ clump. In total we identify 213 compact and UC H II regions, associated with 170 clumps. Unambiguous kinematic distances are derived for these clumps and used to estimate their masses and physical sizes, as well as the Lyman continuum fluxes and sizes of their embedded H II regions. We find a clear lower envelope for the surface density of molecular clumps hosting massive star formation of 0.05 g cm^{-2} , which is consistent with a similar sample of clumps associated with 6.7 GHz masers. The mass of the most massive embedded stars is closely correlated with the mass of their natal clump. Young B stars appear to be significantly more luminous in the ultraviolet than predicted by current stellar atmosphere models. The properties of clumps associated with compact and UC H II regions are very similar to those associated with 6.7 GHz methanol masers and we speculate that there is little evolution in the structure of the molecular clumps between these two phases. Finally, we identify a significant peak in the surface density of compact and UC H II-regions associated with the W49A star-forming complex, noting that this complex is truly one of the most massive and intense regions of star formation in the Galaxy.

Key words: Stars: early-type – stars: formation – ISM: clouds – ISM: H II regions – Galaxy: structure – radio continuum: ISM.

1 INTRODUCTION

Ultracompact H II (hereafter UC H II) regions represent one of the earliest stages in the development of an H II region, where a massive star (or stars) has begun to ionize the surrounding gas and produce a small (diameter, D , $< 0.1 \text{ pc}$) photoionized bubble of gas embedded within a dense molecular cloud clump. UC H II are particularly important tracers of *recent* massive star formation that are readily

[★]The full version of Tables 3 and 4 are 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/> and in the electronic version of the journal, while Figs 11 and 12 are only available in the electronic version of the journal.

†E-mail: j.s.urquhart@gmail.com

detectable across most of the Milky Way via their centimetre-wavelength free–free emission or their strong far-infrared emission (e.g. Wood & Churchwell 1989a; Becker et al. 1990; Giveon et al. 2005; Hoare et al. 2012). As such, UC H II regions give us a reliable snapshot of massive star formation activity within the last $\sim 10^5$ yr (e.g. Davies et al. 2011; Mottram et al. 2011b), enabling a census of massive star formation on the scale of individual complexes throughout the entire Galaxy. UC H II regions are a particularly important signpost of the dense conditions within molecular clouds that give rise to massive star formation, and also as a direct tracer of the luminosity of the recently formed stars via their free–free emission.

Most UC H II regions have been identified by targeted radio interferometry of objects selected by their mid- or far-infrared colours (e.g. Wood & Churchwell 1989a; Kurtz, Churchwell & Wood 1994; Walsh et al. 1998; Urquhart et al. 2007; Urquhart, Morgan & Thompson 2009a). The drawback of such targeted studies is that their infrared selection techniques can be contaminated by other IR-bright objects such as planetary nebulae (PNe) and intermediate-mass young stellar objects (YSOs) (Ramesh & Sridharan 1997), introducing biases into conclusions based on larger samples drawn solely from infrared catalogues. Wide-area radio interferometric surveys remove this constraint and have also been utilized to identify UC H II regions (Becker et al. 1990, 1994; Zoonematkermani et al. 1990; Giveon et al. 2005; Murphy et al. 2010; Hindson et al. 2012). Most of these surveys were carried out in snapshot mode, which limits the uv -coverage of the observations. It is thus difficult to unambiguously identify UC H II regions due to the limited range of angular scales traced by snapshot observations; UC sources detected in snapshot surveys may simply be bright and compact components of more extended emission (Kurtz et al. 1999; Kim & Koo 2001; Ellingsen, Shabala & Kurtz 2005).

Another compounding problem in the identification of UC H II regions is the fact that at their typical 5 GHz flux densities (a few mJy to tens mJy) the source counts are dominated by background radio galaxies (Anglada et al. 1998). In order to conclusively identify UC H II regions one must also show that the UC radio sources have spectra consistent with thermal free–free continuum, are clearly associated with thermal infrared emission from hot dust heated within the ionized nebula or polycyclic aromatic hydrocarbon (PAH) emission from the photodominated region surrounding the nebula, or show signs of being embedded within molecular cloud clumps. Existing studies have mainly used the *IRAS* or *MSX* surveys to fulfil the infrared criterion (Walsh et al. 1998; Giveon et al. 2005). The main drawback with these surveys is their low angular resolution compared to the \sim arcsecond resolution of the radio images, which makes it difficult to unambiguously determine the correct infrared counterpart. While the angular resolution of the ground-based millimetre and submillimetre (submm) observations used to trace the star-forming molecular clumps is higher, wide-area surveys have not come into being until relatively recently. This means that the number of UC H II hosting clumps studied is limited by targeted studies to typically a few tens to a hundred objects (see e.g. Thompson et al. 2006).

We take advantage of three recent Galactic plane surveys to address these issues and compile a robust well-selected sample of molecular clumps associated with deeply embedded H II regions. The surveys are the CORNISH 5 GHz survey (Hoare et al. 2012; Purcell et al. 2013), the GLIMPSE mid-infrared survey (Benjamin et al. 2003) and the APEX Telescope Large Area Survey of the Galaxy (ATLASGAL) 870 μ m survey (Schuller et al. 2009). Each of these surveys covers a common field ($10^\circ \leq \ell \leq 60^\circ$, $|b| \leq 1^\circ$), with

matched angular resolution between the 5 GHz radio and GLIMPSE mid-infrared data sets (~ 2 arcsec). By combining these three surveys we are able to use multiwavelength identification techniques to identify H II regions and confirm their embedded nature via the coincidence of radio and mid-infrared emission (e.g. Hoare et al. 2007) and submm calorimetry (Thompson et al. 2006). Another concern is the possibility of contamination from stellar winds of massive young stellar objects (MYSOs). However, the radio continuum emission arising from these winds is relatively weak ($S_\nu < \text{few mJy at 1 kpc}$; e.g. Tofani et al. 1995; Hoare 2002) and it is therefore unlikely that CORNISH is sensitive enough for this to be an issue. We also determine unambiguous kinematic distances for the majority of our sample. This results in a complete catalogue of molecular clumps with embedded H II regions over the common GLIMPSE, CORNISH and ATLASGAL survey region. This sample will include a mixture of UC ($D \lesssim 0.1$ pc and particle densities, $n, \gtrsim 10^4$ cm $^{-3}$) and compact ($D \lesssim 0.5$ pc and $n \gtrsim 5 \times 10^3$ cm $^{-3}$) H II regions (Kurtz 2005).

This catalogue is used to investigate the physical nature of the UC H II region hosting clumps, the mass–size relationship for massive star-forming clumps (Kauffmann & Pillai 2010), the Galactic distribution of UC H II regions and their statistical lifetimes. Finally, we explore correlations between the UC H II region properties and those of their embedding clumps in order to give insights into the accretion history and star formation efficiency (SFE).

We describe the individual surveys used in this study in Section 2, the multiwavelength identification techniques and observed properties of the UC H II hosting clumps in Section 3 and their kinematic properties in Section 4. Section 5 details the derived kinematic distances and the H I self-absorption techniques that we have used to break the near–far distance ambiguity. In Sections 6 and 7, we derive the physical properties of the molecular clumps and UC H II regions, respectively. We discuss these properties in the context of current star formation thresholds and with the properties of methanol-maser associated clumps discussed in Urquhart et al. (2013, hereafter Paper I). Finally, in Section 9, we summarize our findings and discuss their implications.

This paper is the second in a series of three papers focusing on the properties of molecular clumps from the ATLASGAL survey that are associated with massive star formation. The first paper in the series investigates the properties of clumps associated with 6.7 GHz methanol masers (Paper I) identified by the Methanol Multibeam (MMB; Caswell et al. 2010) survey. In this paper, we deal with UC H II region hosting clumps, and the third will investigate the properties of clumps associated with a sample of mid-infrared-selected massive YSOs (Urquhart et al., in preparation). We also note that there are two forthcoming CORNISH papers focusing on the radio-derived parameters of the UC H II regions (Purcell et al., in preparation) and their morphologies (Hoare et al., in preparation).

2 SURVEY DESCRIPTIONS

2.1 ATLASGAL Survey

The ATLASGAL (Schuller et al. 2009) is the first systematic survey of the inner Galactic plane in the submm band. The survey was carried out with the Large APEX Bolometer Camera (Siringo et al. 2009), an array of 295 bolometers observing at 870 μ m (345 GHz). At this wavelength the APEX Telescope has a full-width at half-maximum (FWHM) beam size of 19.2 arcsec. The initial survey region covered a Galactic longitude region of $300^\circ < \ell < 60^\circ$ and $|b| < 1.5^\circ$, but this was extended to include $280^\circ < \ell < 300^\circ$;

however, the latitude range was shifted to $-2^\circ < b < 1^\circ$ to take account of the Galactic warp in this region of the plane.

Contreras et al. (2013) produced a compact source catalogue for the central part of the survey region (i.e. $330^\circ < \ell < 21^\circ$) using the source extraction algorithm `SEXTRACTOR` (Bertin & Arnouts 1996). This catalogue consists of 6774 sources and is 99 per cent complete at $\sim 6\sigma$, which corresponds to a flux sensitivity of $0.3\text{--}0.4 \text{ Jy beam}^{-1}$. We have used the same source extraction algorithm and method described by Contreras et al. (2013) to produce a catalogue for the currently unpublished $280^\circ < \ell < 330^\circ$ and $21^\circ < \ell < 60^\circ$ regions of the survey. When the sources identified in these regions are combined with those identified by Contreras et al. (2013), we obtain a final compact source catalogue consisting of $\sim 10\,000$ sources (the full catalogue will be presented in Csengeri et al., in preparation). The observations typically had an rms pointing accuracy of ~ 4 arcsec (Schuller et al. 2009), which we adopted as the positional accuracy for the catalogue. This catalogue provides a complete census of dense dust clumps located in the inner Galaxy and includes all potential massive star-forming clumps with masses greater than $1000 M_\odot$ out to 20 kpc (Paper I).

2.2 CORNISH Survey

The CORNISH project has mapped the northern Galactic plane between $\ell = 10^\circ$ and 65° and $|b| < 1^\circ$ for 5 GHz radio-continuum emission. The survey was designed to detect UC H II regions across the Galactic disc (Hoare et al. 2012) and to form a radio counterpart to recent arcsecond-resolution infrared surveys (e.g. UKIDSS, GLIMPSE and MIPS GAL). Utilizing the Karl Jansky Very Large Array (VLA) in B and BnA array configurations, it is possible to resolve radio emission on angular scales ranging from ~ 1.5 to 20 arcsec. The rms noise level in the CORNISH images is better than $0.4 \text{ mJy beam}^{-1}$, sufficient to detect free-free emission from an optically thin H II region around a B0 star on the other side of the Galaxy. The high-reliability CORNISH catalogue (Purcell et al. 2013) contains 2637 distinct sources detected above a 7σ intensity cutoff, ~ 250 of which have been identified as possible H II regions via a visual inspection of infrared images (Purcell et al., in preparation).

The catalogue is 90 per cent complete for point sources at a flux density of 3.9 mJy ; however, this value decreases with increasing source size as the incomplete uv -coverage leads to missing flux in extended objects. Purcell et al. (2013) have used the GLIMPSE $8 \mu\text{m}$ data to identify and re-integrate over-resolved radio emission which has been broken up by the source finder.

3 ATLASGAL–CORNISH ASSOCIATIONS

There are 2337 CORNISH sources located within the region common to both surveys, which corresponds to approximately 90 per cent of the CORNISH catalogue. In this section, we will describe the matching procedure and classification of the matched sources and discuss the properties of the associated clumps.

3.1 Matching statistics and classification

In addition to the source catalogue, `SEXTRACTOR` also produced image masks marking the location and extent of each extracted ATLASGAL source. We have used these masks to identify positional matches with the CORNISH radio catalogue. This procedure identified 285 ATLASGAL–CORNISH matches. This initial sample

includes chance alignments of Galactic dust emission with extragalactic background sources, PNe and radio emission from more extended H II regions. The first two forms of contamination are relatively straightforward to identify and remove, but emission from extended H II regions can be more difficult to identify. Snapshot interferometric observations, such as those used for CORNISH, filter out the larger scale emission (> 14 arcsec), but are still sensitive to more compact, bright spots of emission commonly found towards the edges of more extended H II regions where the ionization front is interacting with the molecular material.

To assist in the classification of radio emission we have created false-colour, mid-infrared images of each of the ATLASGAL–CORNISH matches, using data extracted from the GLIMPSE survey (Benjamin et al. 2003). These images are $5 \text{ arcmin} \times 5 \text{ arcmin}$ in size and combine the 4.5 , 5.8 and $8.0 \mu\text{m}$ IRAC bands (coloured blue, green and red, respectively) centred on the peak of the submm emission. Examples are presented in Figs 1 and 2 with radio and submm emission contours overlaid (see figure captions for more details). Fig. 1 shows examples of the compact and UC H II regions that make up our final sample, while in Fig. 2 we present examples of the contaminating sources that have been excluded.

The correlation of the mid-infrared emission with tracers of the molecular gas and dust has been used in a number of previous studies to aid in the classification of radio emission (e.g. Urquhart et al. 2009b; Hindson et al. 2012; Purcell et al. 2013). Here we use the false-colour images to classify all of the ATLASGAL–CORNISH matches into one of five distinct types using the following criteria.

(i) Radio star: radio emission is ubiquitous from stars across the Hertzsprung–Russell diagram (Güdel 2002) and is thought to originate mainly from dynamo effects in convective envelopes or magnetospheric structures. At the sensitivity of CORNISH these sources are expected to be detected out to distances of several kpc. Stellar sources are easily identified by their point source morphology in GLIMPSE images and their generally blue mid-infrared colours, as can be seen in the upper-left panel of Fig. 2.

(ii) PNe: approximately 1000 PNe are expected to be detected by CORNISH. These are expected to have typical radio continuum flux densities of $5\text{--}50 \text{ mJy}$ and sizes of a few arcseconds (Hoare et al. 2012). These objects have similar mid-infrared colours to the embedded H II regions in which we are interested but their spectral energy distributions (SEDs) tend to peak at shorter wavelengths than the latter and, thus, the majority will not be detected by ATLASGAL. The dusty envelopes of some nearby PNe will be detected. However, these objects are generally isolated in near- and mid-infrared images, having dispersed their natal clump and have a relatively simple infrared morphology making them relatively easy to identify (Purcell et al. 2012). Two examples are presented in the upper-middle and right-hand panels of Fig. 2.

(iii) Extragalactic background sources: these are usually located away from the peak of the submm continuum and are not typically associated with any mid-infrared emission. Their radio emission also tends to be weaker than for UC H II regions and they are generally only marginally resolved and so can be well fitted with a Gaussian profile. Three examples of some of these extragalactic background sources are presented in the middle panels of Fig. 2.

(iv) Evolved H II regions: for these sources the radio emission is found to be coincident with diffuse PAH emission at $8 \mu\text{m}$, associated with the ionization front where UV photons are interacting with the surrounding dust. The radio emission is typically a factor of 3 weaker (see Table 1 for average values), often has an irregular morphology, and tends to be associated with only part of the shell

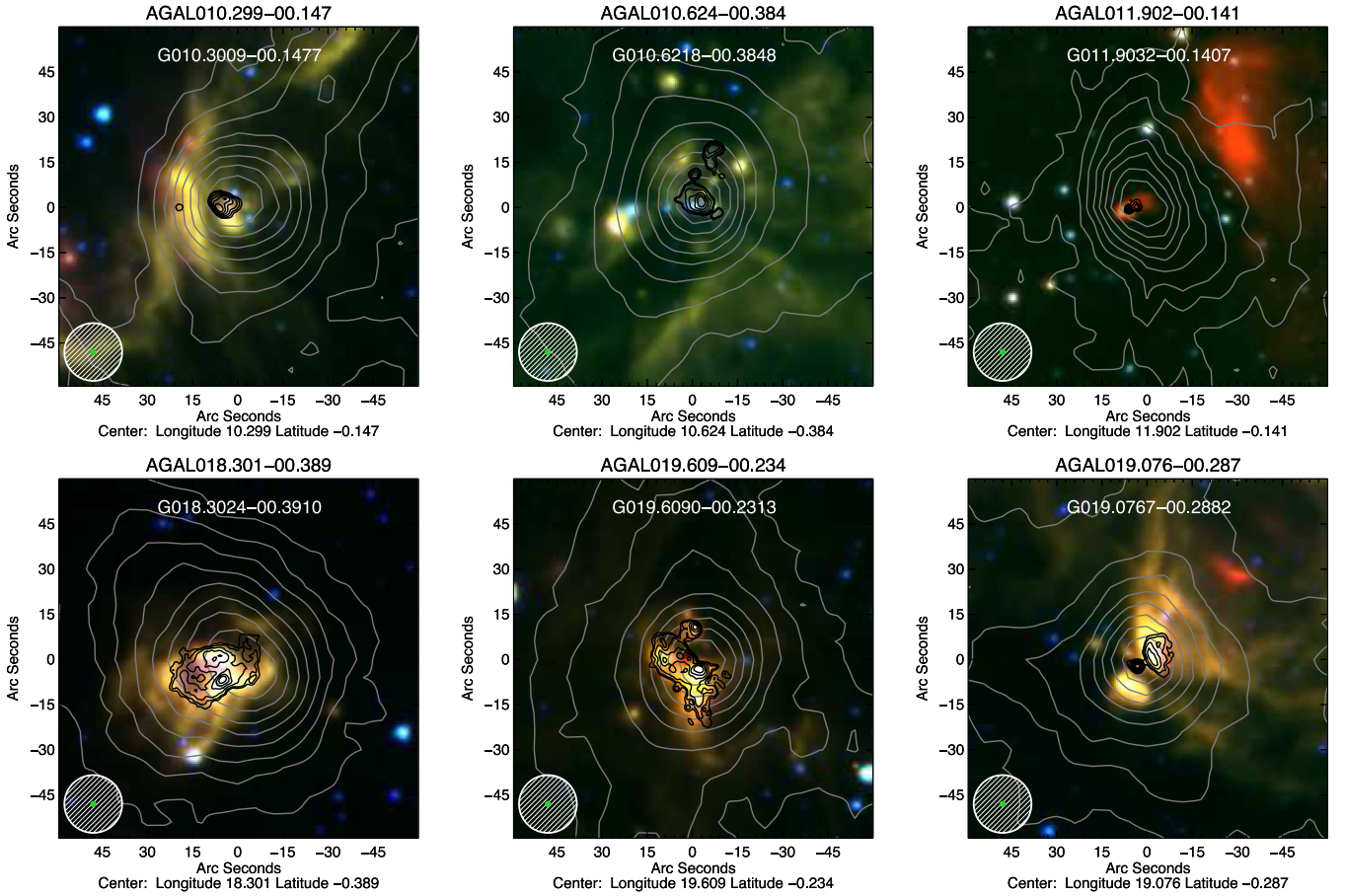


Figure 1. Examples of the local mid-infrared environments found towards the ATLASGAL–CORNISH identified compact and UC H II regions and included in the final catalogue (see the text for details). These images are composed of the GLIMPSE 3.6, 4.5 and 8.0 μm IRAC bands (coloured blue, green and red, respectively) which are overlaid with grey contours showing the ATLASGAL 870 μm thermal dust emission, and black contours showing the CORNISH 5 GHz radio continuum emission. The ATLASGAL source name is given above the image, while the CORNISH source name is given in white in the top part of the image. The CORNISH and ATLASGAL survey resolutions are indicated by the green circles and the white hatched circle in the lower left corner of each image, respectively. The contour levels start at 2σ and increase in steps set by a dynamically determined power law of the form $D = 3 \times N^i + 2$, where D is the dynamic range of the submm emission map (defined as the peak brightness divided by the local rms noise), N is the number of contours used (8 in this case) and i is the contour power-law index. The lowest power-law index used was 1, which results in linearly spaced contours starting at 2σ and increasing in steps of 3σ (see Thompson et al. 2006 for more details). The advantage of this scheme over a linear scheme is its ability to emphasize both emission from diffuse extended structures with low surface brightness and emission from bright compact sources.

structure surrounding the H II region, as viewed in mid-infrared images. Any compact radio emission is therefore more likely to arise from localized dense parts of the shell rather than being a genuine compact or UC H II region. Three examples of these more extended H II regions are presented in the lower panels of Fig. 2 (see also Deharveng et al. 2010 for a detailed study of ATLASGAL sources associated with these more evolved H II regions).

(v) Compact and UC H II regions: radio sources that appear to be self-contained regions of radio continuum emission are coincident with compact mid-infrared sources. These sources are usually found towards the peak of the submm emission seen in the ATLASGAL maps. In the cases where the structure of the radio emission is resolved, their morphologies are also found to be correlated with emission features seen in the mid-infrared images (Hoare et al. 2007). A selection of these H II regions is presented in Fig. 1.

It is the last of these classifications that will be the focus of this study. In a small number of radio sources there is some ambiguity in the classification. We have opted to identify such cases as being extended H II regions in order to avoid the possibility of contaminating

our target sample of compact and UC H II regions. This may result in a small number (<6) of genuine sources being wrongly excluded but these represent only a few per cent of our final sample and are therefore unlikely to affect the results significantly. In Table 1, we present a summary of the classifications, the number of each type identified and the average observed radio continuum parameters.

Since the CORNISH survey is sensitive to radio emission with angular scales up to ~ 20 arcsec, it is likely that our final sample will consist of a mixture of both compact and UC H II regions (the physical sizes will be investigated in Section 7.2). For brevity, we will refer to the contents of this sample collectively as H II regions in the discussion that follows.

