Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **439,** 3719–3754 (2014)
Advance Access publication 2014 February 27

doi:10.1093/mnras/stu224

A multiwavelength study of embedded clusters in W5-east, NGC 7538, S235, S252 and S254-S258

L. Chavarría,[1,2,3,4]★ L. Allen,[5] C. Brunt,[6] J. L. Hora,[2] A. Muench[2] and G. Fazio[2]

[1]*Universidad de Chile, Camino del Observatorio 1515, Santiago, Chile*
[2]*Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA*
[3]*Laboratoire d'Astrophysique de Bordeaux, 2 rue de l'Observatoire, F-33271 Floirac Cedex, France*
[4]*Centro de Astrobiología (CSIC/INTA), Ctra. de Torrejón a Ajalvir, km 4, E-28850, Torrejón de Ardoz, Madrid, Spain*
[5]*National Optical Astronomical Observatory, 950 North Cherry Avenue, Tucson, AZ 85719, USA*
[6]*The School of Physics, University of Exeter, The Queens Drive, Exeter, Devon EX4 4QL, UK*

Accepted 2014 January 31. Received 2014 January 30; in original form 2013 June 19

ABSTRACT

We present *Spitzer*, near-IR (NIR) and millimetre observations of the massive star-forming regions W5-east, S235, S252, S254-S258 and NGC 7538. *Spitzer* data is combined with NIR observations to identify and classify the young population while ^{12}CO and ^{13}CO observations are used to examine the parental molecular cloud. We detect in total 3021 young stellar objects (YSOs). Of those, 539 are classified as Class I, and 1186 as Class II sources. YSOs are distributed in groups surrounded by a more scattered population. Class I sources are more hierarchically organized than Class II and associated with the most dense molecular material. We identify in total 41 embedded clusters containing between 52 and 73 per cent of the YSOs. Clusters are in general non-virialized, turbulent and have star formation efficiencies between 5 and 50 per cent. We compare the physical properties of embedded clusters harbouring massive stars (MEC) and low-mass embedded clusters (LEC) and find that both groups follow similar correlations where the MEC are an extrapolation of the LEC. The mean separation between MEC members is smaller compared to the cluster Jeans length than for LEC members. These results are in agreement with a scenario where stars are formed in hierarchically distributed dusty filaments where fragmentation is mainly driven by turbulence for the more massive clusters. We find several young OB-type stars having IR-excess emission which may be due to the presence of an accretion disc.

Key words: stars: early-type – stars: formation – stars: pre-main sequence – H II regions – infrared: stars.

1 INTRODUCTION

Embedded clusters are truly stellar nurseries, more than 90 per cent of the stars in our Galaxy are formed in such associations (Zinnecker & Yorke 2007). Since they are young (with ages of less than 2–3 Myr), they still contain the imprints of the parental molecular cloud. Moreover, the wide range of number stars (10–10^4, Lada & Lada 2003) and high density of members (more than 20 stars pc^{-2}, Lada & Lada 2003) makes embedded clusters perfect laboratories to study cluster dynamics, stellar evolution and star formation theories.

Among embedded clusters, those harbouring massive stars (hereafter massive embedded clusters) are particularly important since both the formation of massive stars and the impact of massive stars feedback on the other cluster members and the parental molecular cloud are still not well understood. Massive stars begin hydrogen burning while they are still accreting material and the strong stellar winds and ultraviolet (UV) photons emitted will eventually stop the accretion before the star reaches its final mass. In addition, the emitted UV photons ionize the surrounding cloud and create an expanding H II region that disrupts and compress the natal molecular cloud. The feedback effects of massive stars over, for example, their disc lifetimes (in case they have disc) and/or over other cluster members are unclear. It is also unclear under what conditions the H II regions and shock waves will either destroy the molecular cloud or trigger star formation (there are several examples showing molecular gas that has been swept up by expanding H II regions and that contains young stars, e.g. Chavarría et al. 2008; Deharveng et al. 2008; Wang et al. 2011). This feedback into the interstellar medium is absent in the case of low-mass stars and it may play an important role in the star formation rate (SFR) and evolution of the Galaxy.

★ E-mail: luisagustinchavarria@gmail.com

Of the three models proposed to explain the formation of massive stars (competitive accretion in a protocluster environment, monolithic collapse in turbulent cores and stellar collisions and mergers in very dense systems, McKee & Tan 2003; Bonnell, Vine & Bate 2004; Bonnell, Clarke & Bate 2008), competitive accretion and turbulent cores are somehow a scaled-up version of low-mass star formation. Competitive accretion requires that massive stars form at the centre of the cluster (known as primordial mass segregation). This has been observed in some young clusters. However, it can be also achieved by the dynamical interaction between the cluster members and the gas they are embedded in (e.g. Chavarría et al. 2010). The turbulent core model proposes that density enhancements created by turbulent motions allows the high-accretion rates necessary for high-mass stars to form. This requires massive cores highly turbulent which have been lately reported (e.g. Herpin et al. 2012). In addition, rotating toroid-like structures and outflows (which are star formation indicators via gravitational collapse) have been observed for sources with masses up to 25 $M_{\odot}$ (e.g. Cesaroni et al. 2005; Garay et al. 2007). However, there is still no evidence of discs in O-type stars. This suggests that coalescence may be an alternative theory of formation for stars with masses of more than 30 $M_{\odot}$ (Zinnecker & Yorke 2007). However, the high star densities necessary for coalescence to occur have not been yet reported.

Since embedded clusters are located deep inside their natal molecular cloud, they can be observed only at infrared (IR), and millimetre wavelengths. In the last few years, several authors have studied embedded clusters using a combination of *Spitzer* Infrared Array Camera (IRAC) (Allen et al. 2004; Fazio et al. 2004) and near-IR (NIR) data, which is proven to be a powerful tool to identify and classify young stellar objects (YSOs) in regions of star formation (e.g. Ophiuchus, Serpens, Perseus, Taurus and NGC 1333, Winston et al. 2007; Gutermuth et al. 2008; Schmeja, Kumar & Ferreira 2008), most of them in the low-mass range. Schmeja et al. (2008) studied the spatial distribution of different class YSOs in embedded clusters and found that they mostly evolve from a hierarchical to a more centrally concentrated distribution. Gutermuth et al. (2009) analysed 36 low-mass embedded clusters (LEC) and found that YSOs are likely formed by Jeans fragmentation of parsec-scale clumps, in agreement with the accretion scenario. Massive embedded clusters (MEC), on the other hand, have been more elusive to scrutinize mainly due to two reasons; (a) they are less abundant than low-mass clusters and hence usually located at several kpc from the Sun and (b) the early stages of massive star formation last only a few million years. Because of this, only a few MEC have been evenly studied until date using a combination of NIR and *Spitzer* data (e.g. Chavarría et al. 2008; Kirsanova et al. 2008; Koenig et al. 2008; Dewangan & Anandarao 2011; Ojha et al. 2011). Those studies have been carried out by several authors using different data sets and analysis. As a consequence, their results are difficult to compare between each other and the available statistics is still poor.