3.1.1 Contamination from background radio sources

There are approximately 2600 compact radio sources in the CORNISH catalogue, most of which (~ 80 per cent) are likely to be extragalactic in origin. With such a large number of background sources there is a non-negligible probability of contamination

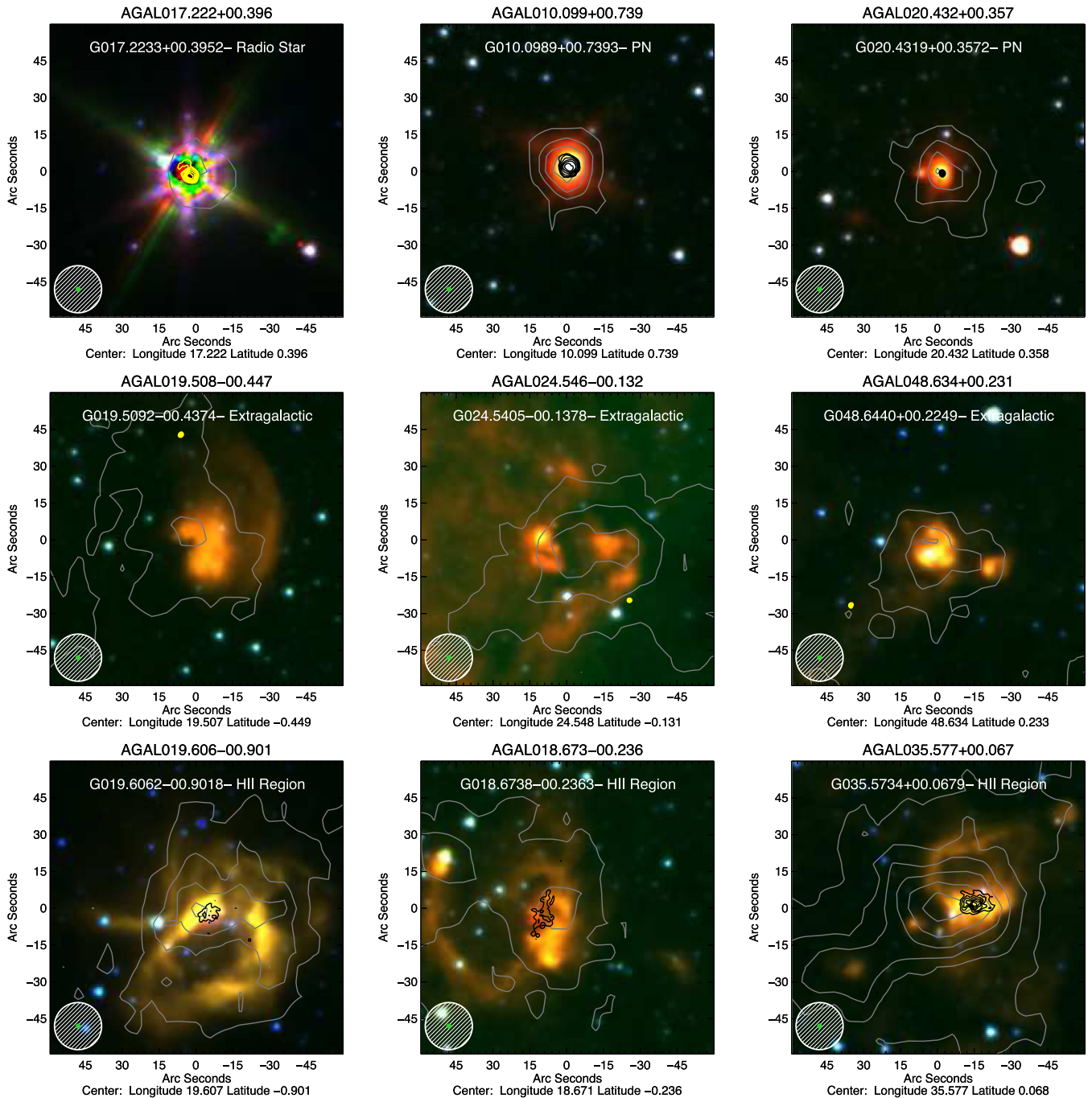


Figure 2. Examples of the local mid-infrared environments found towards the ATLASGAL–CORNISH matches that have been excluded from the final catalogue (see the text for details). The image details are as described in Fig. 1 with the exception of the contours overlaid on the images presented in the middle and upper-left panels; these are shown in yellow to make the radio emission more prominent.

from the chance alignment between these background sources and the dust emission detected by ATLASGAL. In our visual inspection of the mid-infrared images, we have identified eight of the ATLASGAL–CORNISH matches as being due to background source contamination. However, we need to check this number against the probability of such alignments occurring.

To estimate the number of chance alignments, we have conducted a Monte Carlo simulation in which 2000 point sources were randomly distributed in longitude and latitude over the CORNISH survey region. The resulting catalogue was then compared with the ATLASGAL emission masks and the number of matches noted.

After one hundred repeats, the number of chance alignments ranged from 9 to 20 with a mean value of 14 ± 4 , where the error is the standard deviation.

The number of expected chance alignments of background sources with the Galactic dust emission is in reasonable agreement with the actual number identified. It is therefore likely that we have identified the majority of the extragalactic background sources that would otherwise contaminate our sample. There may still be a small number of extragalactic interlopers that have made it into our final H II region sample but contamination from these is at most few per cent. Furthermore, it is likely that these interlopers will

Table 1. Summary of ATLASGAL–CORNISH radio source classifications and average radio continuum parameters.

Classification	Number	Peak flux (mJy beam ⁻¹)	Int. flux (mJy)	Angular diameter (arcsec)
Radio Star	1	54.2	199.9	3.6
PNe	7	19.2	84.3	2.2
Background galaxies	8	5.7	9.5	2.0
H II regions	56	13.4	312.1	9.7
Compact/UC H II regions	213	44.4	373.0	4.4
Total	285	36.6	343.1	5.3

be weeded out at a later stage of the analysis as we expect their derived parameters will deviate from those of the larger sample of H II regions.

3.2 Reliability of the H II region catalogue

As a consistency check on our classifications, we have compared our H II region catalogue to the samples identified from the targeted surveys of Wood & Churchwell (1989a) and Kurtz et al. (1994). These studies identified candidate UC H II regions by selecting sources with similar infrared colours to those of known UC H II regions. The infrared-selected sample was then observed in the radio continuum with the VLA.

Wood & Churchwell (1989a) identified 75 UC H II regions, 60 of which are located in the ATLASGAL–CORNISH region. We recover 48 of these; however, of these we classify four as extended, rather than compact or UC H II regions (i.e. G10.15–0.34, G10.46+0.03B, G35.05–0.52 and G45.48+0.13) and one as a PN (G10.10+0.73; see the upper centre panel of Fig. 2 for an image of this object). Eight of the remaining 12 Wood & Churchwell (1989a) sources are unresolved and are probably a mixture of extragalactic background sources and more distant H II regions that fall below the sensitivity of either CORNISH or ATLASGAL. Using the CORNISH data base, we find that G50.23+0.33, G42.90+0.57A and G42.90+0.57B have been classified as extragalactic radio lobes. Two other sources (G44.26+0.10 and G33.50+0.20) are included in the CORNISH catalogue but there is no significant dust emission coincident with the radio position, which would be expected if these were genuine compact or UC H II regions.

No radio emission is detected in CORNISH towards three more of the Wood & Churchwell (1989a) sample of UC H II regions, G54.0960.059, G15.040.68 and G10.46+0.03A. The first of these is unresolved with a peak flux of 1.8 mJy beam⁻¹ and so is likely to fall below the CORNISH catalogue’s 7 σ threshold (~ 2.8 mJy beam⁻¹). The second is located towards the M17 star-forming complex, which has significant bright, extended radio emission, causing locally increased, artefact-dominated noise (rms ~ 6 mJy beam⁻¹) in the associated CORNISH tiles; hence this object was not recovered. The third source is detected in the CORNISH data, but below the 7 σ threshold. This source was found to have a peak flux of 43.1 mJy beam⁻¹ by Wood & Churchwell (1989a) and so may be variable.

Kurtz et al. (1994) identified an additional 29 UC H II regions in the ATLASGAL–CORNISH area and we find 28 of these in our matched sample. However, one radio source (i.e. G35.580+0.066) is coincident with the edge of what appears to be a more evolved H II region, and so has been excluded from our final sample (see the bottom-right panel of Fig. 2). The missing radio source (i.e. G53.605+0.046) is unresolved in the Kurtz et al. (1994) data, with a 3.6 cm peak and integrated continuum flux density of

1.1 mJy beam⁻¹ and 1.5 mJy, respectively. Assuming a flat spectral index ($\alpha \sim -0.1$ for thermal free–free emission), we expect a similar flux density in the CORNISH data, which would place it below the 7 σ threshold.

Of the 89 UC H II regions identified by Wood & Churchwell (1989a) and Kurtz et al. (1994) in the ATLASGAL–CORNISH region, we have matched 76 with dust emission, though, we classified 6 as either part of the shell of a more evolved H II region, or a PN. The remaining 13 radio sources not matched to any dust emission and are therefore likely to be part of the extragalactic background population.

Another aspect of the reliability of our sample that we need to consider is the possibility of missing genuine H II regions because their associated dust envelope falls below the ATLASGAL sensitivity. As discussed in Section 2.1, the ATLASGAL survey is sensitive to clumps with masses above 1000 M $_{\odot}$ across the Galaxy (~ 20 kpc). If we assume the Kroupa (2001) initial mass function (IMF) and a constant SFE of between 10 and 30 per cent (e.g. Lada & Lada 2003; Johnston et al. 2009), a minimum mass of $\sim 1\text{--}3 \times 10^3$ M $_{\odot}$ is required to form at least one ~ 10 M $_{\odot}$ star (e.g. Zinnecker & Yorke 2007), which is approximately equivalent to a star with a spectral type of B1. Given that the CORNISH survey is sensitive enough to detect an optically thin unresolved UC H II region around a star of spectral type B0 or earlier across the Galaxy (Purcell et al. 2013), it is unlikely we have missed many UC H II regions in this way.

Combining the ATLASGAL and CORNISH data sets has provided a simple and reliable method to identify a large sample of molecular clumps associated with H II regions with relatively uniform noise characteristics for both the radio continuum and submm data. Moreover, this constitutes an unbiased and relatively complete sample of H II regions over a large enough region of the Galaxy to provide statistically robust results.

3.3 Multiplicity

In total, 170 ATLASGAL sources are found to be associated with one or more H II regions. Fig. 3 shows the number of H II-regions associated with each ATLASGAL source. In the majority of cases (147 or 86 per cent), the ATLASGAL source is associated with a single embedded H II region, and only 23 are associated with

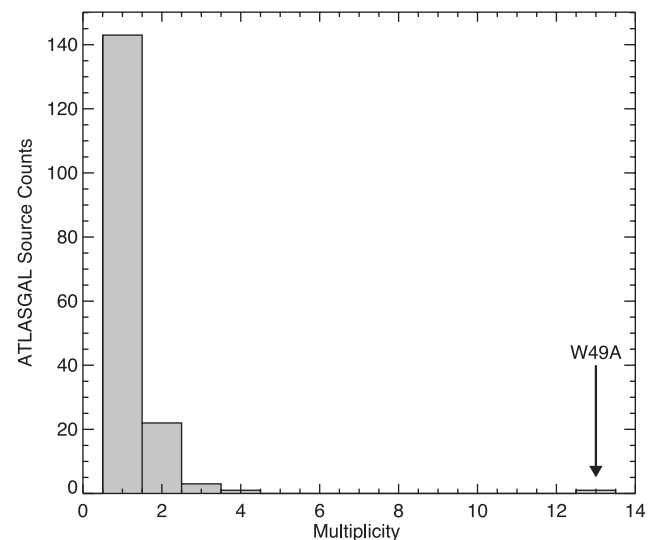


Figure 3. UC H II region density as a function of ATLASGAL clumps. The bin size is 1.

two or more UC H II regions. Only in five cases are ATLASGAL sources associated with more than two H II regions. We have been particularly careful in our analysis to exclude emission that appears likely to arise from larger H II regions. Furthermore, in each of these cases the radio emission is found to be a bright compact source and, in nearly all cases, is coincident with a discrete mid-infrared point source. We are therefore confident that these represent separate H II regions and not a problem with over-resolved emission, which can often be an issue with snapshot surveys.

The one object that stands out is AGAL043.166+00.011, which is associated with 13 compact radio sources. This ATLASGAL detection is part of the W49A star-forming complex that also includes two other ATLASGAL objects, i.e. AGAL043.148+00.014 and AGAL043.164−00.029, which are themselves associated with five more H II regions. In total, this complex is associated with 18 compact and UC H II regions, giving it the highest surface density of H II regions in the part of the Galaxy covered by CORNISH and ATLASGAL.

Three other intense star-forming regions (W31, G34.3+0.1 and AGAL020.081−00.136) contribute another 10 H II regions to the total and so, combined with W49A, these represent a significant fraction of the total number of H II regions identified (~ 13 per cent) but only constitute a few per cent of the total number of associated clumps. We therefore need to be careful that these rather intense star-forming regions do not affect the overall statistical results unduly.

3.4 Galactic longitude and latitude distribution

In the lower panels of Figs 4 and 5, we show the Galactic longitude and latitude distributions of the whole ATLASGAL compact source catalogue (filled grey histogram) and of the associated H II regions identified from the CORNISH catalogue (blue histogram). In the upper panels of these figures, we present the ratio of the two distributions, in each case. The ratio of the population of H II-region associated clumps with the total ATLASGAL source counts can be interpreted as an analogue of, or approximation to, the high-mass SFE of dense cores, evaluated on medium to large scales (0.1–1 kpc) and integrated over the time-scale associated with the UC H II

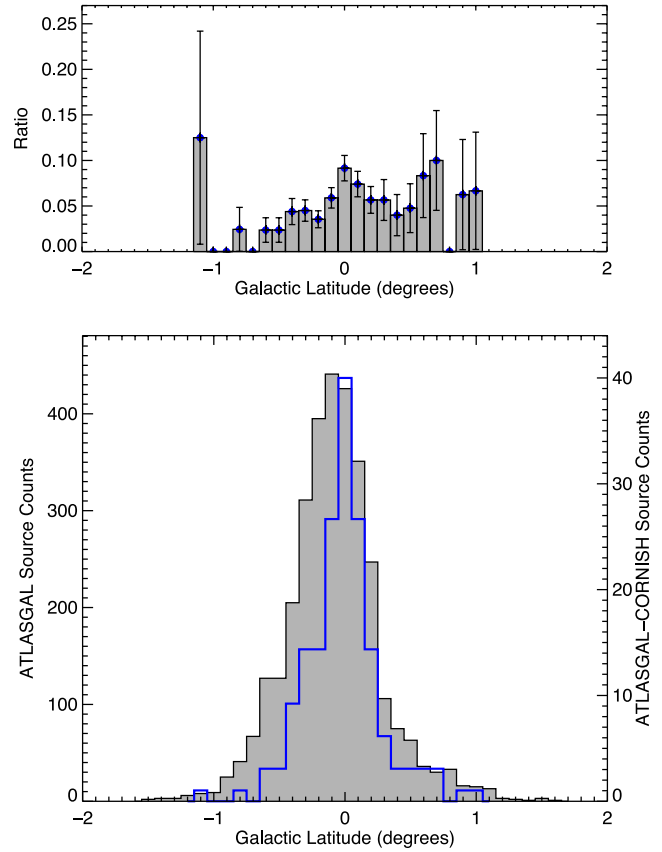


Figure 5. Galactic latitude distribution of ATLASGAL sources (grey filled histogram) and ATLASGAL clumps associated with one or more H II regions (blue histogram) are presented in the lower panel, while in the upper panel we present the ratio of the two histograms presented in the lower panel. The errors plotted in the upper panel are derived using Poisson statistics. The bin size used is 0.1° .

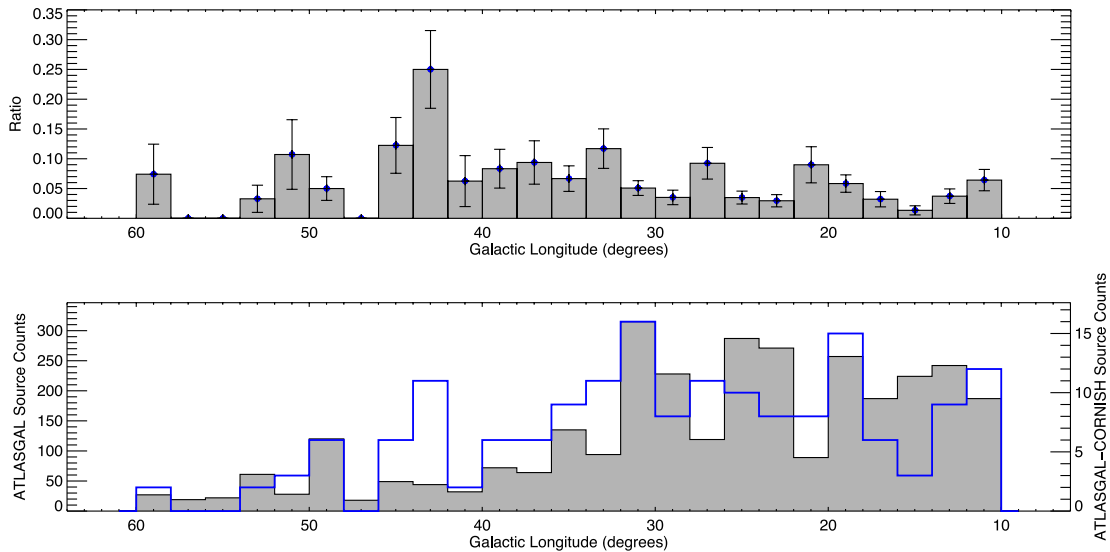


Figure 4. Galactic longitude distribution of ATLASGAL sources (grey filled histogram) and ATLASGAL-CORNISH associated clumps (blue histogram) are presented in the lower panel, while in the upper panel we present the ratio of the two histograms presented in the lower panel. The errors plotted in the upper panel are derived using Poisson statistics. The bin size used is 2° .

region phase. Since the latter is relatively short (a few times 10^5 yr; Davies et al. 2011; Mottram et al. 2011b), we can assume that any differences in the duration of a phase of star formation from region to region are not a significant influence on this ratio.

For the most part, the ratio is relatively flat and low across the longitude range of the sample. The mean value is 0.06 with sample standard deviation of 0.05, standard error on the mean of 0.01 and the median is 0.05. Almost all the values for each 2° longitude bin are within 2σ of the mean and most are within 1σ . The obvious exception comes in the $\ell = 42^\circ\text{--}44^\circ$ bin in which the ratio is 0.25 ± 0.065 , more than 3σ above the mean. This bin contains the ‘mini-starburst’ star-forming region W49A (see Section 3.3). While several massive star-forming regions have been put forward as Galactic-scale analogues of extragalactic starburst systems, W49A may be genuinely different. Its embedded MYSOs or protoclusters have an integrated bolometric luminosity of $6.9 \times 10^6 L_\odot$, dominated by two sources in the infrared-selected Red MSX Source (RMS; Urquhart et al. 2008a) (G043.1679–00.0095 and G043.1650–00.0285) with $L_{\text{bol}} > 10^6 L_\odot$. The infrared-traced SFE in W49A is very high ($L_{\text{bol}}/M_{\text{CO}} = 32 \pm 6 L_\odot M_\odot^{-1}$) and its luminosity function may be flatter than normal (Moore et al. 2012). Gas temperatures and densities are high at 50–100 K and $\sim 10^6 \text{ cm}^{-3}$, respectively (Roberts et al. 2011; Nagy et al. 2012). Also W49A is a source of very high energy gamma-rays (Brun et al. 2011) which may be the cause of the additional gas heating characteristic of starbursts (Papadopoulos et al. 2011).

Several other major star-forming complexes are included in the sample, notably W43 in the Scutum arm tangent at $\ell \sim 30^\circ$, W51 in the Sagittarius arm tangent at $\ell \sim 49^\circ$ and W31 at $\ell = 10^\circ$, and at least the first two of these have also been labelled as mini-starburst regions. On the scale of individual giant molecular clouds, the SFR in these regions may be very high but, on the scales traced by these data, they have little effect on the average SFE (e.g. Eden et al. 2012). W49A is exceptional in this regard (see also Moore et al. 2012).

As noted in previous studies of the ATLASGAL data (i.e. Schuller et al. 2009; Beuther et al. 2012; Contreras et al. 2013), the peak in the latitude distribution of the general population of compact ATLASGAL sources is skewed slightly to a negative value of b . This is consistent with the Sun’s location above the Galactic mid-plane (Humphreys & Larsen 1995). However, we note that the peak in the source counts of ATLASGAL clumps associated with H II regions is peaks at $b = 0$. This difference in latitude distribution is most likely a distance effect, since the H II region sample has a significantly larger mean distance than the unmatched sources (see also Section 4.2).

3.5 Correlation between H II regions and their natal clumps

3.5.1 Clump–H II region angular correlation

Thompson et al. (2006) reported a significant difference in the angular offsets of embedded UC H II regions and of methanol masers, relative to the peak submm flux of their host clumps. They used separations between methanol masers and submm clumps drawn from targeted Submillimetre Common-User Bolometer Array (SCUBA) observations reported by Walsh et al. (2003), and compared positions from 850- μm SCUBA observations to those of UC H II regions taken from the literature (i.e. Wood & Churchwell 1989a; Kurtz et al. 1994; Giveon et al. 2005). Comparing these two samples, Thompson et al. (2006) found that the UC H II regions had, on average, significantly larger offsets from the submm peaks than the methanol masers and discussed several explanations for this.

In the left-hand panel of Fig. 6, we present the 2D distribution of angular offsets for the ATLASGAL–CORNISH sample of H II regions. This plot reveals the presence of a systematic offset of ~ 2 arcsec in Galactic longitude. This offset is approximately six times larger than the standard error on the measurement and is therefore significant. Since the CORNISH data have been referenced to calibrators with positional accuracies much better than 1 arcsec,

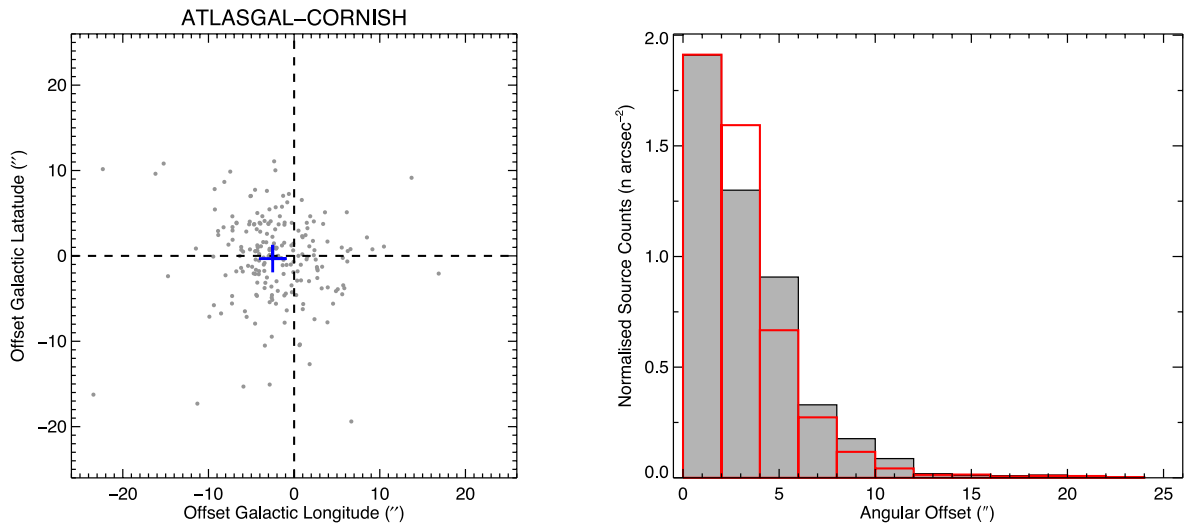


Figure 6. Left-hand panel: two-dimensional distribution of the angular offsets between the peak of the ATLASGAL dust emission and the matched CORNISH sources. The dashed vertical and horizontal lines indicate the x and $y = 0$ axes, respectively. The blue cross indicates the mean offset in longitude and latitude of the whole sample revealing a systematic offset from the 0,0 position ($\Delta\ell = -1.95 \pm 0.28$ and $\Delta b = -0.24 \pm 0.29$). Right-hand panel: normalized source counts for the H II regions (grey filled histogram), and methanol-maser distribution from Paper I (red histogram), are shown as a function of separation between their position and that of the peak of the submm emission of their associated ATLASGAL source. The peak of the methanol-maser distribution has been normalized to that of the H II region distribution to facilitate comparison of the two samples. A correction has been made to compensate for the systematic offset seen in the left-hand panel of this figure. We have truncated the x -axis of this plot at 26 arcsec as there are only eight H II regions that have larger separations and the angular surface density effectively falls to zero. The bin size is 2 arcsec.

this positional discrepancy is likely to be associated with the ATLASGAL data. This positional offset is approximately 10 per cent of the ATLASGAL survey's resolution (19.2 arcsec) and approximately half of the rms pointing error for the ATLASGAL survey (~ 4 arcsec; Schuller et al. 2009) and so is not significant for this work.