We present a homogeneous *Spitzer*-IRAC, NIR and molecular data study on the young stellar population in five high-mass star-forming regions: W5-east, S235, S252, S254-S258 and NGC 7538. Our study aims to address the following questions: What are the physical properties of YSOs in MEC? Are those properties similar to the low-mass case? What are the implications of those properties on the massive star formation scenario? Also, we provide a set of physical quantities with a reasonable statistical weight that will help to constrain theoretical models of star formation and cluster dynamic.

Region S254-S258 was presented by Chavarría et al. (2008). In this work, we use their results for a more recent distance estimate derived from trigonometric parallax of methanol masers (1.6 kpc, Rygl et al. 2010).

Table 1. List of observed regions.

Name	RA (J2000) hh mm ss	Dec. (J2000) dd mm ss	$D_{\odot}$ (kpc)
W5-east	03 01 31.20	60 29 13.0	2.0
S235	05 40 52.00	35 42 20.0	1.8
S252	06 09 04.70	20 35 09.0	2.1
S254-S258	06 12 46.00	18 00 38.0	1.6
NGC 7538	23 13 42.00	61 30 10.0	2.7

The position and distance to the studied regions are summarized in Table 1. Following there is a brief description of the regions (see Chavarría et al. 2008, for a description of region S254-S258). In Section 3, we explain our observations and the data reduction process. Results, including the identification of YSOs, analysis of their spatial distribution and the study of the molecular cloud structure are presented in Section 4. In Section 5, discuss and compare the physical properties of YSOs for low-mass and MEC. Our conclusions are presented in Section 6. The estimation of background contamination, a comparison with previous observations, non-detection estimate and Gaussian decomposition of molecular spectrum are explained in Appendices A through E.

2 DESCRIPTION OF THE STUDIED REGIONS

The luminosities given in this section are normalized by each region assumed distance (see Table 1).

2.1 W5-east

W5-east is located on the west side of H II region Sharpless 199 (S199, also called IC 1848, Sharpless 1959; Koenig et al. 2008). The H II region is powered by an O7V type star (HD18326) and it harbours three young star clusters: AFGL4029 (Price & Walker 1976; Carpenter, Snell & Schloerb 1993; Deharveng et al. 1997), AFGL416 or Sh 2-201 (Carpenter et al. 1993; Ojha et al. 2004) and G138.15+1.69 (Bica, Dutra & Barbuy 2003a). The two former contain young massive stars.

AFGL4029 has a far-IR luminosity ($L_{\rm FIR}$) of 1.7×10^4 $L_{\odot}$ (Snell et al. 1988). It is associated with the *IRAS* source 02575+6017 and sub-millimetre emission (Morgan et al. 2008). The cluster has an estimated number of members of 240 (Carpenter, Heyer & Snell 2000) and around 80 sources with Hα emission (Ogura, Sugitani & Pickles 2002; Nakano et al. 2008). It also contains UCH II regions (Kurtz, Churchwell & Wood 1994; Zapata, Rodríguez & Kurtz 2001) associated with the infrared sources IRS1 and IRS2 (Beichman 1979). IRS1 is believed to be the exciting source of a molecular outflow (Snell et al. 1988) detected also at optical wavelengths (Ray et al. 1990). G138.15+1.69 is located 12 arcmin north-west of AFGL4029 (Bica et al. 2003a). It is associated with the *IRAS* source 02570+6028 ($L_{\rm FIR}$ =380 $L_{\odot}$) and contains at least 80 members (Carpenter et al. 2000) and about 30 stars with Hα emission (Ogura et al. 2002; Nakano et al. 2008). Both G138.15+1.69 and AFGL4029 are immerse in pillar-like structures located along a dusty rim and suggesting a triggered star formation scenario (Koenig et al. 2008; Niwa et al. 2009; Deharveng et al. 2012). This is also supported by the younger ages associated with YSOs inside the rim compared to the ages of YSOs located outside the rim in direction to the ionizing star (Chauhan et al. 2011). AFGL416 is associated with

the H II region Sharpless 201 (S201), the *IRAS* source 02593+6016 and sub-millimetre emission (Morgan et al. 2008). The cluster luminosity ($L_{\rm FIR}$) is 4.2×10^4 $L_\odot$ and it harbours at least 90 members (Carpenter et al. 2000).

The dust content in the region was investigated by Deharveng et al. (2012) using *Herschel* observations in the far-IR. They find that the dust is being collected by the expansion of the ionizing front. In addition to this, they identify around 40 point sources at 100μm over the same field of view (FOV) as in our observations. Those sources are associated with the various clusters in the region as well as with the ionizing front edge and are presumably very YSOs.

The distance to cluster AFGL4029 is estimated between 1.9 and 3.8 kpc (Churchwell, Walmsley & Cesaroni 1990; Hillwig et al. 2006). Since AFGL4029 and AFGL416 are both associated with the H II region S199, and G138.15+1.69 has a similar $V_{\rm LSR}$ as AFGL4029 and AFGL416 (Mampaso et al. 1987), we adopt a distance of 2 kpc (Becker & Fenkart 1971; Mampaso et al. 1987) to the whole region W5-east.

2.2 Sharpless 235

Sharpless 235 (S235) is an H II region from the Sharpless (1959) catalogue first reported as an emission nebula by Minkowski (1946). The H II region is powered by an O9.5V type star (BD+351201, Georgelin, Georgelin & Roux 1973) and also two bright infrared sources (IRS1 and IRS2) with estimated spectral types of early B stars (Evans & Blair 1981; Thompson, Thronson & Campbell 1983). Those are part of the young cluster S235 (Carpenter et al. 1993; Bica et al. 2003a).