The right-hand panel of Fig. 6 shows the normalized source counts (i.e. source counts divided by the area of the bin annulus) distribution for this sample, corrected for the systematic offset mentioned in the previous paragraph (grey filled histogram). In addition, we have plotted the corresponding distribution for the sample of ATLASGAL clumps matched with methanol masers presented in Paper I (red histogram). The two distributions appear very similar and indeed a Kolmogorov–Smirnov (KS) test gives a p -value of 0.03 and is therefore unable to reject the null hypothesis that these two distributions are drawn from the same parent population (to reject the null hypothesis with $\geq 3\sigma$ confidence the p -value must be < 0.003). The positions of the embedded massive stars traced by the methanol masers and H II regions are therefore closely correlated with the peak of the submm emission, which is itself typically found towards the geometric centres of clumps (mean offset is 0.68 arcsec with standard error and deviation of 0.48 and 6.25 arcsec, respectively).

This is in contrast with the results reported by Thompson et al. (2006). Since the angular distribution for the methanol masers and submm clumps is broadly in agreement with that determined by Walsh et al. (2003), we conclude that the difference is associated with the UC H II region positions. We tested the astrometry of the ATLASGAL sources and their associated methanol masers in Paper I and found it to be in excellent agreement. Thompson et al. (2006) were careful to check the astrometry of their submm observations and so the issue is likely to reside in the positional accuracy of the various radio catalogues used by Thompson et al. (2006). We note that 80 per cent of the sample of UC H II regions discussed by Thompson et al. (2006) were drawn from the Giveon et al. (2005) catalogue, which was colour selected from a larger sample of 5 GHz Galactic radio sources presented by White, Becker & Helfand (2005). However, a comparison of the positions given for the same sources in these catalogues (i.e. those of White et al. 2005 and Giveon et al. 2005) reveals the presence of a systematic offset of 5 arcsec in declination. It is unclear where this positional error in Giveon et al. (2005) has come from; however, it has led to the larger offset between the radio sources and the peak submm emission of their host clumps reported by Thompson et al. (2006).

3.5.2 Clump structure and morphology

Fig. 7 shows the distributions of the aspect ratio and the Y -factor (ratio of integrated to peak submm fluxes) for both the ATLASGAL–CORNISH H II regions and methanol-maser associated clumps discussed in Paper I (blue and red histograms, respectively). These have been overlaid on the distributions of the same parameters for the whole ATLASGAL compact source catalogue (grey filled histogram). Examination of these distributions reveals significant differences between the aspect ratios and Y -factors of the H II-region associated clumps and those of the ATLASGAL compact source catalogue, with the H II region clumps having a more spherical structure within which the massive star formation is taking place.

The H II-region and methanol-maser samples have very similar aspect-ratio and Y -factor distributions. The median aspect ratio and Y -factor for the H II-region and methanol-maser associated clumps

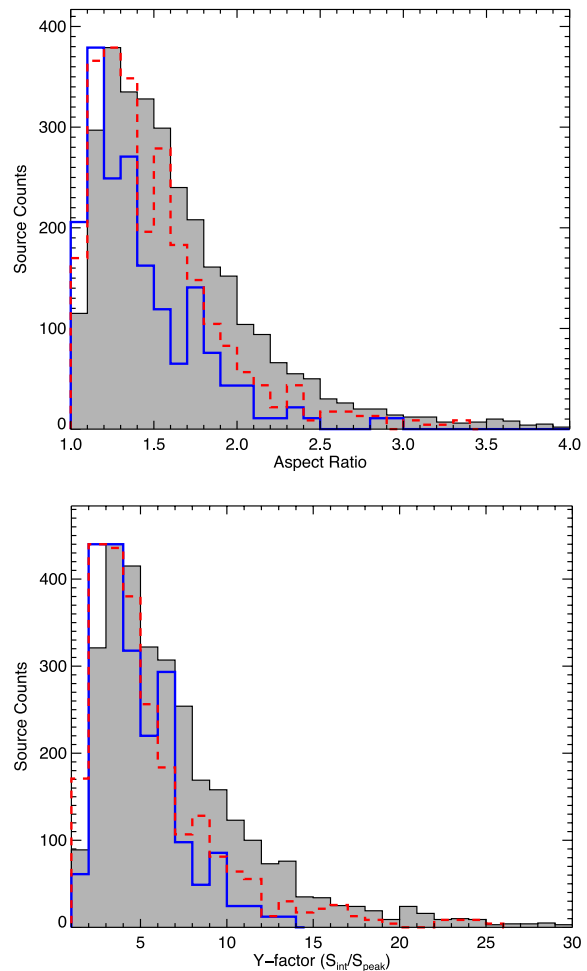


Figure 7. The aspect ratio and Y -factor distributions for the ATLASGAL–CORNISH H II-region associated clumps are shown in the upper and lower panels, respectively (blue histogram). The distribution of the whole sample of ATLASGAL compact sources located between $\ell = 10^\circ$ – 60° and $|b| < 1$ is shown by the grey filled histogram while the ATLASGAL–MMB associated clumps are shown in red. The peaks of the H II and methanol-maser distributions have been scaled to the peak of the ATLASGAL compact source distribution. The bin sizes used for the aspect ratio and Y -factor distributions are 0.1 and 1, respectively.

are 1.33 and 1.4, and 4.41 and 4.6, respectively. For both parameters, a KS test comparing the two samples is unable to reject the null hypothesis that they are drawn from the same parent distribution (p -values are 0.03 and 0.05 for the aspect ratio and Y -factor, respectively).

The fact that there is no significant difference in clump structure between the various embedded stages of massive star formation was noted in Paper I, as well as by previous studies of high-mass protostellar cores (Williams, Fuller & Sridharan 2004) and UC H II regions (Thompson et al. 2006). This has led to the conclusion that the structure of the clump *envelope* does not evolve significantly during the early embedded stages in the massive star formation process, and this is consistent with our present findings.

The right-hand panel of Fig. 6 presents the angular separation between the H II regions and the peak of the submm emission and reveals a strong correlation between the two. However, given the large angular size distribution of the ATLASGAL clumps this plot does not provide a reliable indication of where in the clumps the

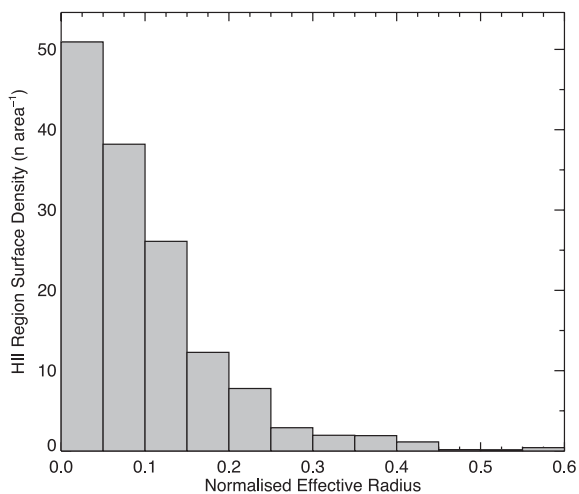


Figure 8. Compact and UC H II region surface density as a function of normalized clump effective radius. We have truncated the x -axis at a normalized effective radius of 0.6 as there are only seven sources with radii larger than this and the distribution effectively falls to zero beyond this point. The bin size is in units of 0.05 times the effective radius.

star formation is taking place. In Fig. 8, we present the surface density as a function of the normalized clump effective radii, which shows the distribution of H II regions with respect to the clump radius.¹ There is a very strong correlation between the positions of the H II regions and the centre of the clump’s where the highest column densities are found, with the majority being located within the inner 10–20 per cent of their clumps radius. Only seven sources are found to have normalized clump effective radii larger than 0.6, which correspond to ~ 4 per cent of the sample.

The aspect ratio of the clumps and the tight angular correlation between the H II regions and dense gas found towards the centre of the clumps suggests that the majority of these massive stars are forming in the inner most parts of centrally condensed, spherical structures. Although the mean offset is almost negligible, we note that for many cometary H II regions the peaks of the radio and submm emission are measurably offset (e.g. G019.0767–00.2882, which is offset by 4 arcsec from the peak in the dust emission – lower-right panel of Fig. 1). However, these offsets are at the limit of the resolution of these data.

4 KINEMATIC PROPERTIES OF THE CLUMPS

4.1 Molecular line data

In order to obtain velocity information for our sample of massive, star-forming clumps, we have searched the literature to find complementary molecular line data. Where available, we have given preference to molecular transitions with high critical densities. The majority of the source velocities are taken from one of three recently reported sets of targeted observations of the lower excitation states of the ammonia inversion transition [NH₃ (1,1), (2,2) and (3,3); Urquhart et al. (2011b); Dunham et al. (2011a); Wien et al. (2012)]. These studies used the Green Bank and Effelsberg 100-m telescopes, with a resolution of ~ 30 and ~ 40 arcsec, respectively, and so are reasonably well matched to the submm clump sizes. Between them, these programmes observed 126 of our H II-region

associated sample of clumps, providing unambiguous velocities and line widths. The NH₃ (2,2) transition was detected towards 115 of these sources allowing their kinetic temperatures to be estimated.

A further 24 clumps have been observed by Schlingman et al. (2011) using HCO⁺ (3–2) and N₂H⁺ (3–2) as part of follow-up observations of sources detected in the Bolocam Galactic Plane Survey (BGPS; Aguirre et al. 2011). The velocities for another 13 clumps were drawn from targeted observations made by Bronfman, Nyman & May (1996), Blum, Conti & Damiani (2000), Urquhart et al. (2009a), Anderson & Bania (2009) and Walsh et al. (2011). In total we have been able to obtain velocities from the literature for 163 of the 170 clumps in our sample.

Of the remaining seven sources, three have been observed as part of an ATLASGAL follow-up programme using the IRAM 30-m telescope. Analysis is still ongoing for these data and their results will be discussed in detail in a future publication (Csengeri et al., in preparation). However, we have been able to extract the source velocities from the preliminary data reduction and these should be reliable enough to estimate the kinematic distance to these sources. For the remaining four sources, we have obtained a ¹³CO (3–2) spectrum from the JCMT HARP survey of the Galactic plane (Moore, private communication) for two sources, and obtained Director’s time on the Mopra radio telescope to observe the other two sources.²

Using a combination of literature values and the ad hoc observations described in the previous paragraph, we have been able to obtain a radial velocity for every object in our sample. In the following two subsections, we will use the observed molecular properties to investigate the Galactic longitude–velocity distribution, velocity dispersion and temperature distribution of these massive, star-forming clumps.

4.2 Galactic longitude–velocity distribution

In Fig. 9, we present the longitude–velocity distribution of our sample with respect to that of the molecular gas as traced by the ¹²CO (1–0) emission mapped by Dame, Hartmann & Thaddeus (2001) and the predicted position of the spiral arms from Taylor & Cordes (1993) and updated by Cordes (2004). The distribution of the compact and UC H II regions agrees well with the location of the model spiral arms, in general. However, the H II-region and molecular-gas distributions correlate significantly less well.

This is in stark contrast to the longitude–velocity distribution of a large sample of ATLASGAL sources observed in ammonia by Wien et al. (2012, i.e. from the general ATLASGAL catalogue). These authors found a much stronger correlation between the distribution of ATLASGAL sources and the molecular gas (see their fig. 3). One possible explanation is that the velocities have been incorrectly assigned to our sample, but the majority of these have been drawn from targeted observations using high-density tracers (e.g. NH₃, CS and HCO⁺ data discussed in the previous subsection). These data all show a single, unambiguous velocity component in the direction of each source and are therefore unlikely to be incorrect.

Inspection of the various distributions shown in Fig. 9 shows that the CO emission associated with the more distant Norma arm is relatively weak and occurs in small, rather isolated sources but that the H II regions here are quite well correlated with the CO. In

¹ The effective radii is as defined in Rosolowsky et al. (2010).

² The Mopra radio telescope is part of the Australia Telescope National Facility which is funded by the Commonwealth of Australia for operation as a National Facility managed by CSIRO.

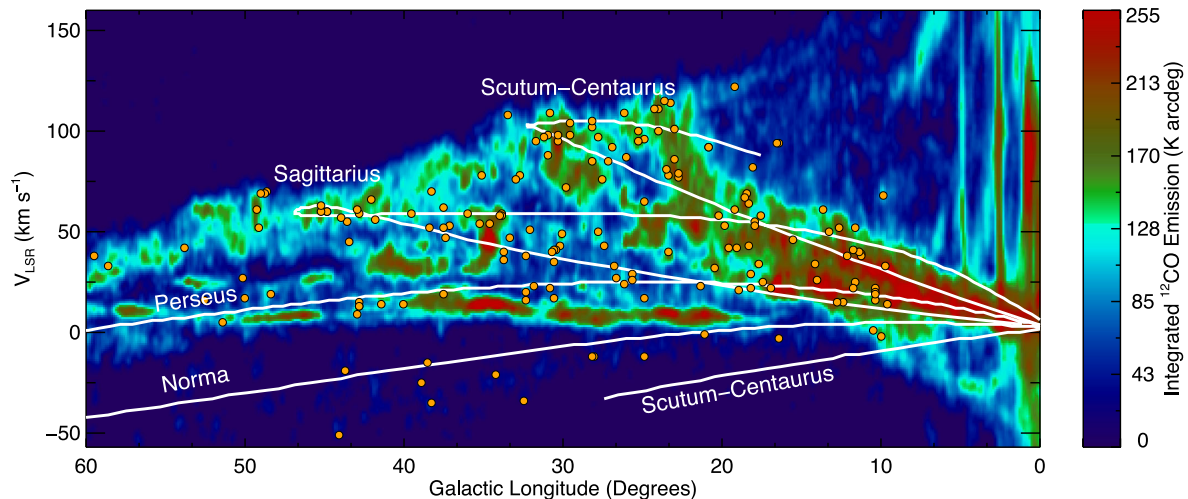


Figure 9. Galactic longitude–velocity distribution of H II-region associated clumps. The background image shows the distribution of molecular gas as traced by the integrated $^{12}\text{CO } J=1-0$ emission (Dame et al. 2001); the colour bar on the right shows the relative intensity of the emission. The orange circles mark the positions of the UC H II regions and the white lines indicate the location of the spiral arms taken from the model by Taylor & Cordes (1993) and updated by Cordes (2004).

Section 5, we derive distances for the whole sample of clumps and find that the majority are located at heliocentric distances greater than ~ 10 kpc. From these two points, we conclude that the apparent anticorrelation between molecular gas and H II region distributions is because the CO emission is dominated by bright, relatively nearby clouds, with emission from the more distant clouds reduced by beam dilution. The higher correlation observed by Wienen et al. (2012) suggests that the general ATLASGAL catalogue is also dominated by a large population of nearby sources (see also the discussion in Section 3.4.3).

4.3 Line widths and temperatures

Given the variety of molecular tracers used to obtain velocity information, it is not straightforward to directly compare the line widths and kinetic temperatures measured from these different tracers. In this subsection, therefore, we will concentrate on the properties

derived from the ammonia observations, and although these are not available for the complete sample, they do include a sufficiently large fraction of the sample (~ 70 per cent) such that the statistical results will be reliable.

In Fig. 10, we present histograms showing the FWHM line width of the ammonia (1,1) inversion transition (left-hand panel) and the kinetic temperature derived from the ratio of the NH_3 (1,1) and (2,2) transitions (see Urquhart et al. 2011b for more details). The statistical values for these two clump properties are tabulated in Table 2. The ammonia line widths have been determined from a simultaneous fit to the main and satellite features of the NH_3 (1,1) line and so these have been corrected for the effects of line broadening due to high optical depths. We note that determination of the kinematic temperature requires the detection of the NH_3 (2,2) transition and therefore the distribution shown in the right-hand panel of Fig. 10 may contain a selection bias to higher temperatures.

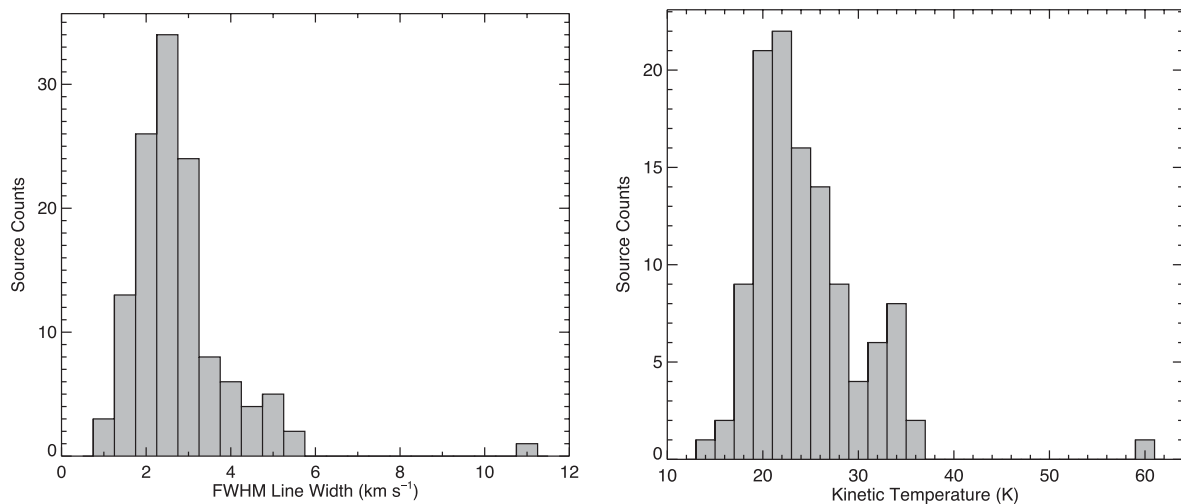


Figure 10. Left-hand panel: FWHM line width distribution of the NH_3 (1,1) inversion transition determined from the 126 clumps observed by Urquhart et al. (2011b), Dunham et al. (2011a) and Wienen et al. (2012). Right-hand panel: distribution of kinetic gas temperatures determined for the 115 clumps towards which both the NH_3 (1,1) and (2,2) transitions are detected.

Table 2. Summary of derived parameters.

Parameter	Number	Mean	Standard error	Standard deviation	Median	Min	Max
Angular offset (arcsec)	213	7.32	0.58	8.42	4.93	0.32	61.28
Aspect ratio	170	1.44	0.03	0.44	1.32	1.01	4.89
Y-factor	170	4.87	0.18	2.34	4.36	1.52	13.82
FWHM line width (km s ⁻¹)	126	2.77	0.11	1.20	2.55	0.86	11.0
Kinetic temperature (K)	115	24.65	0.55	5.89	23.33	14.68	60.76
Heliocentric distance (kpc)	169	9.09	0.32	4.17	9.78	1.87	18.60
Radius (pc)	154	1.52	0.07	0.88	1.38	0.19	6.01
Log[Clump mass (M _⊙)]	169	3.96	0.34	4.38	3.69	1.72	5.46
Log[Viral mass (M _⊙)]	123	3.93	0.37	4.07	3.74	2.08	4.89
Viral ratio	123	2.47	0.17	1.88	2.08	0.25	14.90
Log[Lyman continuum flux (s ⁻¹)]	212	47.66	0.06	0.85	47.73	45.49	49.69
UC H II region observed diameter (arcsec)	212	4.44	0.25	3.61	2.88	1.5	23.42
UC H II region diameter (pc)	212	0.16	0.01	0.17	0.11	0.01	0.89

The mean line width ($\sim 2.8 \text{ km s}^{-1}$) is typical of those found towards other young massive star formation regions identified by methanol masers (e.g. Pandian, Menten & Goldsmith 2009; Wielen et al. 2012) and MYSOs (Urquhart et al. 2011b), but significantly larger than for the general population of BGPS clumps reported by Dunham et al. (2011a) and the subsample of infrared dark ATLASGAL sources reported by Wielen et al. (2012) ($\sim 1.9 \text{ km s}^{-1}$ for both). The difference is approximately seven times the standard error of the mean line width. It is unclear at this point whether the increase in line width seen towards the UC H II region sample is a result of increased feedback from the embedded star formation or simply reflects that the clumps associated with H II regions are typically found at larger distances and consequently the physical scale probed by the telescope beam is larger; in accordance with the Larson size–line width relation (Larson 1981).

We find a mean kinetic temperature of $\sim 25 \text{ K}$, which is again similar to those of the methanol-maser and MYSO samples. However, it is significantly higher than found for starless clumps that tend to have kinetic temperatures of $\sim 15 \text{ K}$ (e.g. Pillai et al. 2006; Dunham et al. 2011a; Wielen et al. 2012). This suggests that young embedded stars are internally heating their host clump on scales large enough to be detected by the present data. It is important to note that all of these ammonia observations have been directed towards the peak of the submm emission or the infrared emission associated with the embedded H II regions. These observations are probing the hottest and most active parts of each clump and therefore the line width and kinetic temperatures are effectively upper limits. It is likely that the kinetic temperature of the gas in the more extended envelope, which contributes the majority of the flux, will be lower. We note that the kinetic temperature of the gas and dust are likely to be significantly higher closer to the H II regions, but the single dish telescopes used for these ammonia observations are unable to probe these size scales. However, the proportion of the total clump mass at higher temperature is very small and unlikely to significantly affect the estimated clump masses.

5 DISTANCES

In the previous section, we identified a large sample of compact and UC H II regions and their host clumps in which they are still deeply embedded. Before we can begin to investigate the properties of these massive star-forming environments (e.g. masses, densities and physical sizes), we need to determine their distances. Many of these H II regions were previously known and so distances are

available in the literature for a significant fraction of the sample. We have therefore conducted a comprehensive review of the literature and have identified 24 clumps that have measured maser parallax or spectroscopic distances (e.g. Reid et al. 2009 and Moisés et al. 2011, respectively), and kinematic distances for ~ 100 additional clumps.

Although maser parallax and spectroscopic distances are the most reliable method for determining distances, these types are only available for a relatively small fraction of our sample. It is therefore necessary to resort to kinematic distances; these will be discussed in more detail in the following subsection.

5.1 Kinematic distances

We use the radial velocities discussed in Section 4.1 in conjunction with the Galactic rotation model of Brand & Blitz (1993, $\theta_0 = 220 \text{ km s}^{-1}$ and $R_0 = 8.5 \text{ kpc}$, where θ_0 is the circular rotation speed and R_0 is the distance to the Galactic Centre) to estimate each source’s kinematic distance. For sources located within the solar circle (i.e. $< 8.5 \text{ kpc}$ from the Galactic Centre), there are two possible distances that are equally spaced on either side of the tangent position. These distances are known as the *near* and *far* distances and the degeneracy is commonly referred to as the kinematic distance ambiguity (KDA). Resolutions of this problem are discussed below. To estimate the associated distance uncertainties, we have assumed an uncertainty of $\pm 10 \text{ km s}^{-1}$ for the radial velocities (i.e. streaming motions; Reid et al. 2009), which corresponds to an error in the kinematic distances of $\sim 0.6\text{--}1 \text{ kpc}$. These distances are given in Table 3 for individual clumps and their overall statistics are summarized in Table 2.

Of the 146 clumps for which we were unable to assign a maser parallax or spectroscopic distance we find: 12 have negative velocities and are therefore located outside the solar circle and so are not affected by KDA; 13 are located within 10 km s^{-1} of the tangent velocity and so have been placed at that distance; 1 was placed at the near distance as a far distance allocation would place it at an unrealistically large height above the Galactic mid-plane for a massive star formation region (i.e. $z > 120 \text{ pc}$, which is four times the scaleheight for massive stars; Reid 2000; Urquhart et al. 2011a). From the literature review, we were able to identify a further 73 clumps included in previous studies that have already resolved their distance ambiguities.

From the literature and the automatic assignment of sources to the near, far and tangent distances, described at the beginning of

Table 3. Derived clump parameters.

ATLASGAL name	Complex name	H II density	Aspect ratio	Peak flux (Jy beam ⁻¹)	Int. flux (Jy)	Y-factor	V_{LSR} (km s ⁻¹)	Ref.	Distance (kpc)	R_{GC} (kpc)	Radius (pc)	Log(N(H ₂)) (cm ⁻²)	Log(M_{clump}) (M_{\odot})	Log(M_{vir}) (M_{\odot})
AGAL010.299–00.147	W31-North	1	1.9	7.67	54.18	7.06	12.8	1	2.4	1	6.2	0.79	23.18	3.44
AGAL010.321–00.257	W31-South	1	2.8	2.39	14.93	6.24	32.2	1	3.5	1	5.1	1.22	22.87	3.28
AGAL010.472+00.027	–	2	1.5	35.01	88.12	2.52	67.0	1,2,3	11.0	2,3	3.1	2.99	23.94	4.77
AGAL010.624–00.384	W31-North	4	1.4	33.10	116.87	3.53	–3.2	1,2,3	2.4	1	6.2	0.76	23.81	3.57
AGAL010.957+00.022	–	1	1.8	3.03	13.55	4.48	21.4	2,3	13.7	2,3	5.6	2.82	22.81	4.15
AGAL010.964+00.011	–	1	1.4	1.33	5.40	4.07	19.2	2,3	2.7	4	5.8	0.2	22.53	2.35
AGAL011.031+00.026	–	1	1.1	0.47	1.39	2.95	18.1	4	2.6	4	5.9	<0.29	22.22	1.72
AGAL011.034+00.061	–	1	1.5	1.77	6.05	3.41	15.4	4	14.7	5	6.6	1.7	22.66	3.86
AGAL011.109–00.397	–	1	1.9	3.53	42.04	11.92	0.1	2,3	16.8	4	8.6	6.01	23.04	4.81
AGAL011.902–00.141	–	2	1.2	3.17	20.03	6.32	38.4	1	4.1	6	4.6	0.99	23.12	3.27
AGAL011.936–00.616	–	1	1.2	7.22	45.64	6.32	37.0	1	4.0	2,3,6	4.6	1.04	23.33	3.61
AGAL011.946–00.036	–	1	1.8	1.75	7.62	4.36	39.1	3,4	12.6	4	4.7	1.98	22.65	3.83
AGAL012.198–00.034	–	1	1.1	2.92	7.42	2.54	50.8	1,2,3	11.9	6	4.0	1.32	22.97	3.76
AGAL012.208–00.102	–	1	1.2	11.58	37.39	3.23	24.5	1	13.6	5	5.6	2.78	23.38	4.58
AGAL012.431–00.049	–	1	2.4	0.72	5.71	7.95	21.1	1	13.9	7	5.9	1.41	22.39	3.78
AGAL012.431–01.114	–	1	1.6	4.28	14.73	3.44	39.7	1,2,3	4.1	4	4.6	0.62	23.00	3.14
AGAL012.804–00.199	W33	2	1.9	36.48	339.23	9.30	36.5	4	2.1	8	6.5	1.04	23.97	3.92
AGAL012.998–00.357	–	1	1.3	2.29	9.86	4.32	14.4	1	1.9	4	6.6	0.3	22.87	2.32
AGAL013.209–00.144	–	1	1.3	3.09	30.64	9.90	51.1	1,3	4.6	5	4.2	1.2	23.05	3.55
AGAL013.384+00.064	–	1	1.4	1.11	5.45	4.92	14.1	4	1.9	4	6.7	0.26	22.45	2.02
AGAL013.872+00.281	–	1	1.3	5.52	29.72	5.39	48.5	1	4.4	2	4.4	1.1	23.13	3.50
AGAL014.246–00.071	–	1	1.5	2.98	11.43	3.84	60.2	1	11.6	4	4.0	2.05	23.06	3.93
AGAL014.607+00.012	–	1	1.6	2.74	16.62	6.06	25.5	1,2,3	2.8	4	5.8	0.58	22.87	2.86
AGAL014.777–00.334	–	1	1.1	0.95	4.02	4.24	33.2	4	13.1	4	5.3	1.16	22.39	3.58
AGAL016.144+00.009	–	1	1.1	0.99	2.51	2.53	44.8	2,3	12.4	4	4.9	1.08	22.62	3.33
AGAL016.942–00.072	–	1	1.1	1.43	6.94	4.86	–4.2	4	17.0	4	9.2	2.04	22.56	4.05
AGAL017.029–00.071	–	1	1.2	1.27	3.42	2.69	93.2	1	10.4	6	3.4	0.93	22.63	3.31
AGAL017.112–00.114	–	1	1.8	0.62	1.82	2.93	93.1	2,3	10.4	6	3.4	<1.17	22.17	3.04
AGAL017.554+00.167	–	1	1.2	0.52	2.38	4.60	20.6	4	14.1	4	6.5	1.22	22.12	3.42
AGAL017.986+00.126	–	1	1.6	0.42	1.27	3.04	24.3	5	–	–	6.3	–	22.17	–

References. Velocity: (1) Wielen et al. (2012), (2) Urquhart et al. (2011b), (3) Urquhart et al. (2008b), (4) Schlingman et al. (2011), (5) this paper, (6) Dunham et al. (2011b), (7) Csengeri et al. in preparation, (8) Kolpak et al. (2003), (9) Blum et al. (2000), (10) Radio recombination line velocity from HOPS (Walsh et al. 2011; private communication A. Walsh), (11) Bronfman et al. (1996), (12) Anderson & Bania (2009).

References. Distance: (1) Moisés et al. (2011), (2) Sewilo et al. (2004), (3) Pandian et al. (2009), (4) this paper, (5) Pandian, Momjian & Goldsmith (2008), (6) Green & McClure-Griffiths (2011), (7) Wood & Churchwell (1989a), (8) Immer et al. (2012), (9) Roman-Duval et al. (2009), (10) Anderson & Bania (2009), (11) Kolpak et al. (2003), (12) Dunham et al. (2011a), (13) Fish et al. (2003), (14) Russeil (2003), (15) Downes et al. (1980), (16) Xu et al. (2011), (17) Araya et al. (2002), (18) Watson et al. (2003), (19) Stead & Hoare (2010), (20) Nagayama et al. (2011), (21) Sato et al. (2010), (22) Urquhart et al. (2012), (23) Zhang et al., in preparation.

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/>.

the previous paragraph, we have obtained distances for 124 clumps (~75 percent of the sample). For the remaining 46 clumps, we have extracted H I spectra from the Southern Galactic Plane Survey (SGPS; McClure-Griffiths et al. 2005) and the VLA Galactic Plane Survey (VGPS; Stil et al. 2006). There are two H I-related methods that can be employed to resolve KDAs; these are known as the absorption against a continuum (Fish et al. 2003; Kolpak et al. 2003; Anderson & Bania 2009; Roman-Duval et al. 2009; Urquhart et al. 2012) and the H I self-absorption technique (Jackson et al. 2002; Roman-Duval et al. 2009; Green & McClure-Griffiths 2011).

5.1.1 H II absorption against a continuum source

This method is the most reliable and works by identifying H I absorption features towards strong continuum sources such as H II regions (see Anderson & Bania 2009 for details and comparison with the H I self-absorption method). The technique works on the principle that H I gas associated with the molecular clouds that contain an H II region will produce an absorption dip at the same velocity as the cloud. There is a high likelihood that numerous H I clouds will lie along the line of sight towards a distant continuum source and

so H I spectra towards sources at far distances will show absorption features at all intervening velocities, including at or near the tangent velocity. On the other hand, H I spectra towards H II regions at near distances will show only absorption features at velocities between the velocity of the H II region and the observer.

We identified 22 sources for which this method could be reliably employed and followed the method defined by Kolpak et al. (2003) and the procedures described in Urquhart et al. (2012). We provide the continuum images and 21-cm emission profiles in Fig. 11 and give the assigned distance in Table 3.

5.1.2 H II self-absorption

This technique is similar to the absorption against a continuum method in that objects located at the near distance are associated with extended envelopes of cool H I that absorb the background emission from warmer H I in the interstellar medium at the same radial velocity as the target. However, for sources located at the far distance there is effectively no background H I to absorb and no absorption signature is seen. This method can be used for the H II regions that do not produce sufficient continuum emission to

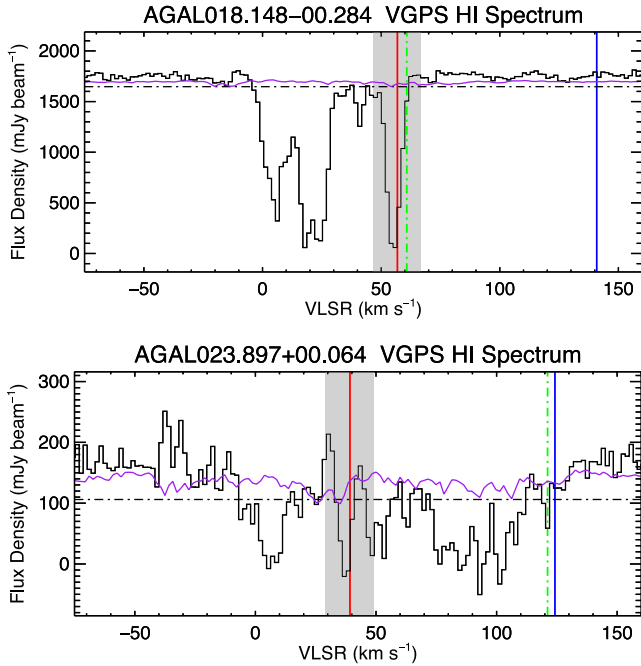


Figure 11. Source-averaged continuum-included H I spectra extracted from the SGPS and VGPS archives towards two H II regions for which a distance is not available in the literature. In the upper and lower panels, we provide an example of a source located at the near and far distances, respectively. The source velocity (v_s), the velocity of the tangent point (v_t) and the position of the first absorption minimum (v_a) are shown by the red, blue and green vertical lines, respectively. The grey vertical band covers the velocity region 10 km s^{-1} either side of the source velocity and is provided to give an indication of the uncertainty associated with it due to streaming motions. The dotted horizontal line shows the $4\sigma_{\text{rms}}$ noise level determined from absorption-free parts of the spectra. The emission profiles of sources located at the far distance will show absorption dips at all velocities up to the tangent velocity.

be detected in the VGPS/SGPS data but is inherently less secure as there are frequently multiple absorption components, most of which are extremely weak against the H I background (~ 80 per cent reliability; Busfield et al. 2006, see also discussion in Anderson & Bania 2009). However, in the absence of more accurate distance measurements and strong continuum emission it is necessary to use this method.

A comprehensive description of this method can be found in Jackson et al. (2002) and Roman-Duval et al. (2009). We provide the H I profiles in Fig. 12 and give the assigned distance in Table 3. Using this method, we are able to resolve the distance ambiguities for 23 of the remaining 24 sources. The only source we have been unable to resolve the distance to is AGAL017.986+00.126.

5.2 Distance distribution

Using a combination of maser parallax, spectroscopic and kinematic distances drawn from the literature and discussed in the previous two subsections we have compiled distances for 169 of the 170 clumps identified in Section 3.1.

Fig. 13 shows the distribution of H II regions (grey filled histogram) and their host clumps (red hatched histogram) as a function of heliocentric distance. The two samples show a similar distribution with five peaks located at 2, 5, 10, 12 and 16 kpc. These roughly coincide with the distances of, respectively, the near side

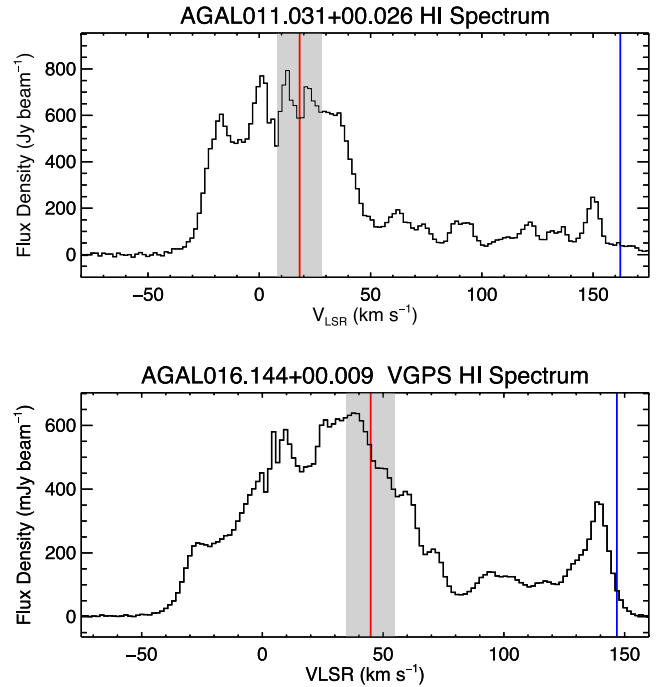


Figure 12. Example H I spectra extracted from the SGPS and VGPS archives. In the upper and lower panels, we show an example of a source located at the near and far distance, respectively. The vertical red and blue lines are as described in Fig. 11.

of the Sagittarius arm, the end of the bar and Scutum–Centaurus arm, the far sides of the Sagittarius and Perseus arms (these two structures are somewhat blended) and the Norma spiral arm. The two distributions differ in source counts at ~ 11 kpc; this is the result of multiple H II regions associated with a single clump where more intense star formation is taking place. The large difference in number of H II regions and clumps seen in this bin is almost entirely due to the W49A star-forming complex.

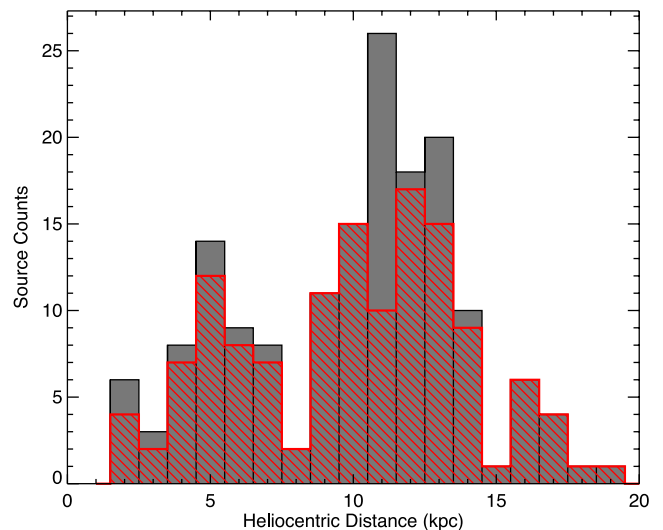


Figure 13. Heliocentric distance distribution for the 212 H II regions and the 169 associated clumps are shown by the filled grey and red hatched histograms, respectively. The strong peak seen in the H II region distribution at ~ 11 kpc is due to the presence of the W49A star-forming complex in this bin. The bin size used is 1 kpc.

The distribution has a mean and median distance of 9.2 ± 0.3 and 9.8 kpc, respectively, and therefore the majority of the sample is located on the far side of the tangent position. This is consistent with the peak in the latitude source counts being slightly positive as discussed in Section 3.4, and their apparent anticorrelation with regions of intense CO emission discussed in Section 4.2.

6 CLUMP PROPERTIES

In this section, we will concentrate on the physical properties of the molecular clumps within which the compact and UC H II regions are embedded. We presented a detailed description of the derivation of many of these parameters and the associated caveats in Paper I and so we only provide a brief outline here. The parameter values for individual clumps are presented in Table 3, while in Table 2 we summarize the global properties of the sample.

6.1 Clump radius

We have estimated the physical radius of the each clump using the heliocentric distances derived in the previous section and their effective radius. In Fig. 14, we show the distribution of effective radii for the H II-region associated clumps (grey histogram) and the methanol-maser associated clumps discussed in Paper I (blue histogram). The dashed vertical lines mark radii of 0.15 and 1.25 pc, indicating the nominal boundary between cores and clumps, and clumps and clouds, respectively (e.g. Bergin & Tafalla 2007). However, as also noted in Paper I, neither sample of clumps shows a sharp break in its distribution at these sizes and so these definitions are probably somewhat arbitrary. The clumps in this sample appear to be rather spherical in structure, are centrally condensed and the vast majority are associated with a single H II region. It is likely these clumps are in the process of forming single clusters of stars and we will therefore continue to refer to them as clumps.

In general, the radii of the clumps hosting H II regions are significantly larger than those associated with methanol masers (me-

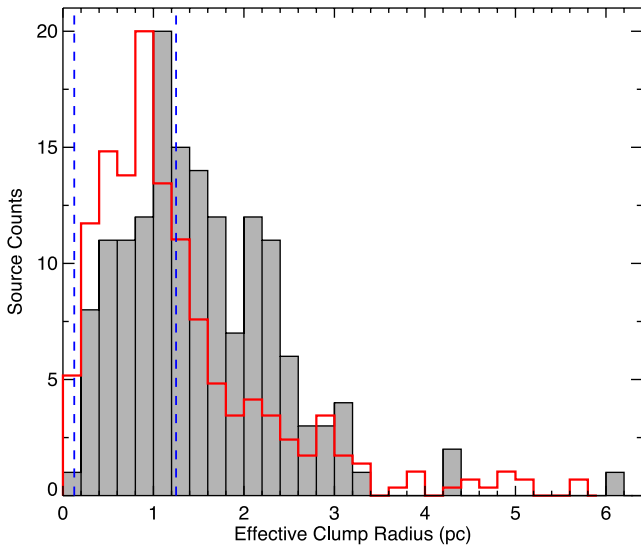


Figure 14. Distribution of the effective radii of the H II-region associated clumps (filled grey histogram) and that of methanol-maser associated clumps (red). The methanol-maser distribution has been scaled to the peak of the H II-region associated clump distribution. The dashed vertical lines indicate the radii separating cores from clumps (0.125 pc), and clumps from clouds (1.25 pc), respectively. The bin size is 0.2 pc.

dian vales of ~ 1.38 and 0.97 pc, respectively). The H II-region and methanol-maser samples of massive star-forming clumps are largely drawn from different quadrants of the Galaxy (first and fourth quadrants, respectively). It is likely that at least some of the difference in physical size is simply a result of the different spiral structures dominating these quadrants. To test this we performed a KS test on a distance-limited sample of methanol and H II-region associated clumps located between 4 and 6 kpc and find that we are unable to reject the null hypothesis that these two samples are drawn from the same parent population (p -value = 0.4).

The distributions of both samples peak at radii of ~ 1 pc and it is therefore likely that the embedded H II regions and methanol masers are tracing the most massive member of a young protocluster. The derived clump properties, bolometric luminosities and Lyman continuum fluxes are therefore likely to be related to the embedded cluster rather than a single star and we need to bear this in mind when interpreting these results.

6.2 Isothermal clump masses and column densities

We estimate isothermal clump masses following the method of Hildebrand (1983) and assuming that the total clump mass is proportional to the total flux density integrated over the source such that

$$\left(\frac{M_{\text{clump}}}{M_{\odot}} \right) = \left(\frac{D}{\text{kpc}} \right)^2 \left(\frac{S_{\nu}}{\text{mJy}} \right) \frac{R}{B_{\nu}(T_{\text{dust}}) \kappa_{\nu}}, \quad (1)$$

where S_{ν} is the integrated 870 μm flux, D is the heliocentric distance to the source, R is the gas-to-dust mass ratio (assumed to be 100), B_{ν} is the Planck function for a dust temperature T_{dust} and κ_{ν} is the dust absorption coefficient taken as $1.85 \text{ cm}^2 \text{ g}^{-1}$, which was derived by Schuller et al. (2009) by interpolating to 870 μm from table 1, column 5 of Ossenkopf & Henning (1994).

For consistency with the masses determined in Paper I, we assume a dust temperature of 20 K. Although we have seen in Section 4.3 the gas temperatures range between 20 and 30 K, the ammonia observations were made towards the peak submm emission, and given that the majority of the dust emission comes from the cooler extended envelope (i.e. mean Y -factor ~ 5 ; see Section 6.1 for details), a lower kinetic temperature is considered more appropriate. Moreover, the choice of 20 K facilitates comparison of the derived masses with similar studies that have used the same temperature (e.g. Hill et al. 2005; Motte et al. 2007).

We estimate column densities from the peak flux density of the clumps using the following equation:

$$\left(\frac{N_{\text{H}_2}}{\text{cm}^{-2}} \right) = \left(\frac{S_{\nu}}{\text{mJy}} \right) \frac{R}{B_{\nu}(T_{\text{dust}}) \Omega \kappa_{\nu} \mu m_{\text{H}}}, \quad (2)$$

where Ω is the beam solid angle, μ is the mean molecular weight of the gas, which we assume to be equal to 2.8 (Kauffmann et al. 2008), m_{H} is the mass of the hydrogen atom, and κ_{ν} and R are as previously defined. We again assume a dust temperature of 20 K.