There are at least 270 YSOs in this region (Dewangan & Anandarao 2011). Of those, around 70 per cent belong to different clusters: S235, S235 East1, S235 East2, S235 NW and S235AB (Bica, Dutra & Barbuy 2003a, Kirsanova et al. 2008, Camargo, Bonatto & Bica 2011, Dewangan & Anandarao 2011). Cluster S235AB harbours three compact H II regions: S235A, S235B and S235C (Israel & Felli 1978). H II regions S235A and S235B are powered by zero-age main-sequence stars with spectral type between B0 and O9.5 (Krassner, Pipher & Sharpless 1979; Olofsson 1983). They are associated with the infrared sources IRS3 and IRS4 (Evans & Blair 1981) as well as water and methanol masers (Felli et al. 2006).

Estimate distances to region S235 are between 1.6 and 2.5 kpc (Israel & Felli 1978; Hunter & Massey 1990). In this paper, we adopt a distance of 1.8 kpc (Evans & Blair 1981).

2.3 Sharpless 252

Sharpless 252 (S252, also called NGC 2175) is an extended H II region (size of ∼1°) from the Sharpless (1959) catalogue. The H II region is powered by an O6.5V type star (HD42088, Grasladen & Carrasco 1975), and it contains four compact H II regions associated with a young population of stars (S252A, S252B, S252C and S252E from Felli, Habing & Israel 1977).

The H II region S252A (also named AFGL5179) is powered by a B1V-O9.5V star (Hiltner 1956; Jose et al. 2012) and it contains an infrared cluster also named S252A with around 80 members (Bica et al. 2003a; Tej et al. 2006). S252A is also associated with the source IRAS 06055+2039 (Carpenter et al. 1995, $L_{\rm FIR} = 4.3 \times 10^3$ $L_\odot$), a water maser (Lada & Wooden 1979) and outflow activity (Xu et al. 2006). The H II region S252B is powered by a B1.0V type star (Jose et al. 2012) which seems to be somehow isolated. The H II region S252C is powered by a B0.5 type star (Jose et al. 2012) and is associated with the infrared cluster S252C (Chavarría-K et al. 1989; Bica et al. 2003a). The H II region S252E (also NGC 2175s or AFGL5184) is powered by a B0V type star (Jose et al. 2012) and is associated with a homonymous infrared cluster and the source IRAS 06068+2030 ($L_{\rm FIR}$=2.9 × 10^3 $L_\odot$, Carpenter et al. 1995).

The distance estimate for S252 are between 1.25 and 2.9 kpc (Mirabel et al. 1987; Wouterloot & Brand 1989). We adopt a distance of 2.1 kpc, derived from trigonometric parallax of methanol masers (Reid et al. 2009).

2.4 NGC 7538

NGC 7538 is an active site of star formation located in the Perseus spiral arm. It was first documented as an emission nebula (H II region S158) by Sharpless (1959) and it contains massive stars in different evolutionary stages; main-sequence stars which ionize the H II region NGC 7538, with spectral types between O3 and O9 (IRS5 and IRS6, Puga et al. 2010); infrared sources IRS1 (associated with a disc and an outflow; Pestalozzi, Elitzur & Conway 2009; Sandell et al. 2009), IRS2 and IRS3 (Wynn-Williams, Beckling & Neugebauer 1974), located south of NGC 7538 and associated with UCH II regions and with the infrared cluster NGC 7538S (Carpenter et al. 1993; Bica et al. 2003a; Sandell & Wright 2010); YSOs like IRS9 and IRS11 (Werner et al. 1979) and sub-millimetre clumps without associated IR emission (Reid & Wilson 2005).

The spatial location of sources at different evolutionary stages suggest a sequence of star formation from north-east to south-west (Puga et al. 2010). Balog et al. (2004) detected 238 sources with NIR excess in NGC 7538. More recent studies in the region have been done by Reid & Wilson (2005), Kraus et al. (2006) and Barriault & Joncas (2007). We suggest these authors and references therein for a more detailed description of this region.

Distance estimations for NGC 7538 are between 2.1 kpc (Balog et al. 2004) and 3.5 kpc (Israel 1977). The most used distance to the region is 2.8 kpc, corresponding to the photometric distance calculated by Crampton, Georgelin & Georgelin (1978). In this paper, we adopt a distance of 2.7 kpc, derived by Moscadelli et al. (2009) using trigonometric parallax.

3 OBSERVATIONS AND DATA REDUCTION

3.1 Mid-IR imaging

All regions were observed with the Infrared Array Camera (IRAC) on the *Spitzer Space Telescope* between 2004 December and 2007 October. We used an integration time of 10.4 s per dither in the High Dynamic Range (HDR) mode, with three dithers per map position. HDR mode also acquires 0.4 s integration time frames for the recovery of bright sources which are saturated in longer exposures. We used S. Carey's artefact correction scripts to remove column pull-down and some of the banding and muxbleed artefacts (Hora et al. 2004). IRAC mosaics were constructed using the Basic Calibrated Data (BCD) frames (S14.0.0, and S15.3.0) with IRACPROC[1] (Schuster, Marengo & Patten 2006). The final image scale is 0.6 arcsec per pixel.

We used IRAF DAOPHOT packages to extract sources and perform photometry in each band. The photometry was done using an

[1] https://www.cfa.harvard.edu/twiki/bin/view/Main/IracProc

aperture radius of 1.8 arcsec (3 pixels) for the 3.6 and 4.5 μm bands and 2.4 arcsec (4 pixels) for the 5.8 and 8.0 μm bands. Inner and outer sky annuli of 4.8 and 6 arcsec radius respectively were used in each IRAC band. We calculated the zero-point magnitudes in each band using Vega fluxes. Their values are 18.443, 17.879, 17.234 and 16.477 mag for the 3.6, 4.5, 5.8 and 8.0 μm bands, respectively (aperture corrections are included).

Mosaics from the Multiband Imaging Photometer for *Spitzer* (MIPS) for all regions were downloaded from the NASA/IPAC Infrared Science Archive.[2] MIPS mosaics were constructed using the BCD frames (S18.12.0). The program ID for region W5-east is 20300 (PI is Lori Allen), for the other regions the program ID is 40005 (PI is Giovani Fazio).