The uncertainties in the derived clump mass and column density are both ~ 70 per cent. When estimating this error, we assume an uncertainty of ± 5 K for the gas temperature and an ~ 10 per cent error in the distance, which we add in quadrature with the absolute flux measurement uncertainty (15 per cent; Schuller et al. 2009). Although quite large, these errors are unlikely to have a significant impact on the overall distribution or the statistical analysis of these parameters.

In the left-hand panel of Fig. 15, we present the isothermal gas mass distribution for both the H II-region and methanol-maser

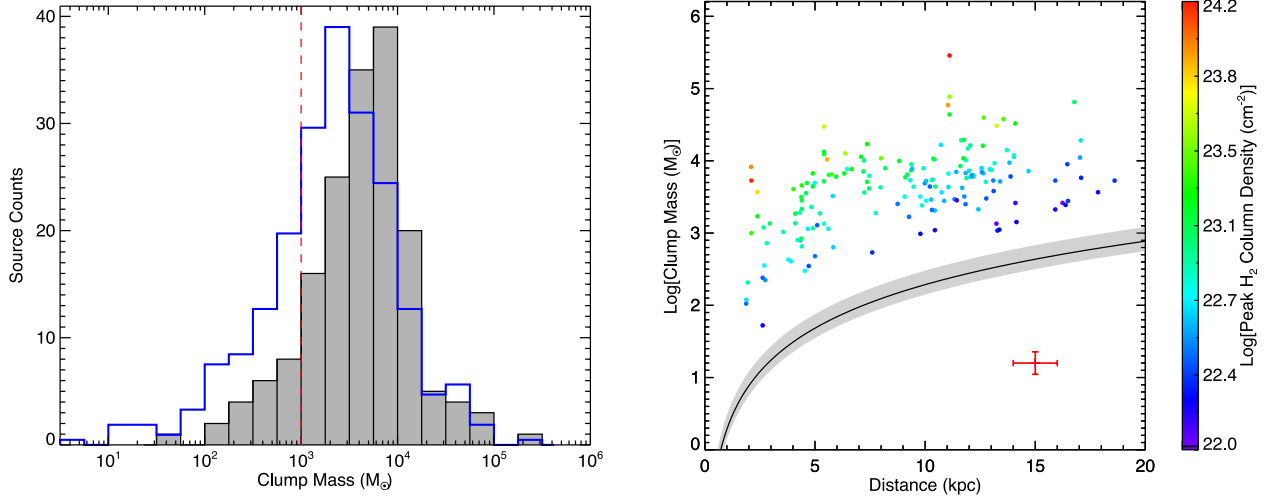


Figure 15. Left-hand panel: the isothermal clump-mass distribution of the H II-region and the methanol-maser associated clumps, shown as filled grey and blue histograms, respectively. A dust temperature of 20 K has been assumed for both samples. The vertical red line indicates the completeness limit (see the text for details). The bin size is 0.5 dex. Right-hand panel: mass distribution of H II-region associated clumps as a function of heliocentric distance. The colours indicate the peak column density of each source (see the colour bar on the right for values) and the solid black line and the grey filled region indicated the mass sensitivity limit of the survey and its associated uncertainty assuming a dust temperature of 20 ± 5 K. Characteristic error bars for these parameters are shown in the lower-right corner of the plot.

associated clumps, while in the right-hand panel we show the mass distribution for only the H II-region associated clumps as a function of heliocentric distance. It is clear from the latter that we are sensitive to all of the H II-region associated clumps with masses above $1000 M_{\odot}$ across the Galaxy within the observed region, and our statistics should be complete above this level; this was also the case for the methanol-maser associated clumps. (This completeness limit is indicated on the left-hand panel of Fig. 15 by the vertical dashed red line.)

There are a small number of H II regions that are associated with clumps below the mass expected for the formation of a massive star (i.e. $< 1000 M_{\odot}$), which may indicate that the distances to these sources have been incorrectly assigned. However, checking the distance assignments for these lower mass sources ($< 300 M_{\odot}$) we find them to be correct given the quality of the data at hand (mostly VGPS data with an angular resolution of ~ 1 arcmin). We note that these lower mass clumps are also the most compact ($R_{\text{eff}} < 0.2$ pc) and therefore may be forming individual or small groups of stars where the IMF is not expected to apply.

The compact and UC H II associated clump-mass distribution has a mean and median value of $\sim 10\,000$ and $\sim 5000 M_{\odot}$, respectively, which are both significantly larger than the completeness limit (see the left-hand panel of Fig. 15) and so the drop off in source counts between the completeness limit and the peak is very likely to be a genuine feature of the distribution. We also note that there are significant differences between the mass distributions of the methanol-maser and H II-region associated clumps. The masses of the methanol-maser associated clumps peak at a few thousand M_{\odot} , which is also above the completeness limit, but a factor of 2 lower than found for the H II-region associated clumps. However, as discussed in the previous subsection these two samples are drawn from different regions of the Galaxy and have different distance distributions as a result.

To mitigate this possible distance bias we again perform a KS test on a distance-limited sample of H II-region and methanol-maser associated clumps (between 4 and 6 kpc); this reveals no significant difference in the mass distribution of the two samples of clumps.

We therefore conclude that the differences in the mass and radii distributions is likely to be a result of the different Galactic structure present in the first and fourth quadrants from which these samples are primarily drawn (e.g. Georgelin & Georgelin 1976).

6.3 Virial mass

To assess the gravitational stability of these clumps, we compare thermal masses and virial masses for the 123 resolved clumps towards which ammonia (1,1) emission is been detected using

$$\left(\frac{M_{\text{vir}}}{M_{\odot}}\right) = \frac{5}{8 \ln 2} \frac{1}{a_1 a_2 G} \left(\frac{R_{\text{eff}}}{\text{pc}}\right) \left(\frac{\Delta v_{\text{avg}}}{\text{km s}^{-1}}\right)^2, \quad (3)$$

where R_{eff} is the effective radius of the clump, Δv_{avg} is the velocity dispersion in the gas calculated from the measured line width via equation (5), G is the gravitational constant and a_1 and a_2 are corrections for the assumptions of uniform density and spherical geometry, respectively (Bertoldi & McKee 1992). For aspect ratios less than 2, the value $a_2 \sim 1$, which is suitable for this sample of clumps, and the value of a_1 is given by

$$a_1 = \frac{1 - p/3}{1 - 2p/5} \text{ for } p < 2.5, \quad (4)$$

where p is the power-law index of the density profile where $[n(r) \propto r^{-p}]$. We have adopted a value for p of 1.6 (i.e. $a_1 = 1.3$) as determined for a similar sample of massive star-forming clumps by Beuther et al. (2002).

The measured ammonia line width needs to be corrected to estimate the average velocity dispersion of the total column of gas. Following Fuller & Myers 1992,

$$\Delta v_{\text{avg}}^2 = \Delta v_{\text{corr}}^2 + 8 \ln 2 \times \frac{k_b T_{\text{kin}}}{m_{\text{H}}} \left(\frac{1}{\mu_p} - \frac{1}{\mu_{\text{NH}_3}}\right), \quad (5)$$

where Δv_{corr} is the observed NH₃ line width corrected for the resolution of the spectrometer (i.e. $\Delta v_{\text{corr}}^2 = \Delta v_{\text{obs}}^2 - \Delta v_{\text{channel width}}^2$; 0.4 and 0.7 km s⁻¹ for the GBT and Effelsberg observations, respectively), k_b is the Boltzmann constant, T_{kin} is the kinetic temperature

of the gas (again taken to be 20 K) and μ_p and μ_{NH_3} are the mean molecular mass of molecular hydrogen and ammonia, respectively; these are taken as 2.33 (i.e. Fuller & Myers 1992) and 17, respectively. The error in the virial mass is of the order of 20 per cent allowing for an ~ 10 per cent error in the distance, the fitted line width, and spectrometer channel width and the error in the measured source size.

In calculating the virial mass, we are making the assumption that the observed NH_3 line width is representative of the whole clump; however, Zinchenko, Henning & Schreyer (1997) found that the line width decreases towards the edges of the clumps hosting embedded protostellar objects. This may lead to the virial mass being overestimated (see Dunham et al. 2011a for a more comprehensive discussion).

In Table 3, we present the estimated virial mass for each source and in Fig. 16 we plot the virial ratio ($M_{\text{clump}}/M_{\text{vir}}$) versus the clump mass for the subsample of sources with ammonia data. The dashed red horizontal line marks the line of gravitational stability. Clumps located below this line are likely to be unbound while those above this should may be unstable against gravity. The clumps have a mean and median virial ratio values of 3.45 ± 0.30 and 3.35, respectively, which are both significantly above unity.

However, the equation used to estimate the virial mass does not take the magnetic field strength into account, which can provide a significant amount of magnetic support to counteract gravity. The magnetic field has not been measured; however, if we assume an equipartition of energy between the kinetic and magnetic energy, we find that $M_{\text{clump}} = 2M_{\text{vir}}$ (e.g. Bertoldi & McKee 1992). The dashed blue horizontal line shown on Fig. 16 indicates the virial ratio for gravitational stability when magnetic support is included. We might therefore expect clumps to be unbound if their virial ratio were $\lesssim 2$; however, even allowing for an equal amount of magnetic support, we find the mean and median values for the virial ratio are significantly larger than 2, which would suggest that the majority of clumps are likely to be unstable against gravity and collapsing.

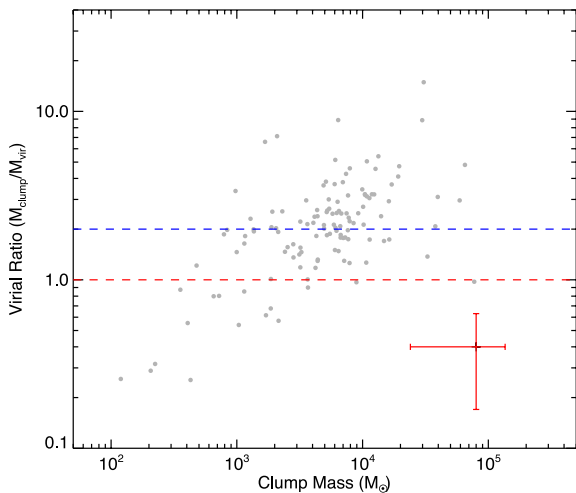


Figure 16. Virial ratio ($M_{\text{clump}}/M_{\text{vir}}$) as a function of clump mass (M_{clump}) for the subsample of 121 clumps detected in ammonia (see Section 4.1 for details). The red horizontal lines indicate the locus of gravitational equilibrium for thermal and kinematic energies, while the blue dashed lines indicates the locus of gravitational equilibrium assuming an equipartition of kinetic and magnetic energy. Characteristic error bars for these parameters are shown in the lower-right corner of the plot.

7 H II REGION PROPERTIES

In this paper, we will only investigate the relationships between the ionization rates and bolometric luminosities of the embedded H II regions and the masses of the surrounding clumps, plus the overall Galactic distribution of the sample. The morphologies, infrared and radio properties of the H II regions themselves will be presented in two future publications (i.e. Purcell et al., and Hoare et al., both in preparation). The physical properties of the individual H II regions are presented in Table 4 and a summary of their overall statistics is presented in Table 2.

7.1 Lyman continuum flux

The Lyman continuum output rate from a massive star, N_i , can be estimated using equation 7 from Carpenter, Snell & Schloerb (1990) relating N_i to the measured radio continuum flux density, i.e.

$$\left(\frac{N_i}{\text{photon s}^{-1}} \right) = 9 \times 10^{43} \left(\frac{S_\nu}{\text{mJy}} \right) \left(\frac{D}{\text{kpc}} \right)^2 \left(\frac{\nu^{0.1}}{5 \text{ GHz}} \right), \quad (6)$$

where S_ν is the integrated radio flux density measured at frequency ν and D is the heliocentric distance to the source. This assumes that the H II regions are optically thin and will significantly underestimate the Lyman flux for more compact H II regions that are optically thick at 5 GHz. The typical error in the derived Lyman flux is ~ 20 per cent and takes into account the 10 per cent error in both the distance and integrated flux measurement. Using the CORNISH 7σ limit of ~ 3 mJy and maximum distance of 20 kpc we find, we are complete to unresolved optically thin H II regions with Lyman fluxes of $\geq 10^{47}$ photon s^{-1} , which corresponds to a star of mass $15 M_\odot$ (i.e. a star of spectral type B0 or earlier). In the case of hypercompact (HC) H II regions, where the ionized nebula is likely to be optically thick below ~ 50 GHz (e.g. Kurtz 2005), the measured flux would be two orders of magnitude lower at 5 GHz. We are therefore only sensitive to HC H II regions associated with stars with masses of $\geq 40 M_\odot$, equivalent to a spectral type of O6–O5 or earlier at a distance of 20 kpc.

In the left-hand panel of Fig. 17, we present the Lyman photon flux as a function of heliocentric distance. The right-hand panel of this figure shows the distribution for the whole sample with the completeness limit indicated by the red vertical dashed line. A peak in the Lyman flux distribution occurs at approximately 10^{48} photon s^{-1} , significantly above the completeness limit for unresolved, optically thin H II regions (i.e. 10^{47} photon s^{-1}). However, given a typical IMF, we should expect the peak in the distribution to coincide with the completeness level. This would suggest that we are missing a significant number of H II regions with Lyman fluxes between 10^{47} and 10^{48} photon s^{-1} .

The colours of the symbols in the left-hand panel of Fig. 17 indicate the observed angular sizes of each H II region (see the horizontal colour bar for corresponding source diameters). While the solid black curve shows the point source sensitivity of the CORNISH survey, it is clear that many of the H II regions are resolved. As an UC H II region expands, its total flux density is conserved but its surface brightness decreases. We have therefore added the dash-dotted lines to show the sensitivity to more extended sources and this shows that we are not sensitive to H II regions with $N_i < 10^{47}$ photon s^{-1} that are larger than 12, 6 and 3 arcsec at distances greater than 2, 5 and 12 kpc, respectively. For H II regions with $N_i < 10^{48}$ photon s^{-1} , we are not sensitive to sources larger than 12 and 6 arcsec at more than 7 and 15 kpc, respectively, but are sensitive to UC H II regions with sizes of 3 arcsec or smaller across the Galaxy.

Table 4. Derived H II region parameters.

CORNISH name	ATLASGAL name	Offset (arcsec)	Peak flux (mJy beam ⁻¹)	Integrated flux (mJy)	Diameter (arcsec)	Diameter (pc)	Log(N_{Lyman}) (photon s ⁻¹)	Log(L_{bol}) (L_{\odot})
G010.3009–00.1477	AGAL010.299–00.147	6.2	56.37	631.39	5.24	0.06	47.51	–
G010.3204–00.2586	AGAL010.321–00.257	4.1	14.61	18.20	0.74	0.01	46.31	3.11
G010.4724+00.0275	AGAL010.472+00.027	0.5	22.34	38.43	1.67	0.09	47.62	4.82
G010.4736+00.0274	AGAL010.472+00.027	4.1	12.30	19.30	1.36	0.07	47.33	–
G010.6218–00.3848	AGAL010.624–00.384	8.8	16.81	37.06	1.73	0.02	46.28	–
G010.6223–00.3788	AGAL010.624–00.384	20.5	97.90	483.33	5.56	0.06	47.40	–
G010.6234–00.3837	AGAL010.624–00.384	3.2	305.58	1952.22	4.39	0.05	48.00	3.90
G010.6240–00.3813	AGAL010.624–00.384	10.4	38.23	71.65	1.45	0.02	46.57	–
G010.9584+00.0221	AGAL010.957+00.022	3.6	109.11	195.97	1.61	0.11	48.52	5.36
G010.9656+00.0089	AGAL010.964+00.011	8.7	7.86	51.75	4.65	0.06	46.55	3.55
G011.0328+00.0274	AGAL011.031+00.026	9.2	3.58	5.69	1.15	0.01	45.55	–
G011.0339+00.0616	AGAL011.034+00.061	3.0	21.01	103.36	6.93	0.49	48.30	5.14
G011.1104–00.3985	AGAL011.109–00.397	6.0	51.55	305.37	8.22	0.67	48.89	5.68
G011.9032–00.1407	AGAL011.902–00.141	3.6	5.59	25.57	2.84	0.06	46.59	–
G011.9039–00.1411	AGAL011.902–00.141	6.3	11.11	16.81	1.07	0.02	46.41	–
G011.9368–00.6158	AGAL011.936–00.616	4.5	163.93	1155.90	5.69	0.11	48.23	4.77
G011.9446–00.0369	AGAL011.946–00.036	5.1	61.58	943.58	14.56	0.89	49.13	5.50
G012.1988–00.0345	AGAL012.198–00.034	4.0	19.68	62.71	2.22	0.13	47.90	4.64
G012.2081–00.1019	AGAL012.208–00.102	2.5	72.18	207.87	2.41	0.16	48.54	–
G012.4294–00.0479	AGAL012.431–00.049	7.2	13.78	45.17	2.26	0.15	47.89	–
G012.4317–01.1112	AGAL012.431–01.114	9.8	9.67	69.01	4.13	0.08	47.02	–
G012.8050–00.2007	AGAL012.804–00.199	6.2	287.90	12 616.40	16.16	0.16	48.70	–
G012.8131–00.1976	AGAL012.804–00.199	32.5	131.04	1500.39	5.22	0.05	47.77	–
G012.9995–00.3583	AGAL012.998–00.357	7.8	4.76	20.14	2.70	0.03	45.84	–
G013.2099–00.1428	AGAL013.209–00.144	5.6	40.72	946.76	8.21	0.18	48.25	4.52
G013.3850+00.0684	AGAL013.384+00.064	15.3	10.49	603.94	19.12	0.17	47.28	3.53
G013.8726+00.2818	AGAL013.872+00.281	3.5	24.71	1447.55	15.36	0.33	48.40	–
G014.2460–00.0728	AGAL014.246–00.071	7.3	11.40	51.26	3.64	0.20	47.79	–
G014.5988+00.0198	AGAL014.607+00.012	40.6	2.89	4.39	1.08	0.01	45.49	3.28
G014.7785–00.3328	AGAL014.777–00.334	6.4	11.02	18.25	1.75	0.11	47.45	4.43

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/>.

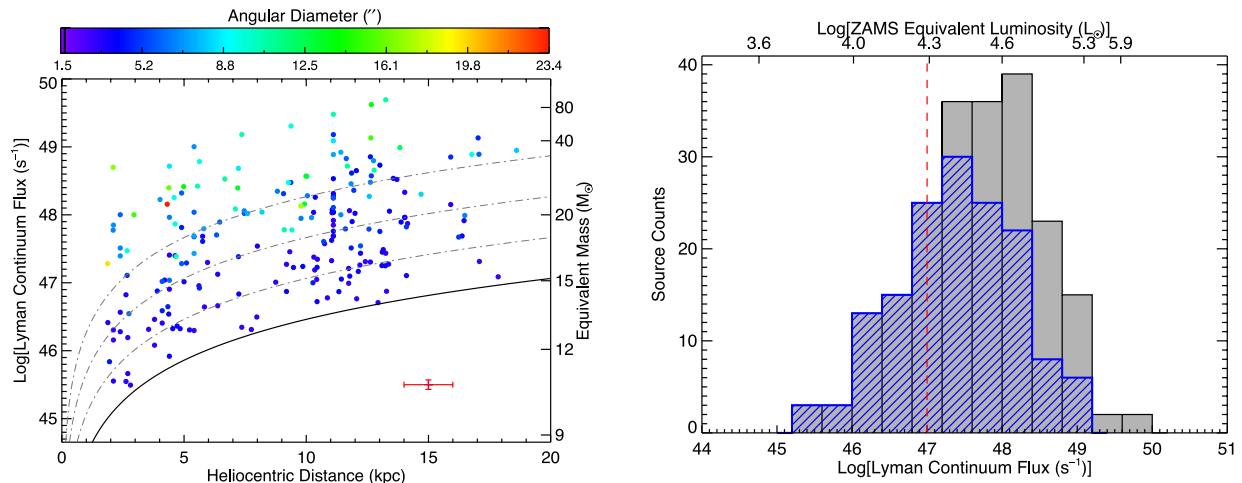


Figure 17. Left-hand panel: Lyman continuum flux as a function of heliocentric distance. The colours correspond to the angular sizes of the H II regions (see colour bar for approximate values) and the solid and dash-dotted curved lines indicate the CORNISH sensitivity to different angular size scales; these correspond to FWHM of 1.5, 3, 6 and 12 arcsec, respectively. Characteristic error bars for these parameters are shown in the lower-right corner of the plot. Right-hand panel: (grey histogram) the ionizing photon flux distribution of the whole ATLASGAL–CORNISH matched sample of embedded H II regions; (blue histogram) the same, but for a subset of sources with sizes < 5 arcsec. The dashed red vertical line indicates the CORNISH completeness level. The bin size is 0.4 dex.

It is clear that we are less sensitive to more extended H II regions and that this has a bigger impact on the nearby early B-type star population of the sample than the generally more distant O-type star population. To illustrate this, we overlaid the N_i distribution of the

more compact H II regions (< 5 arcsec) on the right-hand panel of Fig. 17 (blue hatched histogram). This shows that the distribution of the most compact H II regions peaks nearer the completeness limit than that of the full sample. It is this more limited sensitivity to

the less compact H II regions with ionizing fluxes below 10^{48} photon s^{-1} that produces the higher than expected turnover for the whole sample seen in the right-hand panel of Fig. 17.

Another interesting feature arising from Fig. 17 is that the proportion of compact to less-compact H II regions decreases steeply as the ionizing photon flux increases.

7.2 Physical sizes

In Fig. 18, we show the physical size distribution for this sample of H II regions (grey histogram). Additionally, we separate this sample into two subsamples using an N_i threshold of 10^{48} photon s^{-1} . This is approximately the ionizing photon output rate expected from a $\sim 20 M_{\odot}$ zero-age main-sequence (ZAMS) star (Martins, Schaerer & Hillier 2005; Davies et al. 2011; Mottram et al. 2011b), roughly equivalent to an O9 spectral type and thus effectively separates late O- and early B-type star populations. For simplicity, we will refer to these two subsamples as O- and B-type H II regions. The distribution of H II regions above and below this threshold are plotted in Fig. 18 as red and blue hatched histograms, respectively.

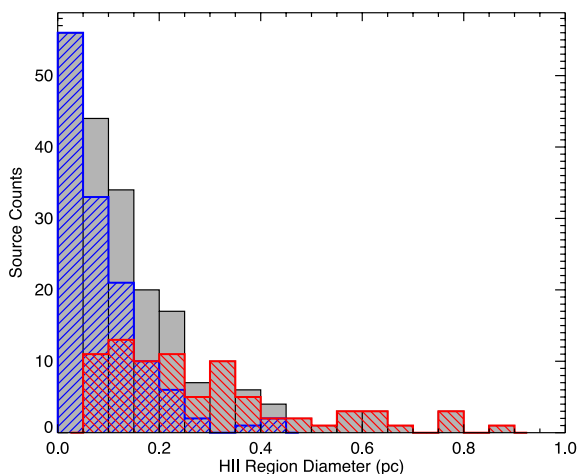


Figure 18. Distribution of H II region sizes. The whole sample is traced by the filled grey histogram, while the red and blue histograms show the H II regions that have Lyman continuum fluxes above and below 10^{48} photon s^{-1} , respectively. The bin size used is 0.05 pc.

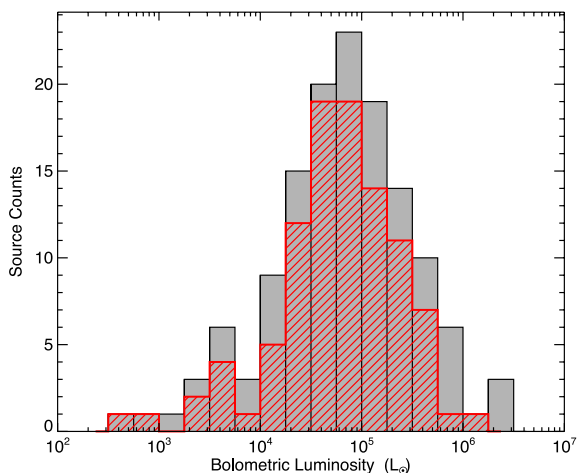


Figure 19. Bolometric luminosity distribution of the H II region sample. The bin size is 0.25 dex.

The overall distribution of H II region sizes shows a peak for diameters less than 0.05 pc, after which it decreases rapidly before flattening into a long tail at ~ 0.4 pc; the largest measured size being ~ 0.8 pc. The classification as a UC H II region is generally reserved for sources with diameters less than 0.1 pc and applying this to sample we find that only 100 H II regions satisfy this requirement (slightly less than 50 per cent). Although the overall distribution looks fairly smooth, in that there are no obvious breaks or evidence of bimodality, the size distributions of O- and B-type H II regions show some striking differences.

The B-type H II region size distribution peaks at the same position as that of the whole sample, but drops off much more quickly, with the largest size measured being ~ 0.4 pc. There are 131 H II regions in this subsample with the majority falling into the UC H II region classification (89 or ~ 68 per cent). The lower number of more extended B-type H II regions can be understood by considering the relationship between the angular size of the H II region and the surface brightness sensitivity of CORNISH as discussed in the previous subsection.

The distribution of O-type H II regions is significantly flatter and more skewed towards larger diameters, with a peak at 0.0–0.05 pc. This subset provides the majority of the more extended H II regions (> 0.2 pc). Noticeably, there are no O-type H II regions smaller than 0.05 pc and only 11 of the 81 H II regions in this sample could be classified as UC H II regions (~ 14 per cent). Due to the lower sensitivity of CORNISH to more extended H II regions, particularly those associated with B-type stars, we are likely to be incomplete to these types of H II regions. However, we should be sensitive to all optically thin UC H II regions driven by O stars across the Galaxy and therefore the low number of these in the sample is significant (this is discussed in more detail in Section 8.3). Additionally, although not explicitly shown in Fig. 18, we find no compact or UC H II regions with $N_i > 10^{49}$ photon s^{-1} and physical diameters smaller than 0.2 pc.

7.3 Bolometric luminosity

Many of our sample of H II regions are also found in the RMS catalogue (Urquhart et al. 2008a) and therefore have estimated bolometric luminosities (i.e. Mottram et al. 2010, 2011a). The RMS survey has characterized the Galactic population of mid-infrared-selected H II regions and MYSOs and is complete to a few $10^4 L_{\odot}$. This is well matched to the CORNISH completeness and makes the RMS sample well suited for comparison.

Using a search radius of 10 arcsec, we have matched 135 H II regions with their RMS counterparts and adopted the bolometric luminosities from the RMS data base.³ The fluxes used to estimate these luminosities are effectively clump-average values and are therefore a measure of the total luminosity of the embedded protocluster.

The bolometric luminosities for H II regions and associated clusters are shown in Fig. 19 and range from $\sim 1000 L_{\odot}$, which is effectively the lower limit required for a B3 star, to $10^6 L_{\odot}$, which is approximately equivalent to an O3 star (Martins et al. 2005). These stellar classes have been estimated assuming that a single star is responsible for the measured luminosity but in reality the luminosity of the most massive star will be a few subclasses later than those stated. Mottram et al. (2011a) estimates the typical errors for the derived luminosities to be approximately 34 per cent.

³ http://rms.leeds.ac.uk/cgi-bin/public/RMS_DATABASE.cgi.

8 DISCUSSION

8.1 Empirical mass–size relationship for massive star formation

In Fig. 20, we reproduce the clump mass–radius plot from fig. 15 of Paper I, with the addition of the H II-region data presented in this paper. The dashed red and blue lines show the least-squares fits to the H II-region and methanol-maser associated clump samples, respectively. The similarity between their overall distributions and the fit parameters, which agree within the errors, provides strong evidence that the methanol-maser sources and the H II regions form in the same population of dense clumps. The two data sets provide us with a combined sample of over 500 massive, star-forming clumps, forming a fairly continuous distribution over almost four orders of magnitude in mass and two orders of magnitude in radius. The least-squares fit to the combined sample is $\text{Log}(M_{\text{clump}}) = 3.4 \pm 0.013 + (1.67 \pm 0.036) \times \text{Log}(R_{\text{eff}})$.

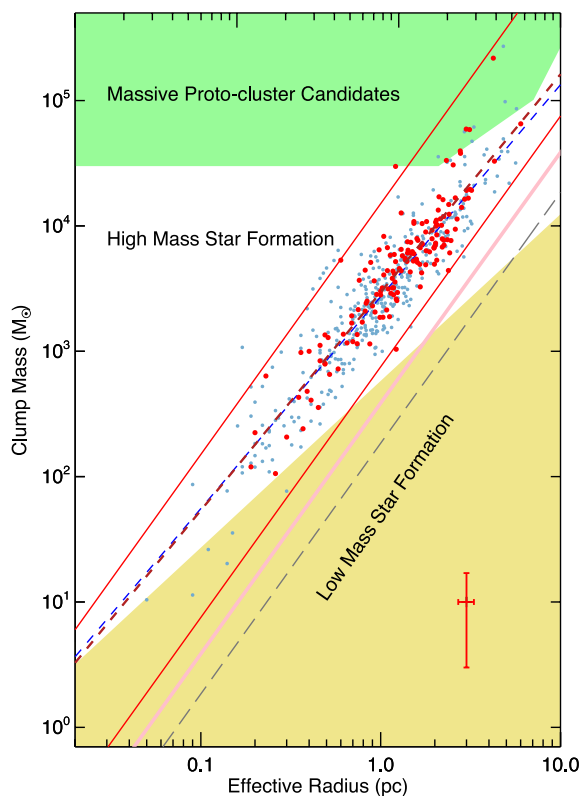


Figure 20. The mass–size relationship of the current H II-region sample (red dots) and methanol-maser associated clumps from Paper I (blue dots). The yellow shaded region shows the part of the parameter space found to be devoid of massive star formation that satisfies the relationship $m(r) \leq 580 M_{\odot} (R_{\text{eff}}/\text{pc})^{1.33}$ (cf. Kauffmann & Pillai 2010). The green shaded region indicates the region in which YMC progenitors are expected to be found (i.e. Bressert et al. 2012). The dashed blue and red lines show the result of linear least-squares fits to the methanol-maser and H II-region associated clumps, respectively. The grey dashed line shows the sensitivity of the ATLASGAL survey and the upper and lower solid red line shows the surface densities of 1 and 0.05 g cm^{-2} , respectively. The diagonal pink band fills the gas surface density ($\Sigma(\text{gas})$) parameter space between 116 and $129 M_{\odot} \text{ pc}^{-2}$ suggested by Lada, Lombardi & Alves (2010) and Heiderman et al. (2010), respectively, to be the threshold for ‘efficient’ star formation. Characteristic error bars for these parameters are shown in the lower-right corner of the plot.

The lower and upper solid diagonal red lines in Fig. 20 denote surface densities of 0.05 and 1 g cm^{-2} , the empirical lower and upper bounds for the clump-averaged surface densities of the methanol-maser associated clumps of Paper I. The upper bound is equal to the value of 1 g cm^{-2} predicted to be a lower limit for massive star formation (i.e. McKee & Tan 2003; Krumholz, McKee & Tumlinson 2008). However, the latter was determined for core-sized structures and is clearly not applicable to the larger clumps resolved by ATLASGAL. These limits also provide reasonable constraints on the H II-region associated clump distribution and thus on the surface density required for massive star formation.

To facilitate the comparison of our results with other studies, we also include the threshold derived by Lada et al. (2010) and Heiderman et al. (2010) (116 and $129 M_{\odot} \text{ pc}^{-2}$, respectively, which approximately corresponds to $A_V \simeq 8$ mag; hereafter LH threshold) for ‘efficient’ star formation (solid pink line) and the mass–size criterion for massive star formation determined by Kauffmann et al. (2010b, i.e. where $m(r) \geq 580 M_{\odot} (R_{\text{eff}}/\text{pc})^{1.33}$; yellow shaded region).⁴ The LH threshold was determined from studies of nearby molecular clouds and is therefore not specific to massive star formation. The Kauffmann et al. (2010b) criterion was determined from an investigation of the mass–radius relationship of nearby molecular clouds ($< 500 \text{ pc}$, i.e. Ophiuchus, Perseus, Taurus and the Pipe Nebula; Kauffmann et al. 2010a,b) and known samples of high-mass star formation such as those of Beuther et al. (2002), Hill et al. (2005), Motte et al. (2007), Mueller et al. (2002) and Kauffmann et al. (2010b). However, as stated in Paper I the lower bound of 0.05 g cm^{-2} is approximately twice the LH-threshold and provides a better constraint than Kauffmann et al. for the high-mass end of the distribution (i.e. $R_{\text{eff}} \gtrsim 0.5 \text{ pc}$ or $M_{\text{clump}} \gtrsim 500 M_{\odot}$).

So while the LH threshold can be considered a reasonable criterion for ‘efficient’ star formation the value of 0.05 g cm^{-2} , found to fit the lower envelope of the mass–radius relationship for both methanol-maser and H II-region associated clumps, can be considered the threshold required for ‘efficient’ high-mass star formation.

The area highlighted in green in the upper portion of Fig. 20 shows the region of parameter space where massive protocluster (MPC) candidates are expected to be located (Bressert et al. 2012; see Paper I for details). Six of our sources are located in this region. Three are associated with W49A (AGAL043.148+00.014, AGAL043.164–00.029 and AGAL043.166+00.011) and have a combined mass $\sim 4 \times 10^5 M_{\odot}$. One clump is associated with each of the W31 and W51 star-forming regions (AGAL010.472+00.027 and AGAL049.489–00.369, respectively). The remaining source is AGAL019.609–00.234. All six have masses above $10^4 M_{\odot}$ and are therefore possible progenitors of future young massive clusters (YMCs) such as the Arches and Quintuplet clusters (Portegies Zwart, McMillan & Gieles 2010). The MPC candidates in W31, W49A and W51 were previously identified by Ginsburg et al. (2012) from an analysis of 1.1 mm continuum data from the BGPS and so

⁴ Note that when deriving this relationship, Kauffmann et al. (2010b) reduced the dust opacities of Ossenkopf & Henning (1994) by a factor of 1.5. This reduced value for the opacities has not been applied when determining the clump masses presented here and therefore we have rescaled the original coefficient given by Kauffmann et al. (2010b) by this factor (i.e. from 870 to $580 M_{\odot}$; cf. Dunham et al. 2011a).

AGAL019.609–00.234 is the only new potential MPC identified by this work.⁵

In Paper I, we identified seven MPC candidates, six of which were new, and together with those identified by Longmore et al. (2012) and Ginsburg et al. (2012) we estimated that the total population of MPC candidates is likely to be $\leq 20 \pm 6$. However, comparing this estimate with the number of known YMCs, we found that there are far more MPCs than would be expected, assuming a lifetime of 10 Myr and formation time of 1 Myr for the YMCs (e.g. Longmore et al. 2012). We suggested a couple of possible explanations for this apparent discrepancy: either there are far more YMCs yet to be discovered or the SFE is significantly lower than the 30 per cent assumed by Bressert et al. (2012) when defining their mass–size criterion. The first of these possibilities would require the total number of YMCs in the Galaxy to be an order of magnitude larger than currently known, which seems unlikely. The second seems more plausible given that an SFE of ~ 10 per cent (e.g. Johnston et al. 2009 and see Section 8.5 of this paper) would reduce the number of MPC candidates to no more than a handful.

8.2 Galactic distribution of compact and UC H II regions

We have already seen some hints of the large-scale structural feature of the Galaxy in the longitude–velocity diagram presented in Section 4.2 and the heliocentric distribution presented in Section 5.2. In this subsection, we will use the distances and Galactic longitude to trace the two-dimensional distribution of compact H II regions for the first quadrant. As these H II regions are still deeply embedded in their dust cocoons and have lifetimes of at most a few times 10^5 yr (Davies et al. 2011; Mottram et al. 2011b), we can assume that they have not travelled far from where they formed. Furthermore, since massive star formation is found to be almost exclusively associated with the spiral arms in nearby spiral galaxies (e.g. Kennicutt 2005), the Galactic distribution of this sample of H II regions may be able to trace the spiral arms of the Milky Way.

In Fig. 21, we present the 2D Galactic distribution of ATLASGAL–CORNISH associations. The positions of the ATLASGAL–CORNISH associations are in reasonable agreement with the main structural features of the Galaxy, as determined by other tracers (taking into account the kinematic distance uncertainty due to peculiar motions is of the order of ± 1 kpc; Reid et al. 2009). We note that there is a large amount of scatter in these data, particularly at the far side of the Sagittarius arm. However, the width of the interarm region is approximately twice the uncertainty in the distance measurement and therefore it is possible, and even likely, that sources located in this interarm region are actually associated with either the Sagittarius or Perseus arms. Two previous studies have performed a similar analysis using methanol masers (Paper I) and a sample of infrared-selected MYSOs and H II regions (Urquhart et al. 2011a), and with both finding a higher degree of correlation with the spiral arms. However, both studies had larger samples and so sample size may be an issue here.

The Galactocentric and latitude distributions of the ATLASGAL–CORNISH matched sources are presented in the upper and lower panel of Fig. 22, respectively. As with the heliocentric distance distribution presented in Fig. 13, we plot the distribution of the

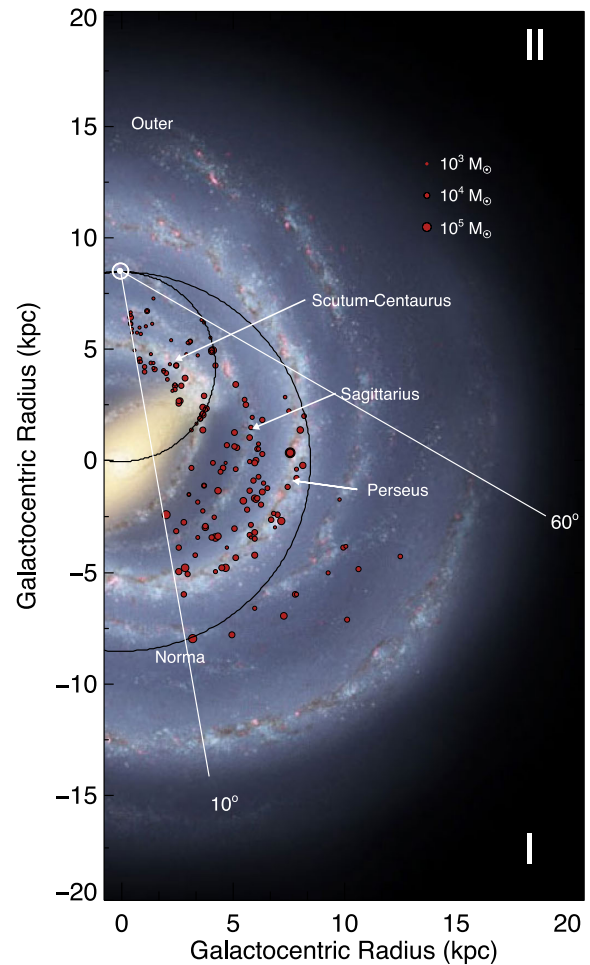


Figure 21. Galactic distribution of the ATLASGAL–CORNISH associations with known distances and masses above the completeness limit ($1000 M_{\odot}$). The Galactic positions of our sample derived from the maser parallax, photometric and kinematic distances and their Galactic longitude are plotted as red circles, the sizes of which give an indication of their mass, as depicted in the upper-right corner. We have superimposed the clump distribution over the sketch of the Galaxy produced by Robert Hurt of the Spitzer Science Center in consultation with Robert Benjamin (Courtesy NASA/JPL-Caltech). The position of the Sun is shown by the small white circle above the Galactic Centre. The two white solid lines enclose the region of the Galactic plane overlapped by the ATLASGAL and CORNISH surveys. The black dot–dashed circles represent the locus of tangent points and the solar circle, respectively.

H II regions and clumps separately (filled grey and red hatched histogram, respectively); this is primarily to evaluate the impact of extreme star formation regions, such as W49A, have on the overall H II region distribution.

Although the association with individual spiral arms is fairly weak in the 2D distribution shown in Fig. 21, the Galactocentric distribution shows three distinct peaks at ~ 4 , 6 and 8 kpc. Peaks at similar radii have been previously noted in the northern Galactocentric radial distribution from a sample of *IRAS* sources observed in CS ($J=2-1$) (Bronfman et al. 2000) and the distribution of the RMS sample of young massive stars and their associated CO emission (Urquhart et al. 2011a; Moore et al. 2012). The first of these corresponds to the position of the Scutum–Centaurus arm tangent and the end of the long bar, which is located at a Galactic longitude of $\sim 25^{\circ}$. The peak at ~ 6 kpc is largely associated with the Sagittarius

⁵ AGAL019.609–00.234 was also identified by Ginsburg et al. (2012) as a massive clump but its derived mass was slightly below $3 \times 10^4 M_{\odot}$ and so it did not make into their final sample. However, the mass derived by us would suggest this source is a good MPC candidate.

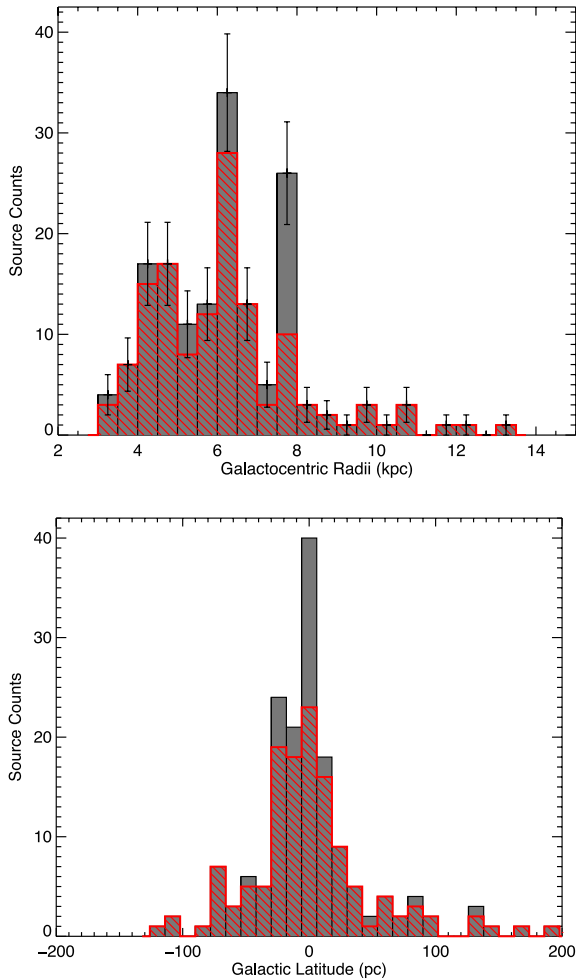


Figure 22. Galactocentric and Galactic latitude distributions of the H II regions (grey) and clumps (red hatching) are shown in the upper and lower panels, respectively. In both of these plots, we have only included UC H II regions with ionizing fluxes and masses above the completeness limit ($N_i > 10^{47}$ photon s^{-1} and $M_{\text{clump}} > 1000 M_{\odot}$; 163 H II regions and 132 clumps). The strong peak seen at ~ 8 kpc is almost entirely due to the presence of W49A in this bin. The bins sizes are 0.5 kpc and 12 pc in the upper and lower panels, respectively.

arm. The third peak at ~ 8 kpc is again almost entirely due to the W49A star-forming complex; however, this peak is still significant and is coincident with the expected Galactocentric radius of the far side of the Perseus arm.

The peaks at 4 and 6 kpc in the Galactocentric radius distribution have also been reported by Anderson & Bania (2009) from a study of compact and diffuse H II regions; however, the relative intensity of the peaks is reversed, presumably due to a larger proportion of nearby sources in their sample. The lack of sources at distances less than 3 kpc is not significant as it simply results from the minimum longitude range used in this study.

An interesting feature of this plot is that the vast majority of the massive star formation is taking place within ~ 8 kpc of the Galactic Centre, with relatively little taking place outside the solar circle. Similarly, significantly lower surface densities of methanol masers (discussed in Paper I), H II regions and MYSOs identified by the RMS survey (Lumsden et al. 2013) have been reported in the outer Galaxy than found inside the solar circle. Since we have selected only sources above our completeness limit, this is unlikely to be a

selection effect and suggests that high-mass clumps are more likely to be found inside the solar circle where the majority of the Galactic molecular material resides. The sharp drop in surface density occurs at approximately the same distance from the Galactic Centre as the corotation radius (~ 8 kpc), which also coincides with a sharp drop in the metallicity gradients of open clusters (Lépine et al. 2011). This drop in the metallicity gradients is interpreted by the authors as being due to a gap in the dense gas at the corotation radius, with smaller gas flows and consequently fewer mergers producing massive clumps outside the corotation radius than inside (see Lumsden et al. 2013 for a more detailed discussion). While spiral arms are thought to play an important role in forming dense molecular structures within the corotation radius (e.g. Heyer & Terebey 1998; Dobbs, Bonnell & Pringle 2006), it is likely that this process is regulated by a different mechanism in the outer Galaxy, such as supernovae (Dib et al. 2009). Alternatively, the lack of massive clumps in the outer Galaxy could simply be due to statistics rather than any physics since there is a power-law mass function for clumps (CMF) and so a smaller population of clumps will have far fewer massive clumps, and none, if the clump population density is below some level (e.g. Snell, Carpenter & Heyer 2002).