3.2 Near-IR imaging

Observations with Flamingos were performed at the 2.1 m telescope located at Kitt Peak National Observatory in 2004 December and 2006 January. We observed *J, H* and *K* bands (centred at 1.24, 1.65 and 2.21 μm, respectively) for W5-east and S252, *J* and *K* bands for NGC 7538 and *K*-band for S235. Flamingos has a 2048 × 2048 pixel Hawaii II HgCdTe detector array with a plate-scale of 0.611 arcsec per pixel which gives an FOV of 20×20 arcmin. The seeing during the observations was approximately 1.0 arcsec. The observations were done in dithering mode with 15–30 arcsec shifts for a total integration time of approximately 1000 s per band.

The NIR data reduction was performed using IDL,[3] IRAF[4] and WCS tools.[5] Image distortion at the edge of Flamingos FOV were corrected by applying a second-order transformation in the *x* and *u* axis using the position of matching sources from the 2MASS catalogue. This has an impact of less than 1 arcsec in the source position at the edges of the detector which produces a variation in the point spread function over the mosaic. To avoid systemic errors in the flux estimation of detected sources, we performed aperture photometry with the following parameters: threshold of 3 sigma for detection and 1.8, 4.8 and 6 arscec in radii for aperture, inner annulus and inner plus outer annulus, respectively. Calibration was performed by minimizing residuals to corresponding 2MASS detections. No colour terms were assumed in the zero-point minimization. The rms for residuals between the data sets used was less than 0.08 mag in all bands.

We also obtained NIR data with the SAO Widefield Infrared Camera (SWIRC) at the 6.5 m Multiple Mirror Telescope (MMT) telescope located at Fred Lawrence Whipple Observatory in 2005. We acquired data in the *J* and *H* bands for S235 and in the *H* band for NGC 7538. SWIRC has a 2048 × 2048 pixel Hawaii II detector array with a plate-scale of 0.15 arcsec per pixel which gives an FOV of 5 × 5 arcmin. For NGC 7538, the observations were performed in dithering mode with 15–20 arcsec shifts for a total integration time of 100 s in the *H* band. For S235, we observed in dithering mode for a total integration time of 30 s in *H* and 60 s in *J*. We performed aperture photometry of 0.9 arcsec in radii for both bands, using a detection threshold of 3 sigma, and 1.0 and 2.0 arcsec radii for inner annulus and inner plus outer annulus, respectively. As for Flamingos, we calibrated SWIRC photometry by minimizing residuals to corresponding 2MASS detections.

Finally, we combined IRAC, Flamingos and SWIRC data by merging the photometry of the seven bands. The maximum tolerance in positional offsets between bands was 2.0 arcsec. We use detections with error of less than 0.2 mag in our analysis.

3.3 Millimetre observations

Millimetre wavelength spectral line observations of the target regions were conducted at the Five College Radio Astronomy Observatory (FCRAO) 14 m telescope in New Salem, Massachusetts between 2003 November and 2006 April. We observed the $J = 1$–0 lines of ^{12}CO and ^{13}CO simultaneously in on-the-fly mapping mode, using the 32 pixel Second Quabbin Optical Imaging Array (SEQUOIA) focal plane array (Erickson et al. 1999) and the dual channel correlator (DCC). Regions S252 and S235 were observed as part of the Extended Outer Galaxy Survey (E-OGS; Brunt 2004), that extends the coverage of the FCRAO Outer Galaxy Survey (Heyer et al. 1998) to Galactic longitude $l = 193$, over a the latitude range $-3.5 \leq b \leq +5.5$. Observations of W5-east were conducted in May–June of 2004 and 2005, and observations of NGC 7538 were conducted in 2003 November.

For all observations, pointing and focus checks were carried out every 3–4 h, shortly after dawn–dusk or after a significant change in source coordinates. The data were initially converted to the T_A^* scale using the standard chopper wheel method (Kutner & Ulich 1981). We used the OTFTOOL software, written by M. Heyer, G. Narayanan and M. Brewer, to place the spectra on a regular 22.5 arcsec grid in Galactic *l*, *b* coordinates. Individual spectra contributing to a given coordinate were assigned a $1/\sigma^2$ weighting during the griding, after first fitting and subtracting a first-order baseline from signal-free regions of each spectrum. The gridded data were scaled to the main beam temperature scale by dividing by the main beam efficiency of 0.45 (^{12}CO) or 0.48 (^{13}CO). The FCRAO beam size is 45 arcsec for ^{12}CO and 46 arcsec for ^{13}CO. For S252, S235 and W5-east the 1024-channel DCC was configured with a total bandwidth of 50 MHz, yielding a channel separation of 0.126 km s^{-1} (^{12}CO) or 0.132 km s^{-1} (^{13}CO). For NGC 7538, the total bandwidth was 25 MHz and the channel separation was 0.063 km s^{-1} (^{12}CO) or 0.066 km s^{-1} (^{13}CO). The velocity resolution is 1.21 times the channel spacing.

4 RESULTS

Spitzer-IRAC colour images for W5-east, S235, S252 and NGC 7538 are shown in Figs 1–4, respectively (for region S254-S258, see fig. 1 from Chavarría et al. 2008). All regions show abundant extended emission in reddish colour corresponding mainly to fluorescence from polycyclic aromatic hydrocarbon (PAH) molecules. Regions W5-east and S252 show PAHs ridges (more prominent in W5-east) at the interface between the expanding H II region and the neutral molecular material. It is possible to see several pillar-like structures along the ridges pointing towards the ionizing source. Some of them seem to have a single source at their tip (e.g. between AFGL4029 and G138.15+1.69) and others have several (e.g. clusters AFGL4029 and G138.15+1.69). The dust around the H II regions (as traced by PAHs emission) shows at least two different morphologies: single and bipolar cavity. Single cavities range from less than 1 (e.g. H II region S235C) to around 5 pc in size (e.g. H II region S254) and they have the ionizing source(s) located in their interior. On the other hand, bipolar cavities (e.g. H II regions

[2] http://sha.ipac.caltech.edu

[3] Linearization, developed by Robert A. Gutermuth (http://www.astro.umass.edu/~rguter/ Rob_Gutermuth_Astronomy/IDL_Page.html).

[4] Darks, flat-field, bad-pixel mask and background frame creation and application.

[5] Distortion and astrometry correction, developed by Doug Mink (http://www.harvard.edu/wcstools/).