Peaks in both the H II region and clump latitude distribution are coincident with the Galactic mid-plane ($|z| \sim 0$). The peak of the H II distribution is noticeably higher than that found for the clumps; however, again this is primarily the presence of W49A in this bin. The presence of a few extreme star-forming regions close to the mid-plane, not only increases the overall peak in the H II region distribution with respect to the clump distribution, but also has a knock on effect to the measured scaleheight. The scaleheight of the H II region and clump samples is 20.7 ± 1.7 and 29.1 ± 3.0 pc, respectively. The scaleheight derived from the methanol-maser distribution reported by Green & McClure-Griffiths (2011) is 27 ± 1 pc, which is in good agreement with the scaleheight derived from our clump distribution.

As seen in the heliocentric distance distribution discussed in Section 5.2 and for the Galactocentric and latitude distributions presented here, it is clear that the presence of a small number of extreme star formation regions, such as W31, W33, W42 and W49A, can have a significant impact of the overall distributions of H II regions.

8.3 Estimating the Galactic population of H II regions and their lifetimes

Our catalogue of H II regions covers a large fraction of the Galaxy and has been compiled from two unbiased surveys.⁶ It should therefore constitute a good representation of the Galactic population as a whole and be able to provide reliable estimates of the number of H II regions. We begin by calculating the surface density of our sample of H II regions as a function of Galactocentric distance (as shown in Fig. 23) by dividing the source counts given in the upper panel of Fig. 22 by the surface area of the portion of each annulus included in survey. Assuming that this surface density is representative of the Galactic population as a whole, we can simply multiply each bin by the area of the corresponding complete annulus to estimate the total number of H II regions in each bin, and then sum all of the bins to get the total population. Using this method, we estimate that

⁶ More precisely this includes ~ 36 per cent of the Galactic surface area within the solar circle and ~ 25 per cent of the area within 15 kpc of the Galactic Centre.

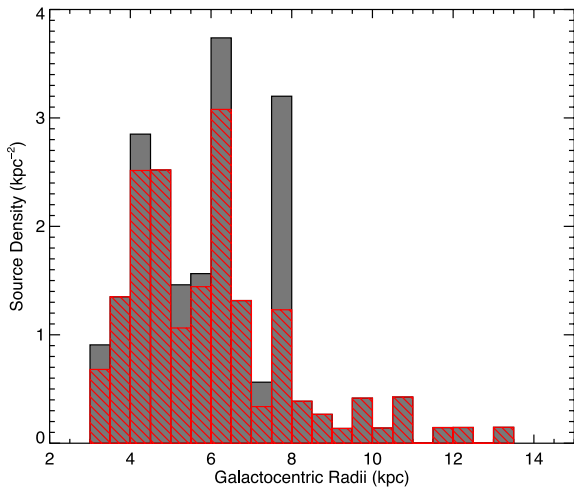


Figure 23. Surface density of H II regions and their host clumps. Colours same as for Fig. 22. The strong peak seen at ~ 8 kpc is almost entirely due to the presence of W49A in this bin. The bin size is 0.5 kpc.

the total number of H II regions around B0 and earlier type stars (the nominal CORNISH sensitivity limit) to be approximately 400. Performing the same procedures using the surface density of O6 stars and earlier would suggest that there are only ~ 45 detectable H II regions of these types in the Galaxy.

The total number of H II regions is slightly lower than the Galactic population estimated by the RMS survey. However, the RMS sample includes some more extended H II regions that have been identified from their mid-infrared morphology and whose radio continuum emission falls below the CORNISH surface brightness sensitivity. More interesting is the low number of H II regions associated with a star of spectral type O6 and earlier. Wood & Churchwell (1989b) estimated the number of embedded O6 stars expected from the H II expansion time and the number of field O stars to be of the order of a few hundred. Comparing this with the number expected from the 1600 candidate UC H II regions identified from their *IRAS* colours led Wood & Churchwell (1989b) to conclude that the embedded phase is much longer than would be expected from an uninhibited expansion of the H II region – this is often referred to as the *lifetime problem*. It is likely that their sample of H II region candidates is significantly contaminated from evolved stars and nearby B stars. Furthermore, it is very likely that given the *IRAS* resolution (ranging from 0.5 to 5 arcmin for the 12–100 μm bands) that many of the luminosities calculated for their sample of UC H II regions have been significantly overestimated. This would reduce the total number of embedded O6 stars, which in turn would reduce the need for an extended embedded UC H II region phase.

Wood & Churchwell (1989a) estimated the UC H II region phase to be $\sim 3\text{--}6 \times 10^5$ yr, which is supported by a Galactic population synthesis analysis (Davies et al. 2011) and estimates of their statistical lifetimes (Mottram et al. 2011b), both of which reported similar H II region lifetimes (a few 10^5 yr). However, we find that only a small proportion of H II regions associated with late O-type stars, and none of those associated with early O-type stars, can be classified as UC H II regions. This may suggest that their expansion is extremely rapid and that the UC H II region lifetime is relatively short for late O-type stars, and possibly so short for early O-type stars that their UC H II region stage is effectively unobservable. Alternatively, the early O-type stars evolve from late O-type stars via accretion of material through their UC H II and compact H II

region (Keto 2007). In this picture, the H II region first forms when the MYSO reaches $\sim 30 M_{\odot}$ (approximately equivalent to a ZAMS O7-type star) and contracts down to the main sequence and expands rapidly while continuing to accrete material. By the time the central star has accreted enough mass to be classified as an early O-type star, the H II region has expanded beyond the sizes of typical of UC and compact H II regions.

Another possibility is that there is a significant number of O-type stars surrounded by optically thick ionized nebulae intermixed with the sample of B-type H II regions; these extremely compact, optically thick H II regions are classified as HC H II regions. The Lyman continuum flux of these HC H II regions may be underestimated by up to two orders of magnitude (see Section 7.1). Although it is conceivable that very high accretion rates might impede the expansion of the H II region, it can do nothing to suppress the infrared luminosity. We might therefore expect to find a number of radio-quiet, embedded objects with bolometric luminosities of $10^5 L_{\odot}$ and higher. However, the fact that the RMS survey (Mottram et al. 2011a,b) has not found any embedded MYSOs more luminous than $\sim 10^5 L_{\odot}$ rules this possibility out.

8.4 Comparison between methanol-maser and H II-region associated clumps

Comparing the derived parameters for the H II-region associated clumps presented in this paper with those for the methanol-maser associated clumps presented in Paper I, we find them to be very similar. The overall clump structure (aspect ratios and Y -factors) for the two samples are indistinguishable from each other. This confirms the finding reported in earlier studies of high-mass protostellar objects (Williams et al. 2004) and UC H II regions (Thompson et al. 2006) that the *envelope* structure does not significantly change during the embedded stages of the massive star's or associated cluster's evolution. This would imply that the clumps are already in place before the star formation begins (as also suggested by Lumsden et al. 2013) and that evolution on clump scales has a much longer time-scale than the formation process for massive stars.

We find no significant difference in the location within clumps of massive stars as traced by H II regions and methanol masers. Both tracers are closely associated with the peak of the submm emission which is likely to be the highest column density. This would place the massive stars at the centres of the clumps at the bottom of the gravitational potential, which is broadly in agreement with the predictions of the competitive accretion model of massive star formation (Bonnell et al. 1997, 2001). However, as mentioned in Paper I, higher resolution observations are required to investigate the internal structure and mass distribution of these clumps before any detailed comparison with the models can be performed. Comparing the mass–radius distribution of these two samples of clumps, we find them to be again indistinguishable from each other, both covering the same narrow band of parameter space and both satisfying the criteria for massive star formation.

The similarities between the properties of clumps associated with methanol masers and H II regions, given that only ~ 20 per cent of the methanol masers are associated with the later stages of formation (i.e. MYSOs and H II regions; Paper I), provide strong support for the hypothesis that methanol masers are almost exclusively associated with the earliest stages of massive star formation. Approximately 80 per cent of the detected methanol masers are not bright enough at mid-infrared to have been detected as a discrete source by *MSX*. These could be deeply embedded accreting protostars in a very early evolutionary stage having not yet warmed their dust envelope

sufficiently to be observed by *MSX*. However, the majority of the methanol masers, although not bright enough in the mid-infrared to have been detected by *MSX*, are detected in *Spitzer* IRAC images (Gallaway et al. 2013). These weaker mid-infrared sources may still be very embedded MYSOs but alternatively could also be more evolved intermediate-mass YSOs that have started to clear their surroundings, lowering the line-of-sight extinction and allowing them to become visible in the mid-infrared.

We also note that the total number of H II regions around stars of stellar type B0 and earlier is similar to the expected number of methanol masers (~ 600 ; Paper I), which would suggest that these two stages have similar statistical lifetimes.

8.5 Correlation between UC H II regions and their natal clumps

8.5.1 Monte Carlo simulation

In the previous section, we derived the ionizing photon fluxes and bolometric luminosities for a large proportion of our sample of H II regions. These parameters can provide a useful indication of the current level of star formation taking place within each clump. As discussed in Sections 6.1 and 6.2, the clump median radius and mass are ~ 1.4 pc and $5000 M_{\odot}$, respectively. It is likely that these clumps are host to embedded protoclusters (e.g. Lada & Lada 2003; Motte, Schilke & Lis 2003) rather than single massive stars. Furthermore, massive stars are almost exclusively found to be associated with clusters (e.g. de Wit et al. 2004; Gvaramadze et al. 2012) and we therefore want to compare these integrated parameters with the emission expected from an embedded cluster.

To do this, we have followed the procedure described by Walsh et al. (2001) using a Monte Carlo simulation to randomly sample a Kroupa (2001) IMF (i.e. $N \propto M^{\alpha}$, where $\alpha = 1.3$ for $0.08 M_{\odot} \leq M \leq 0.5 M_{\odot}$ and $\alpha = 2.3$ for $M > 0.5 M_{\odot}$). We used a stellar mass range of 0.1 to $120 M_{\odot}$ and have assumed an average massive SFE of 10 per cent (e.g. Johnston et al. 2009).

A random number is generated and compared to the probability assigned to each mass interval in the range to determine the mass of star. This process is repeated until the total mass of the stars is approximately equal to the clump mass \times SFE. We have used polynomial fits to the stellar atmosphere models (i.e. Davies et al. 2011, see Fig. 24 for details) to determine each star's contribution to the total cluster luminosity and Lyman continuum flux. This process was repeated 1000 times for each mass increment for clump masses in the range 100 to $5 \times 10^5 M_{\odot}$. The results of these simulations will be used in the following subsections.

8.5.2 Luminosity–Lyman continuum flux relation

Fig. 25 shows the Lyman photon flux as a function of bolometric luminosity for the H II region sample. In cases where multiple H II regions are within 10 arcsec of the RMS position, we have integrated the Lyman flux and used the size of the largest and most evolved H II region in the plot. We also show the luminosity–Lyman photon flux relation for OB stars from table of values given in Davies et al. (2011), the results of the Monte Carlo cluster simulations discussed in the previous subsection (red curve and grey shaded region, respectively).

The distribution of compact and UC H II regions reveals a possible correlation between the bolometric luminosity and ionizing photon flux. However, this is effectively a luminosity–luminosity plot with both of these parameters having a distance dependence and we

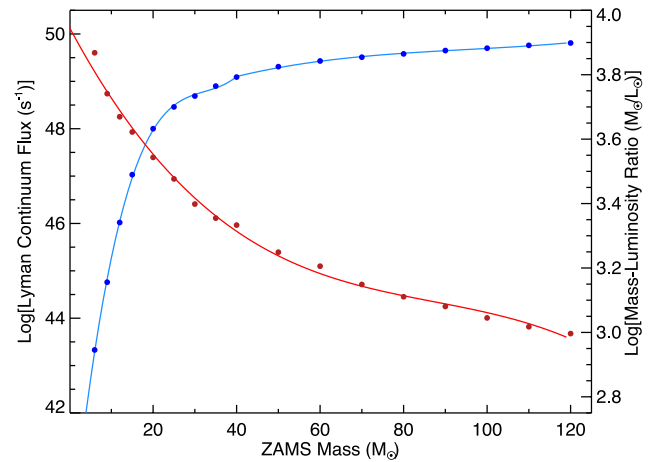


Figure 24. Mass–Lyman continuum flux and mass–luminosity ratio for OB stars derived from stellar atmosphere models are shown in blue and red, respectively. The filled circles show the data taken from Davies et al. (2011, and reference therein), while the curves show the polynomial fit to these points.

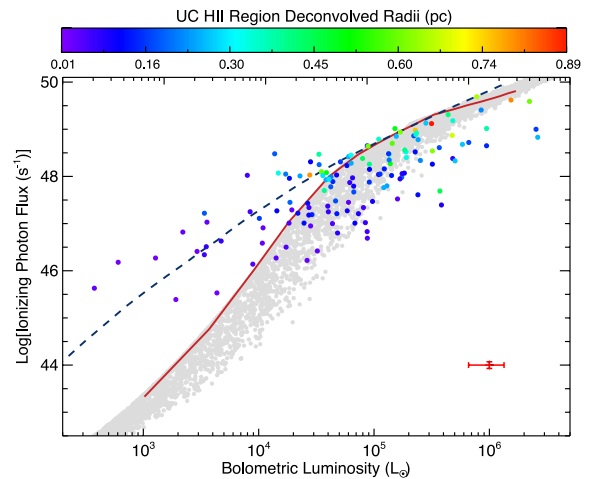


Figure 25. H II region Lyman continuum photon flux as a function of bolometric luminosity of the associated embedded cluster. The coloured circles show the source parameters derived from the data and indicate H II region sizes (see colour bar for values). The thick red curved line indicates the relationship for a single OB ZAMS star (from Martins et al. 2005) and the light grey circles show the result of the Monte Carlo simulation for clusters. The dashed black curve shows the ionizing photon flux that would be expected from a blackbody with the same radius and temperature as a ZAMS star (Sánchez-Monge private communication). Characteristic error bars for these parameters are shown in the lower-right corner of the plot.

need to be careful that the observed correlation is not the result of the Malmquist bias. We have therefore used a partial Spearman correlation test to remove the dependence of these two parameters on distance (see Paper I for more details). This gives a correlation coefficient of 0.69 for the whole sample with a p -value $< 1 \times 10^{-7}$, and so the correlation is significant and points to a strong correlation between these parameters.

Fig. 25 reveals some interesting features. First, the distribution of the measured parameters agrees well with the predictions of the Monte Carlo simulations for clusters, for which the single OB star locus forms an upper envelope. But the agreement only holds for the high-luminosity sources. For H II regions with luminosities equivalent to B stars, the observed distribution strongly disagrees

with the theoretical relationship and the measured values lie well above the predicted upper envelope. This discrepancy has also been noted by Lumsden et al. (2013) and Sánchez-Monge et al. (2013) for the RMS sample of H II regions (with which our sample shares a number of sources) and a sample of ~ 200 IRAS-selected sources, respectively. This excess Lyman continuum luminosity is unlikely to be the result of an overestimate of the radio flux as this is more likely to be underestimated due to the nature of interferometric observations, which filters out some of the larger spatial frequencies. This excess is, therefore, not easy to explain.

The sample of Sánchez-Monge et al. (2013) is split into three evolutionary types (millimetre source only – Type 1; millimetre with associated infrared source – Type 2; and infrared source only – Type 3). They find that 70 per cent of their Type 1 and 2 sources have excess ionizing photon fluxes and suggest that this may preferentially occur for *young* B-type stars, a conclusion independently arrived at by Lumsden et al. (2013). Following Sánchez-Monge et al. (2013), Fig. 25 also includes the locus of a blackbody with the same radius and temperature as ZAMS OB stars (dashed black curve), giving the theoretical upper limit to the Lyman continuum flux. The data for B-type H II regions agree significantly better with the blackbody prediction than with the ZAMS stellar atmosphere models, albeit with considerable scatter in these data.

Another feature to note in Fig. 25 is in the O-star distribution, where the larger H II regions appear closer to the Lyman luminosities expected from the single-star stellar atmosphere models. For late O-type stars, the ionizing flux is still consistent with what we might expect if the bolometric luminosity were due to a cluster rather than a single star; however, this is not the case for the early O stars where there are clearly fewer ionizing photons than expected, even allowing that the luminosity is the result of a cluster. One explanation could be that the interferometer has filtered out some of the larger scale radio continuum flux; however, this is more likely to be an issue for the more extended H II regions and most of the sources found to have a deficit of ionizing photons are the genuine UC H II regions. Another possibility is that some of these nebulae are actually optically thick which would result in the ionizing flux being underestimated and/or there could be significant absorption due to dust which would lower the overall number of ionizing photons in the H II region.

8.5.3 Clump mass–luminosity relationships

In Fig. 26, we present the ionizing photon flux and bolometric luminosity as a function of clump mass. Since the main focus of this paper is the clump properties and, given that we are unable to subdivide the masses of clumps with multiple H II regions, we simply integrate the radio flux for all embedded sources and their RMS counterparts to obtain the total bolometric luminosities and Lyman continuum flux for each clump. The yellow hatched region shown on these plots corresponds to the range of parameter values that includes 90 per cent of the Monte Carlo simulations, while the grey region indicates the region of parameter space covered by the most massive member of each cluster.

Inspecting these plots, it appears that the bolometric luminosity and ionizing photon flux are correlated with clump mass, even though there is a large scatter in the data. As above, we use the partial Spearman correlation test to remove any potential Malmquist bias. This gives a coefficient of 0.54 for the correlation between clump mass and Lyman continuum flux, and 0.53 for the clump mass

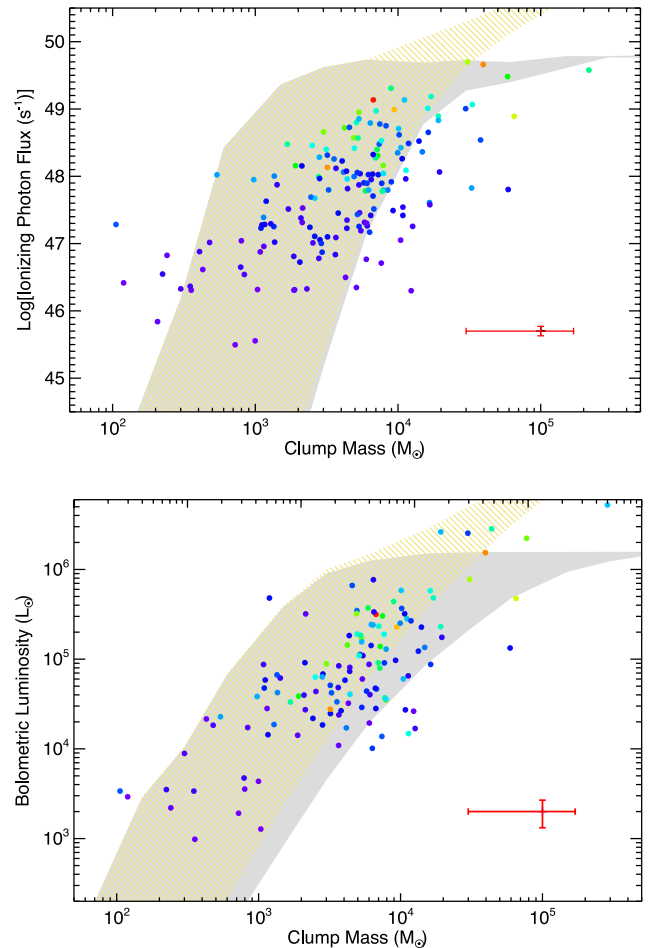


Figure 26. Upper panel: comparison of clump mass to ionizing photon flux emitted from their embedded UC H II regions. Lower panel: clump mass–bolometric luminosity relationship for the ATLASGAL–CORNISH sample of H II regions. The colours are the same as for Fig. 25 and correspond to the physical radius of the brightest embedded UC H II-region associated with each clump. The region of parameter space encompassed by the yellow hatching indicates the area in which 90 per cent of the simulated clusters are located, while the region highlighted in grey indicates the area where the most massive star of each cluster was located. Characteristic error bars for these parameters are shown in the lower-right corner of each plot.

and bolometric luminosity;⁷ the sample sizes are 164 and 130 and associated p -value for both correlations is $< 1 \times 10^{-7}$. The bolometric luminosity and Lyman continuum are therefore correlated with clump mass, with the most massive stars forming within the most massive clumps. Presumably the high accretion rate required to form the more massive stars can only be sustained by the most massive clumps. There is clearly an upper bound on cluster mass set by the clump mass, but the lower bound suggests that extremely low SFEs are rare (i.e. the SFE has a probability distribution that is not flat).

⁷ These correlation coefficients are slightly higher at 0.57 if we include only H II regions with sizes less than 0.2 pc in diameter in the mass–ionizing flux relation and 0.58 for the mass–bolometric luminosity relation if we include only sources where an SED fit has been made to the data. The sample sizes drop to 97 and 108, respectively, for these correlations but the p -values remain unchanged.

The dependence on clump mass of the Lyman continuum and bolometric luminosities derived from the data fit the Monte Carlo simulation very well. If the SFE were higher than the 10 per cent assumed in our simulations, the predicted Lyman and bolometric cluster luminosities (shaded regions) would be shifted to the left, worsening the match between measured and predicted values, requiring a further increase in cluster luminosity to restore the correspondence. The fact that the Lyman continuum flux and bolometric luminosities of the more evolved H II regions are close to those prediction by the stellar atmosphere models suggests that their luminosities are unlikely to change significantly and that their evolution is at an end.

Since the larger H II regions effectively set the upper envelope to the Lyman continuum flux and this agrees with predictions derived using an SFE of 10 per cent, we would conclude that an average SFE is more likely.

9 SUMMARY AND CONCLUSIONS

We have combined the ATLASGAL and CORNISH surveys to identify a large sample of molecular clumps associated with compact and UC H II regions. The ATLASGAL survey (Schuller et al. 2009; $280^\circ < \ell < 60^\circ$ and $|b| < 1.5^\circ$) traces the thermal emission from dense clumps at 870 μm and is complete to all massive clumps above $1000 M_\odot$ to the far side of the inner Galaxy (~ 20 kpc). In total the ATLASGAL survey has identified some 10 000 compact sources (Contreras et al. 2013 and Csengeri et al., in preparation), many of which have the potential to form the next generation of massive stars. The CORNISH project has mapped the northern Galactic plane between $\ell = 10^\circ$ and 65° and $|b| < 1^\circ$ for 5 GHz radio-continuum emission (Hoare et al. 2012; Purcell et al. 2013).

By comparing the positions of the ~ 2300 CORNISH radio sources with the dust emission maps and archival mid-infrared images extracted from the GLIMPSE data base, we have identified 213 compact and UC H II regions between $\ell = 10^\circ$ and 60° and $|b| < 1^\circ$. These are associated with 170 clumps with multiple UC H II regions being associated with a single clump in 23 cases. Radial velocities have been assigned for every source, largely from values available in the literature, with the majority being drawn from high-density tracers such as NH_3 (1,1) and HCO^+ (3–2) transitions. Where available, distances have been taken from maser parallax measurements; however, the majority are kinematically derived with distance ambiguity solutions provided from the literature or derived here from archival H I data.