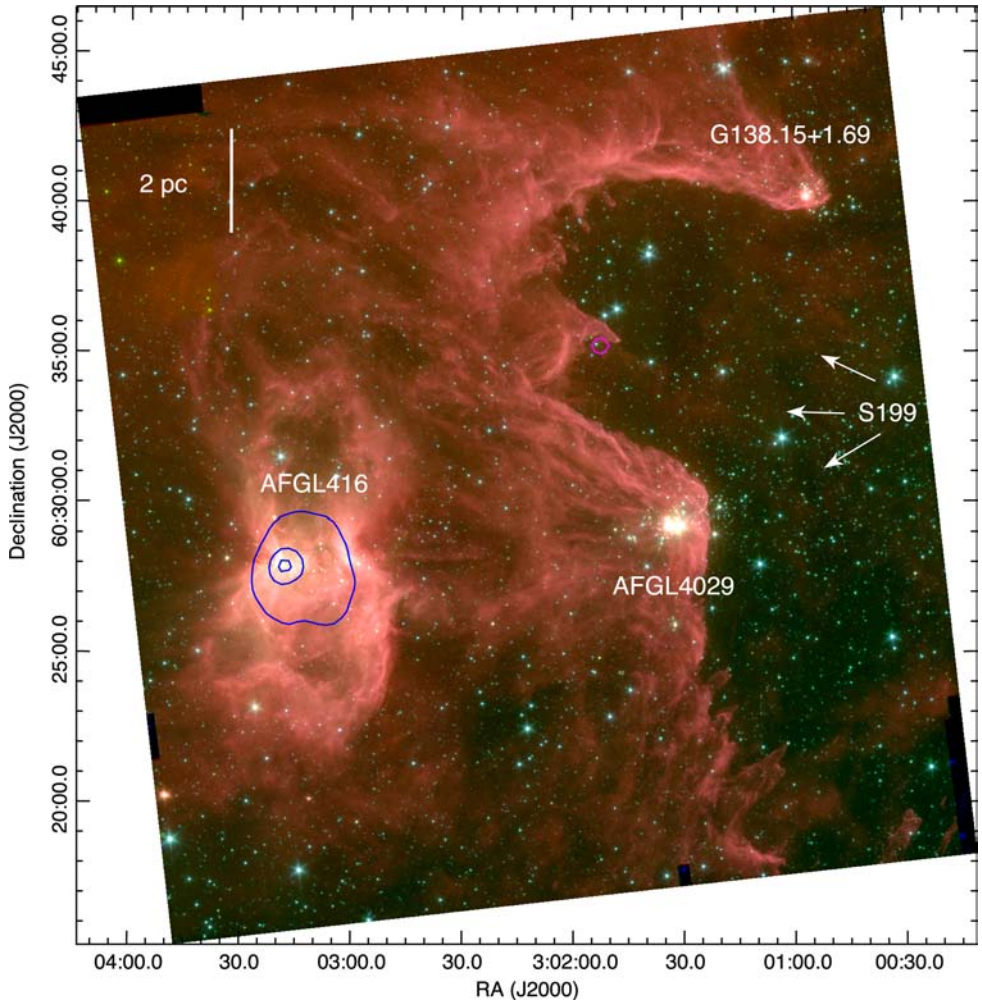

Figure 1. IRAC three-colour image of region W5-east (blue: 3.6 μm, green: 4.5 μm, red: 8.0 μm). Blue and magenta contours correspond to the 1.4 GHz (21 cm) emission from the NRAO VLA Sky Survey (Condon et al. 1998). Contours are at 10, 50 and 90 per cent of peak emission. Blue contours are associated with W5-east, while magenta contours are likely background sources. The white arrows show the ionizing front advancing direction. The ionizing star HD18326 is out of the field of view (FOV).

AFGL416, S252E and S252C) have lobe sizes of a couple of parsecs and the ionizing source located between lobes. It is possible that bipolar cavities look like single cavities when seen through the end of one of their lobes. Another feature seen in the IRAC colour images are several greenish blobs (e.g. around IRS9 and IRS1-3 in NGC 7538, east side of S252A, S235-E1 and S235-E2) known as extended green objects. Those correspond mainly to shocked H_2 detected in the IRAC 4.5 μm band and are used as tracers for outflow activity (Noriega-Crespo et al. 2004; Cyganowski et al. 2008). Finally, it is possible to identify several areas with a high density of sources, some of them associated with known embedded clusters (e.g. AFGL4029, S235A-C, S252A). In the following section, we explain how we identify the young population among the studied regions.

4.1 Identifying and classifying young stars

We combine NIR, IRAC and MIPS data to identify and classify YSOs in the observed regions. First, we identify YSOs from sources detected in the IRAC and NIR+IRAC bands. Then, we identify more YSOs from sources detected only in the NIR bands by their colour excess in the $[J-H]$ versus $[H-K]$ colour–colour diagram. Finally, sources detected in the IRAC-8 μm and MIPS-24 μm bands but not detected in the IRAC-3.6 μm band are also identified as YSOs.

4.1.1 Sources with IRAC four-band detections

Sources with detections in all IRAC bands are classified using their observed IRAC spectral energy distribution slope $\alpha_{\rm IRAC}$ as Class I ($\alpha_{\rm IRAC} > 0$), Class II ($-2 < \alpha_{\rm IRAC} < 0$) and stellar photospheres or Class III ($\alpha_{\rm IRAC} < -2$) (Lada 1987). This classification scheme follows an evolutionary trend that begins with a Class 0 (collapsing starless envelope), continues with Class I (star plus collapsing envelope), then Class II (star plus a disc) and Class III (star with a thin disc). If well there may be a few Class 0 sources among our Class I sample, this has no implications in our analysis since they are less evolved than the Class II sources as well. Among the sources detected in the IRAC bands, those detected also in the NIR bands are classified using their derreddened $\alpha_{\rm IRAC}$ value. The percentage of sources that changed their classification between the observed and derreddened $\alpha_{\rm IRAC}$ value is of less than 5 per cent in each region. After background contaminants subtraction, we identify 495 Class I

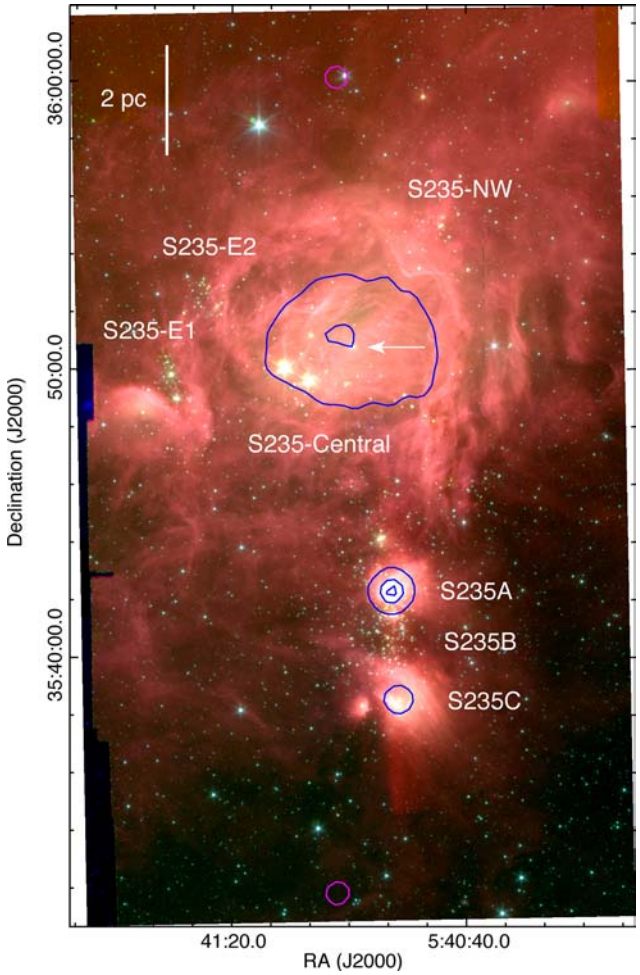

Figure 2. IRAC 3-colour image of region S235 (blue: 3.6 μm, green: 4.5 μm, red: 8.0 μm). The contours are the same as for Fig. 1. The white arrow indicates the ionizing star BD+351201.

and 1186 Class II out of the 3531 detected sources. The distribution of observed $\alpha_{\rm IRAC}$ values is shown in Fig. 5.

4.1.2 Adding MIPS detections

Some sources in our sample that are detected only at the longer IRAC wavelengths may correspond to embedded Class I or Class 0 sources. We include those YSOs in our Class I sample by adding sources detected at 8 with a 24 μm counterpart but not detected at 3.6 μm. This way we identify five additional Class I sources in region W5-east, 8 in S235, 18 in S252, 6 in S254-S258 and 7 in NGC 7538.

4.1.3 Sources with IRAC and NIR detections

Additional YSOs are identified by combining the IRAC-4.5 μm and the *H* and *K* bands (e.g. Winston et al. 2007; Gutermuth et al. 2008). Sources with IR-excess due to presence of a dusty disc around them are located to the right of the reddening vector in the $[H - K]$ versus [K–4.5] colour–colour diagram (Fig. 6). From this diagram, we find 285 sources with IR-excess in region W5-east, 439 in S235, 419 in S252, 364 in S254-S258 and 387 in NGC 7538. Of those, many were already classified as YSOs from their $\alpha_{\rm IRAC}$ value (see Section 4.1.1). Therefore, the number of additional YSOs identified per region are 135, 230, 237, 229 and 291 for W5-east, S235, S252, S254-S258 and NGC 7538, respectively. Since the additional YSOs are not detected in all IRAC bands, those are not classified in Class I or Class II. The same happens with the YSOs detected only in the NIR (see Section 4.1.4).

4.1.4 Sources detected only in the NIR

The NIR data is used to identify YSOs which are not detected in the IRAC bands due to bright PAHs emission and/or saturation in the IRAC bands (e.g. cluster S255-2 in region S254-S258, Chavarría et al. 2008). We apply the method from Chavarría et al. (2010) to identify YSOs by their NIR colour excess in the $[H - K]$ versus $[J - H]$ colour–colour diagram. First, we use the colour–colour diagram to estimate the reddening law E_{J-H}/E_{H-K} towards each region and then we classify sources as YSOs if they are located at least 1 sigma to the right of the reddening vector. We identify 39 additional YSOs in W5-east, 59 in S235, 34 in S254-S258 and 42 in NGC 7538. The derived E_{J-H}/E_{H-K} values are 1.38 for W5-east, 1.50 for S235, 1.69 for S252, 1.60 for S254-S258 and 1.68 for NGC 7538, all with errors up to 0.1.

By adding the identified YSOs from the previously described methods, the total number of YSOs per region is 478, 690, 679, 512 and 662 for W5-east, S235, S252, S254-S258 and NGC 7538, respectively. Their coordinates, magnitudes and classification can be found in the electronic version of this paper (see Table 2 for an example). The properties of YSOs per region are listed in Table 3. Previous identification of YSOs using IRAC+NIR data have been done by Koenig et al. (2008) and Dewangan & Anandarao (2011) in regions W5-east and S235, respectively. A comparison between their findings and our work is presented in Appendix B.

4.2 Spatial distribution of YSOs

Having identified the young population of stars, we investigate their spatial distribution by searching for concentrations of YSOs and by comparing the arrangement of different YSOs classes.

4.2.1 Minimum spanning tree (MST)

The MST is defined as a network of branches connecting points such as the total length of the branches is minimized and there is no loops (Battinelli 1991). This algorithm has lately become a popular tool to search for clusters of stars since it is independent from the stars density number. Nevertheless, it requires the definition of some maximum critical distance between cluster members (d_c). Gutermuth et al. (2009) used the MST to study embedded clusters in low-mass star-forming regions with d_c given by the turnover of the MST branches length distribution (see fig. 1 in their paper). In our case, since our observations cover larger fields/regions including several clusters, a single critical distance per region may not be representative for all clusters. Therefore, we divide each region in areas containing concentrations of YSOs either clearly separated (like AFGL4029, AFGL416 and G138.15+1.69 in W5-east, see Fig. 7) or associated with molecular material at different $V_{\rm LSR}$ (see Section 4.3). Then, we calculate the critical distance d_c for each area using the method from Gutermuth et al. (2009). The MST for each region is shown in Figs 7–11 for W5-east, S235, S252, S254-S258 and NGC 7538, respectively.