These distances have been used to estimate clump masses and their effective radii, as well as the Lyman continuum fluxes and sizes of the embedded H II regions. The clump masses range from 100 to a few $10^5 M_\odot$ with a median value of $\sim 5000 M_\odot$, with sizes from a couple of tenths of a pc to a few pc and so cover the full range of size scales (i.e. cores, clumps and clouds). The median clump radius is ~ 1.4 pc and so the majority would be considered clumps; however, classifications based on size seem somewhat arbitrary. We have estimated the virial masses from the effective radii and molecular-line width of a large fraction of the clumps and find that, in most clumps, these are lower than the derived clump masses. This may suggest that the majority of clumps are gravitationally unstable and therefore likely to be collapsing. This is still true if we assume an equipartition of thermal and magnetic energy within the clumps.

We find that we are complete to optically thin UC H II regions driven by a ZAMS star of spectral type B0 or earlier embedded within a $1000 M_\odot$ to the far side of the Galaxy (~ 20 kpc and between $\ell = 10^\circ$ and 60°). Assuming a standard IMF (i.e. Kroupa

2001) and an SFE of 10 per cent, the minimum mass required to form at least one massive star is $\geq 1000 M_\odot$. It is therefore unlikely that any embedded B0 and earlier spectral type stars are associated with lower mass clumps. The H II region Lyman photon fluxes range between a 10^{45} and 10^{50} photon s^{-1} , and radii range from 0.01 to 0.9 pc, with median values of $\sim 10^{48}$ photon s^{-1} and 0.11 pc, respectively. Bolometric luminosities have been obtained by matching the H II regions with sources identified by the RMS survey; these have median value of $\sim 10^5 L_\odot$, which roughly corresponds to a $30 M_\odot$ star or ZAMS star of spectral type O7.

We use the derived parameters to investigate the correlation between the properties of these massive star-forming clumps and the star formation taking place within them. We use these data to test several density criteria suggested in the literature for massive star formation and correlate the distribution of this sample of embedded massive stars with the various features of Galactic structure present in the first quadrant. Our main findings are as follows.

(i) We find that the lower envelope of the mass–radius relation for methanol- and H II-region associated clumps is well modelled with a mean surface density of 0.05 g cm^{-2} . As stated in Paper I, this density threshold provides a better estimate of the lower envelope of the distribution for masses and radii greater than $500 M_\odot$ and 0.5 pc, respectively, than the Kauffmann et al. (2010b) criterion. This surface density threshold is also approximately twice that determined by Lada et al. (2010) and Heiderman et al. (2010) for ‘efficient’ star formation. So while the Lada et al. (2010) and Heiderman et al. (2010) criteria may be considered a minimum requirement for ‘efficient’ star formation, the threshold determined for this sample of massive star-forming clumps can be considered a minimum requirement for the ‘efficient’ formation of *massive* stars.

(ii) We find a clear correlation between the mass of the most massive embedded stars and the masses of their natal clumps, with the most massive stars forming in the most massive clumps. We find good agreement between the predicted Lyman and bolometric luminosities of an embedded cluster and those measured from our data assuming a standard IMF and an SFE of ~ 10 per cent. This is similar to the mean value of 7 ± 8 per cent reported by Johnston et al. (2009) for a sample of 31 intermediate- and high-mass clumps, but significantly lower than the upper limit for cluster formation of 30 per cent suggested by Lada & Lada (2003).

(iii) Of the 213 H II regions we have identified, we find that only half of these satisfy the size criterion to be classified as UC H II regions. This proportion decreases to less than 20 per cent when we consider H II regions associated with O stars and to zero when considering only early O stars.

(iv) The Galactocentric distribution of our sample reveals strong peaks at approximately 4 and 6 kpc, which corresponds to the expected radial positions of the Scutum–Centaurus arm, the end of the long bar and the tangent of the Sagittarius and far side of the Perseus arms. We find another peak at 8.5 kpc; however, this can be attributed solely to the W49A star-forming complex.

(v) We measure significantly different Galactic scaleheights for the H II regions and their host clumps (20.7 ± 1.7 and 29.1 ± 3.0 pc, respectively), with more extreme star-forming environments (i.e. W31, W43 and W51) being located closer to the Galactic mid-plane. The star formation associated with these prominent regions is intense enough that they have a significant effect on the derived Galactic parameters. Using the surface density of H II regions above our completeness limit (i.e. optically thin H II regions associated with stars of spectral types of B0 or earlier), we estimate the Galactic population to be ~ 400 . However, we note that we are not complete

to more extended H II regions around nearby B0 stars, as these fall below the CORNISH surface brightness sensitivity.

(vi) We have found H II regions associated with B stars have a higher Lyman continuum photon flux relative to that predicted from standard stellar atmosphere models compared to their bolometric luminosities. This is even more pronounced for later B stars. This would suggest that *young* B-type stars are significantly more luminous in the ultraviolet than predicted by current stellar atmosphere models.

(vii) Although the region of the Galactic plane discussed in this paper includes the major star-forming regions W31, W43 and W51 (the latter two of these have been labelled as mini-star bursts), it is W49A that stands out as being exceptional. It contains ~ 7 per cent of the H II regions identified in this study and has an integrated luminosity of $\sim 7 \times 10^6 L_{\odot}$, a mass of $\sim 4 \times 10^5 M_{\odot}$ and a very high SFE ($15\text{--}32 L_{\odot} M_{\odot}^{-1}$). This is one of the most massive and intense regions of star formation in the Galaxy.

(viii) The properties of the clumps associated with methanol masers and those associated with H II regions are in excellent agreement despite being drawn from different quadrants. The masses, sizes and structure of the clumps are all very similar and the embedded young massive stars identified by these have a similar distribution with respect to the clump structure. We find that the massive stars traced by the methanol masers and H II regions are found towards the centre of their host clumps, with their positions tightly correlated with the peaks in the dust emission. The similarities between the two samples and the fact that only 20 per cent of the methanol masers are associated with the MYSO and H II region stages provides strong support for the hypothesis that methanol masers are indeed tracing some of the earliest stages of massive star formation.

(ix) From the similarities between the *envelope* structure of clumps associated with both the early and late embedded stages of massive star formation (as traced by the methanol masers and radio emission), we conclude that the structure does not change significantly during the massive star's or its associated cluster's evolution. It is likely then that the clumps are already in place before the collapse and the subsequent star formation begins, and that once the collapse has begun the clumps effectively decouple from any external influences.

This is the second in a series of three papers planned to use the ATLASGAL survey to conduct a detailed and comprehensive investigation of massive star formation environments. The main aim of these papers is to use the unbiased nature of the dust emission mapped by ATLASGAL over the inner Galactic plane to connect the results derived from samples selected using different high-mass star formation tracers. In the final paper in this series, we will investigate the dust properties of a well-selected sample of MYSOs identified by the RMS survey team.

ACKNOWLEDGEMENTS

We would like to thank the anonymous referee for their many comments and suggestions that have helped to clarify and improve this work. The ATLASGAL project is a collaboration between the Max-Planck-Gesellschaft, the European Southern Observatory (ESO) and the Universidad de Chile. The University of New South Wales Digital Filter Bank used for the observations with the Mopra Telescope was provided with support from the Australian Research Council. This research has made use of the SIMBAD data base operated at CDS, Strasbourg, France. This work was partially funded

by the ERC Advanced Investigator Grant GLOSTAR (247078) and was partially carried out within the Collaborative Research Council 956, subproject A6, funded by the Deutsche Forschungsgemeinschaft (DFG). This paper made use of information from the Red MSX Source survey data base at www.ast.leeds.ac.uk/RMS which was constructed with support from the Science and Technology Facilities Council of the UK. LB acknowledges support from CONICYT project Basal PFB-06.

REFERENCES

- Aguirre J. E. et al., 2011, *ApJS*, 192, 4
 Anderson L. D., Bania T. M., 2009, *ApJ*, 690, 706
 Anglada G., Villuendas E., Estalella R., Beltrán M. T., Rodríguez L. F., Torrelles J. M., Curiel S., 1998, *AJ*, 116, 2953
 Araya E., Hofner P., Churchwell E., Kurtz S., 2002, *ApJS*, 138, 63
 Becker R. H., White R. L., McLean B. J., Helfand D. J., Zoonematkermani S., 1990, *ApJ*, 358, 485
 Becker R. H., White R. L., Helfand D. J., Zoonematkermani S., 1994, *ApJS*, 91, 347
 Benjamin R. A. et al., 2003, *PASP*, 115, 953
 Bergin E. A., Tafalla M., 2007, *ARA&A*, 45, 339
 Bertin E., Arnouts S., 1996, *A&AS*, 117, 393
 Bertoldi F., McKee C. F., 1992, *ApJ*, 395, 140
 Beuther H., Walsh A., Schilke P., Sridharan T. K., Menten K. M., Wyrowski F., 2002, *A&A*, 390, 289
 Beuther H. et al., 2012, *ApJ*, 747, 43
 Blum R. D., Conti P. S., Damineli A., 2000, *AJ*, 119, 1860
 Bonnell I. A., Bate M. R., Clarke C. J., Pringle J. E., 1997, *MNRAS*, 285, 201
 Bonnell I. A., Clarke C. J., Bate M. R., Pringle J. E., 2001, *MNRAS*, 324, 573
 Brand J., Blitz L., 1993, *A&A*, 275, 67
 Bressert E., Ginsburg A., Bally J., Battersby C., Longmore S., Testi L., 2012, *ApJ*, 758, L28
 Bronfman L., Nyman L.-A., May J., 1996, *A&AS*, 115, 81
 Bronfman L., Casassus S., May J., Nyman L.-A., 2000, *A&A*, 358, 521
 Brun F., de Naurois M., Hofmann W., Carrigan S., Djannati-Ataï A., Ohm S., 2011, in Alecian G., Belkacem K., Samadi R., Valls-Gabaud D., eds, *SF2A-2011: Proceedings of the Annual Meeting of the French Society of Astronomy and Astrophysics*, p. 545
 Busfield A. L., Purcell C. R., Hoare M. G., Lumsden S. L., Moore T. J. T., Oudmaijer R. D., 2006, *MNRAS*, 366, 1096
 Carpenter J. M., Snell R. L., Schloerb F. P., 1990, *ApJ*, 362, 147
 Caswell J. L. et al., 2010, *MNRAS*, 404, 1029
 Contreras Y. et al., 2013, *A&A*, 549, A45
 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
 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
 de Wit W. J., Testi L., Palla F., Vanzi L., Zinnecker H., 2004, *A&A*, 425, 937
 Deharveng L. et al., 2010, *A&A*, 523, A6
 Dib S., Walcher C. J., Heyer M., Audit E., Loinard L., 2009, *MNRAS*, 398, 1201
 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
 Ellingsen S. P., Shabala S. S., Kurtz S. E., 2005, *MNRAS*, 357, 1003
 Fish V. L., Reid M. J., Wilner D. J., Churchwell E., 2003, *ApJ*, 587, 701

- Fuller G. A., Myers P. C., 1992, *ApJ*, 384, 523
- Gallaway M. et al., 2013, *MNRAS*, 430, 808
- Georgelin Y. M., Georgelin Y. P., 1976, *A&A*, 49, 57
- Ginsburg A., Bressert E., Bally J., Battersby C., 2012, *ApJ*, 758, L29
- Giveon U., Becker R. H., Helfand D. J., White R. L., 2005, *AJ*, 129, 348
- Green J. A., McClure-Griffiths N. M., 2011, *MNRAS*, 417, 2500
- Güdel M., 2002, *ARA&A*, 40, 217
- Gvaramadze V. V., Weidner C., Kroupa P., Pflamm-Altenburg J., 2012, *MNRAS*, 424, 3037
- Heiderman A., Evans N. J., II, Allen L. E., Huard T., Heyer M., 2010, *ApJ*, 723, 1019
- Heyer M. H., Terebey S., 1998, *ApJ*, 502, 265
- Hildebrand R. H., 1983, *QJRAS*, 24, 267
- Hill T., Burton M. G., Minier V., Thompson M. A., Walsh A. J., Hunt-Cunningham M., Garay G., 2005, *MNRAS*, 363, 405
- Hindson L., Thompson M. A., Urquhart J. S., Faimali A., Clark J. S., Davies B., 2012, *MNRAS*, 421, 3418
- Hoare M. G., 2002, in Crowther P., ed., *ASP Conf. Ser. Vol. 267, Hot Star Workshop III: The Earliest Phases of Massive Star Birth*. Astron. Soc. Pac., San Francisco, p. 137
- Hoare M. G., Kurtz S. E., Lizano S., Keto E., Hofner P., 2007, in Reipurth B., Jewitt D., Keil K., eds, *Protostars and Planets V*. University of Arizona Press, Tucson, p. 181
- Hoare M. G. et al., 2012, *PASP*, 124, 939
- Humphreys R. M., Larsen J. A., 1995, *AJ*, 110, 2183
- Immer K., Schuller F., Omont A., Menten K. M., 2012, *A&A*, 537, A121
- Jackson J. M., Bania T. M., Simon R., Kolpak M., Clemens D. P., Heyer M., 2002, *ApJ*, 566, L81
- Johnston K. G., Shepherd D. S., Aguirre J. E., Dunham M. K., Rosolowsky E., Wood K., 2009, *ApJ*, 707, 283
- Kauffmann J., Pillai T., 2010, *ApJ*, 723, L7
- Kauffmann J., Bertoldi F., Bourke T. L., Evans N. J., II, Lee C. W., 2008, *A&A*, 487, 993
- Kauffmann J., Pillai T., Shetty R., Myers P. C., Goodman A. A., 2010a, *ApJ*, 712, 1137
- Kauffmann J., Pillai T., Shetty R., Myers P. C., Goodman A. A., 2010b, *ApJ*, 716, 433
- Kennicutt R. C., 2005, in Cesaroni R., Felli M., Churchwell E., Walmsley M., eds, *Proc. IAU Symp. 227, Massive Star Birth: A Crossroads of Astrophysics*. Cambridge Univ. Press, Cambridge, p. 3
- Keto E., 2007, *ApJ*, 666, 976
- Kim K.-T., Koo B.-C., 2001, *ApJ*, 549, 979
- Kolpak M. A., Jackson J. M., Bania T. M., Clemens D. P., Dickey J. M., 2003, *ApJ*, 582, 756
- Kroupa P., 2001, *MNRAS*, 322, 231
- Krumholz M. R., McKee C. F., Tumlinson J., 2008, *ApJ*, 689, 865
- Kurtz S., 2005, in Lis D. C., Blake G. A., Herbst E., eds, *Proc. IAU Symp. 231, Astrochemistry: Recent Successes and Current Challenges*. Cambridge Univ. Press, Cambridge, p. 47
- Kurtz S., Churchwell E., Wood D. O. S., 1994, *ApJS*, 91, 659
- Kurtz S. E., Watson A. M., Hofner P., Otte B., 1999, *ApJ*, 514, 232
- Lada C. J., Lada E. A., 2003, *ARA&A*, 41, 57
- Lada C. J., Lombardi M., Alves J. F., 2010, *ApJ*, 724, 687
- Larson R. B., 1981, *MNRAS*, 194, 809
- Lépine J. R. D. et al., 2011, *MNRAS*, 417, 698
- Longmore S. N. et al., 2012, *ApJ*, 746, 117
- 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*, preprint (arXiv:1308.0134)
- Martins F., Schaerer D., Hillier D. J., 2005, *A&A*, 436, 1049
- McClure-Griffiths N. M., Dickey J. M., Gaensler B. M., Green A. J., Haverkorn M., Strasser S., 2005, *ApJS*, 158, 178
- McKee C. F., Tan J. C., 2003, *ApJ*, 585, 850
- Moisés A. P., Damineli A., Figuerêdo E., Blum R. D., Conti P. S., Barbosa C. L., 2011, *MNRAS*, 411, 705
- Moore T. J. T., Urquhart J. S., Morgan L. K., Thompson M. A., 2012, *MNRAS*, 426, 701
- Motte F., Schilke P., Lis D. C., 2003, *ApJ*, 582, 277
- Motte F., Bontemps S., Schilke P., Schneider N., Menten K. M., Brogière D., 2007, *A&A*, 476, 1243
- Mottram J. C., Hoare M. G., Lumsden S. L., Oudmaijer R. D., Urquhart J. S., Meade M. R., Moore T. J. T., Stead J. J., 2010, *A&A*, 510, A89
- Mottram J. C. et al., 2011a, *A&A*, 525, A149
- Mottram J. C. et al., 2011b, *ApJ*, 730, L33
- Mueller K. E., Shirley Y. L., Evans N. J., II, Jacobson H. R., 2002, *ApJS*, 143, 469
- Murphy T., Cohen M., Ekers R. D., Green A. J., Wark R. M., Moss V., 2010, *MNRAS*, 405, 1560
- Nagayama T., Omodaka T., Handa T., Honma M., Kobayashi H., Kawaguchi N., Ueno Y., 2011, *PASJ*, 63, 719
- Nagy Z., van der Tak F. F. S., Fuller G. A., Spaans M., Plume R., 2012, *A&A*, 542, A6
- 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
- Papadopoulos P. P., Thi W.-F., Miniati F., Viti S., 2011, *MNRAS*, 414, 1705
- Pillai T., Wyrowski F., Carey S. J., Menten K. M., 2006, *A&A*, 450, 569
- Portegies Zwart S. F., McMillan S. L. W., Gieles M., 2010, *ARA&A*, 48, 431
- Purcell C. R. et al., 2012, *MNRAS*, 426, 1972
- Purcell C. R. et al., 2013, *ApJS*, 205, 1
- Ramesh B., Sridharan T. K., 1997, *MNRAS*, 284, 1001
- Reed B. C., 2000, *AJ*, 120, 314
- Reid M. J. et al., 2009, *ApJ*, 700, 137
- Roberts H., van der Tak F. F. S., Fuller G. A., Plume R., Bayet E., 2011, *A&A*, 525, A107
- Roman-Duval J., Jackson J. M., Heyer M., Johnson A., Rathborne J., Shah R., Simon R., 2009, *ApJ*, 699, 1153
- Rosolowsky E. et al., 2010, *ApJS*, 188, 123
- Russeil D., 2003, *A&A*, 397, 133
- Sánchez-Monge Á., Beltrán M. T., Cesaroni R., Fontani F., Brand J., Molinari S., Testi L., Burton M., 2013, *A&A*, 550, A21
- Sato M., Reid M. J., Brunthaler A., Menten K. M., 2010, *ApJ*, 720, 1055
- Schlingman W. M. et al., 2011, *ApJS*, 195, 14
- Schuller F. et al., 2009, *A&A*, 504, 415
- Sewilo M., Watson C., Araya E., Churchwell E., Hofner P., Kurtz S., 2004, *ApJ*, 154, 553
- Sringo G. et al., 2009, *A&A*, 497, 945
- Snell R. L., Carpenter J. M., Heyer M. H., 2002, *ApJ*, 578, 229
- Stead J. J., Hoare M. G., 2010, *MNRAS*, 407, 923
- Stil J. M. et al., 2006, *AJ*, 132, 1158
- Taylor J. H., Cordes J. M., 1993, *ApJ*, 411, 674
- Thompson M. A., Hatchell J., Walsh A. J., MacDonald G. H., Millar T. J., 2006, *A&A*, 453, 1003
- Tofani G., Felli M., Taylor G. B., Hunter T. R., 1995, *A&AS*, 112, 299
- Urquhart J. S., Busfield A. L., Hoare M. G., Lumsden S. L., Clarke A. J., Moore T. J. T., Mottram J. C., Oudmaijer R. D., 2007, *A&A*, 461, 11
- Urquhart J. S., Hoare M. G., Lumsden S. L., Oudmaijer R. D., Moore T. J. T., 2008a, in Beuther H., Linz H., Henning T., eds, *ASP Conf. Ser. Vol. 387, Massive Star Formation: Observations Confront Theory*. Astron. Soc. Pac., San Francisco, p. 381
- Urquhart J. S. et al., 2008b, *A&A*, 487, 253
- Urquhart J. S., Morgan L. K., Thompson M. A., 2009a, *A&A*, 497, 789
- Urquhart J. S. et al., 2009b, *A&A*, 501, 539
- Urquhart J. S. et al., 2011a, *MNRAS*, 410, 1237
- Urquhart J. S. et al., 2011b, *MNRAS*, 418, 1689
- Urquhart J. S. et al., 2012, *MNRAS*, 420, 1656
- Urquhart J. S. et al., 2013, *MNRAS*, 431, 1752 (Paper I)
- Walsh A. J., Burton M. G., Hyland A. R., Robinson G., 1998, *MNRAS*, 301, 640
- Walsh A. J., Bertoldi F., Burton M. G., Nikola T., 2001, *MNRAS*, 326, 36
- Walsh A. J., Macdonald G. H., Alvey N. D. S., Burton M. G., Lee J.-K., 2003, *A&A*, 410, 597
- Walsh A. J. et al., 2011, *MNRAS*, 416, 1764

- Watson C., Araya E., Sewilo M., Churchwell E., Hofner P., Kurtz S., 2003, *ApJ*, 587, 714
- White R. L., Becker R. H., Helfand D. J., 2005, *AJ*, 130, 586
- Wienen M., Wyrowski F., Schuller F., Menten K. M., Walmsley C. M., Bronfman L., Motte F., 2012, *A&A*, 544, A146
- Williams S. J., Fuller G. A., Sridharan T. K., 2004, *A&A*, 417, 115
- Wood D. O. S., Churchwell E., 1989a, *ApJS*, 69, 831
- Wood D. O. S., Churchwell E., 1989b, *ApJ*, 340, 265
- Xu Y., Moscadelli L., Reid M. J., Menten K. M., Zhang B., Zheng X. W., Brunthaler A., 2011, *ApJ*, 733, 25
- Zinchenko I., Henning T., Schreyer K., 1997, *A&AS*, 124, 385
- Zinnecker H., Yorke H. W., 2007, *ARA&A*, 45, 481
- Zoonematkermani S., Helfand D. J., Becker R. H., White R. L., Perley R. A., 1990, *ApJS*, 74, 181

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table 3. Derived clump parameters.

Table 4. Derived H II region parameters.

Figure 11. Source-averaged continuum-included H I spectra extracted from the SGPS and VGPS archives towards two H II regions for which a distance is not available in the literature.

Figure 12. Example H I spectra extracted from the SGPS and VGPS archives (<http://mnras.oxfordjournals.org/lookup/suppl/doi:10.1093/mnras/stt1310/-/DC1>).

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.

This paper has been typeset from a \TeX/L\AA\TeX file prepared by the author.