We use the MST to identify clusters as groups of $n = 10$ or more YSOs connected by branches shorter than d_c (for region

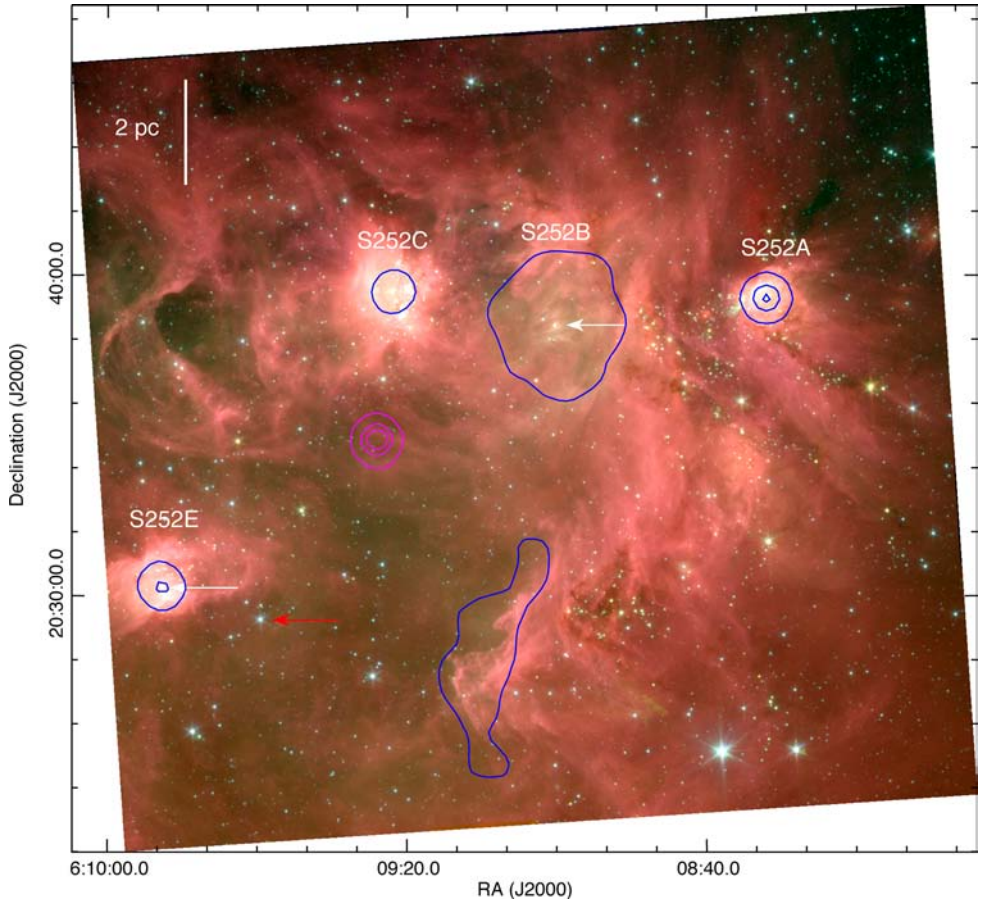

Figure 3. IRAC 3-colour image of region S252 (blue: 3.6 μm, green: 4.5 μm, red: 8.0 μm). The contours are the same as for Fig. 1. The ionizing star HD42088 is indicated by a red arrow. Other known ionizing stars are indicated by white arrows.

S254-S258, we used $n = 7$ in order to identify all clusters from Chavarría et al. 2008). We find in total 41 embedded clusters. Out of those, 15 have not been found in the literature until date. All clusters and their properties are listed in Table 4: columns 1 to 7 list the cluster name, centre position (given by the members coordinates mean), total number of members ($N_{\rm YSO}$), number of Class I and Class II and ratio of Class I to Class II sources; column 8 list the cluster members average $\alpha_{\rm IRAC}$ value ($\widetilde{\alpha}_{\rm IRAC}$); column 9 list the radius of the minimum area circle enclosing all cluster members ($R_{\rm C}$); column 10 list the hull radius ($R_{\rm H}$, see Section 4.2.3); column 11 list the aspect ratio (AR or cluster roundness) $R_{\rm C}^2/R_{\rm H}^2$; columns 12 and 13 list, respectively, the mean and peak surface density of YSOs inside the cluster convex hull area; columns 14 and 15 list the critical distance (d_c) and the mean branches length between cluster members ($s_{\rm YSO}$); columns 16 and 17 list the structural $\mathcal{Q}$ parameter value and error (see Section 4.2.3) and column 18 list the corrected number of cluster members ($N*_{\rm YSO}$, see Appendix C).

4.2.2 IR-excess sources surface density map

We use YSOs surface density maps to complement the MST cluster finding algorithm. For each region, the FOV is divided into a 3 arcsec grid. Then, at each point of the grid, the YSOs surface density ($\Sigma_{\rm YSO}$) is calculated from:

$$\Sigma_{\rm YSO} = \frac{N}{\pi r_N^2}, \quad (1)$$

where r_N is the distance to the $N = 5$ nearest neighbour (NN). Surface density maps for the studied regions are shown in Figs 7–11. Both MST and $\Sigma_{\rm YSO}$ methods identify basically the same YSOs groups as clusters. In general, MST clusters are enclosed by surface density contours of around 10–50 stars per pc^{-2}. For a more detailed comparison between these and other cluster finding algorithms, we recommend the work by Schmeja (2011).

4.2.3 The structural $\mathcal{Q}$ parameter and K–S tests

In this section, we describe the tools used to compare in a more quantitatively way the distribution of different class YSOs in the observed regions and clusters: the $\mathcal{Q}$ parameter and the Kolmogorov–Smirnov (K–S) test. The $\mathcal{Q}$ parameter (Cartwright & Whitworth 2004; Schmeja & Klessen 2006) is used to measure the level of hierarchical versus radial distribution of a set of (x_N, y_N) points, and it is defined by

$$\mathcal{Q} = \frac{\bar{l}_{\rm MST}}{\bar{s}}, \quad (2)$$

where $\bar{l}_{\rm MST}$ is the MST normalized mean branches length and $\bar{s}$ is the normalized mean separation between points. The MST mean branches length $l_{\rm MST}$ is calculated directly from the MST total length divided by the number of branches $(N - 1)$. Then, it is normalized by $\sqrt{A/N}$, where the convex hull area A is the area of a polygon enclosing all points, with internal angles between two contiguous sides of less than 180°. In a similar way, the mean separation

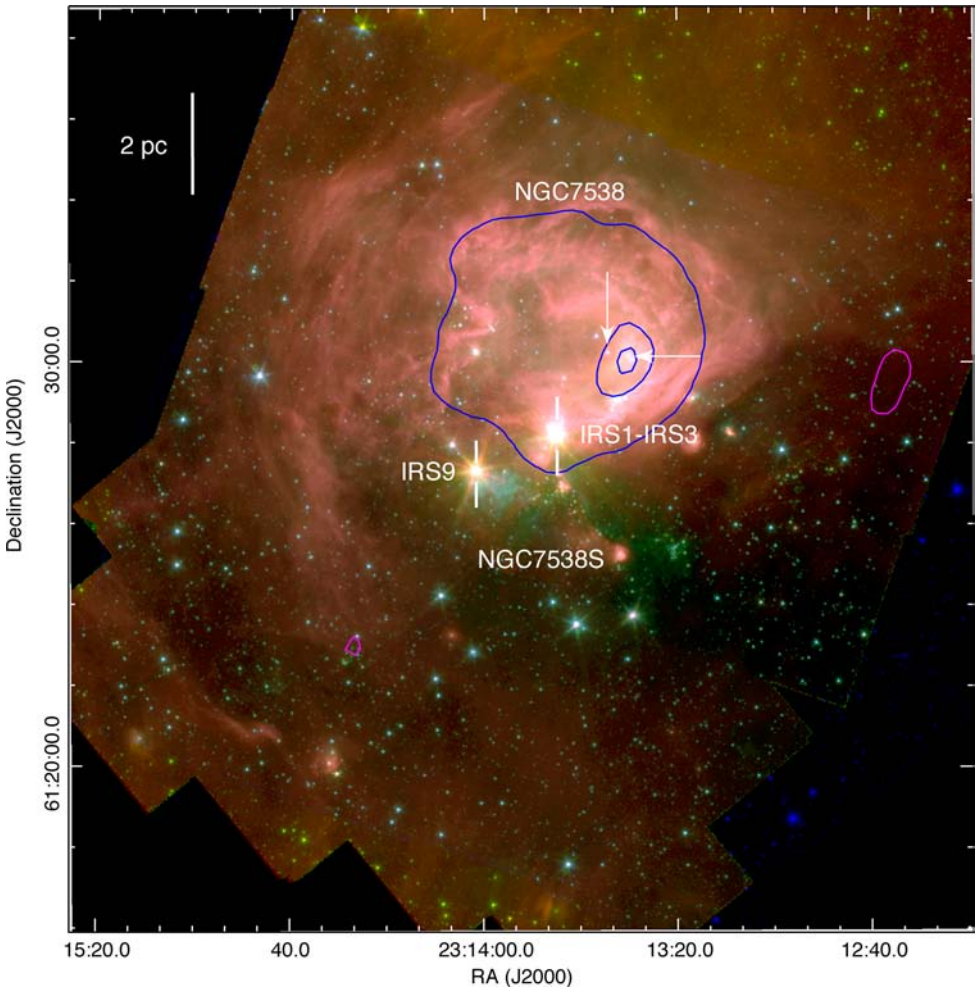

Figure 4. IRAC three-colour image of region NGC 7538 (blue: 3.6 μm, green: 4.5 μm, red: 8.0 μm). The contours are the same as for Fig. 1. The white arrows indicate the known ionizing stars. Infrared sources IRS1-IRS3 and IRS9 are indicated by bars.

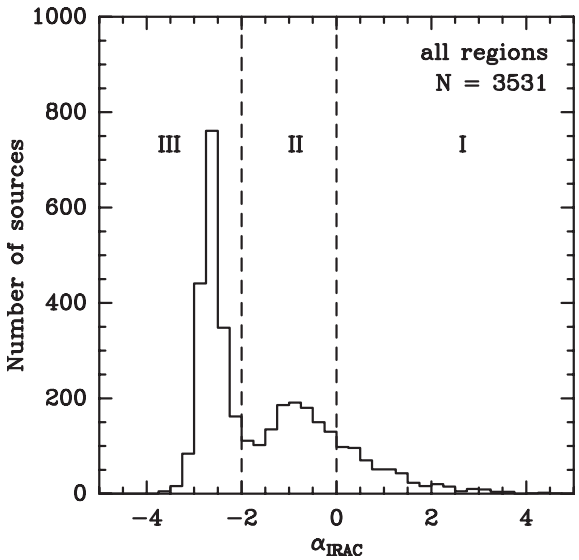

Figure 5. Observed spectral indices distribution for sources detected in the four IRAC bands (see Section 4.1.1). Dashed lines separate Class I and Class II sources. The minimum around $\alpha_{IRAC} = -2$ is also seen in other studies (Kumar et al. 2007; Muench et al. 2007).

between points s is normalized by R_H, the radius of a circle with area A. Using the normalized values $\bar{l}_{MST}$ and $\bar{s}$, the $\mathcal{Q}$ parameter becomes independent from the cluster size (Schmeja & Klessen 2006). According to Cartwright & Whitworth (2004), both $\bar{l}_{MST}$ and $\bar{s}$ values decrease as the degree of radial concentration increases (or the level of hierarchy decreases). However, $\bar{s}$ decreases faster than $\bar{l}_{MST}$. This way, a group of (x_N, y_N) points distributed radially will have a high $\mathcal{Q}$ value ($\mathcal{Q} > 0.8$) while clusters with a more fractal distribution will have a low $\mathcal{Q}$ value ($\mathcal{Q} < 0.8$) (Cartwright & Whitworth 2004). We also assign an error to $\mathcal{Q}$ given by the variation of its value over the range d_c and $d_c + 0.2 \times d_c$.

The K–S test is a statistical tool commonly used to investigate if two data sets are significantly different. In practice, the K–S test gives the probability that the hypothesis statement 'both data sets are drawn from the same distribution', is true. We define that two data sets are significantly different if their K–S probability is of less than 1 per cent ($P < 0.01$).

4.3 The molecular content around embedded clusters

We use extinction maps together with ^{12}CO and ^{13}CO maps to derive the physical properties of the molecular material associated with embedded clusters.

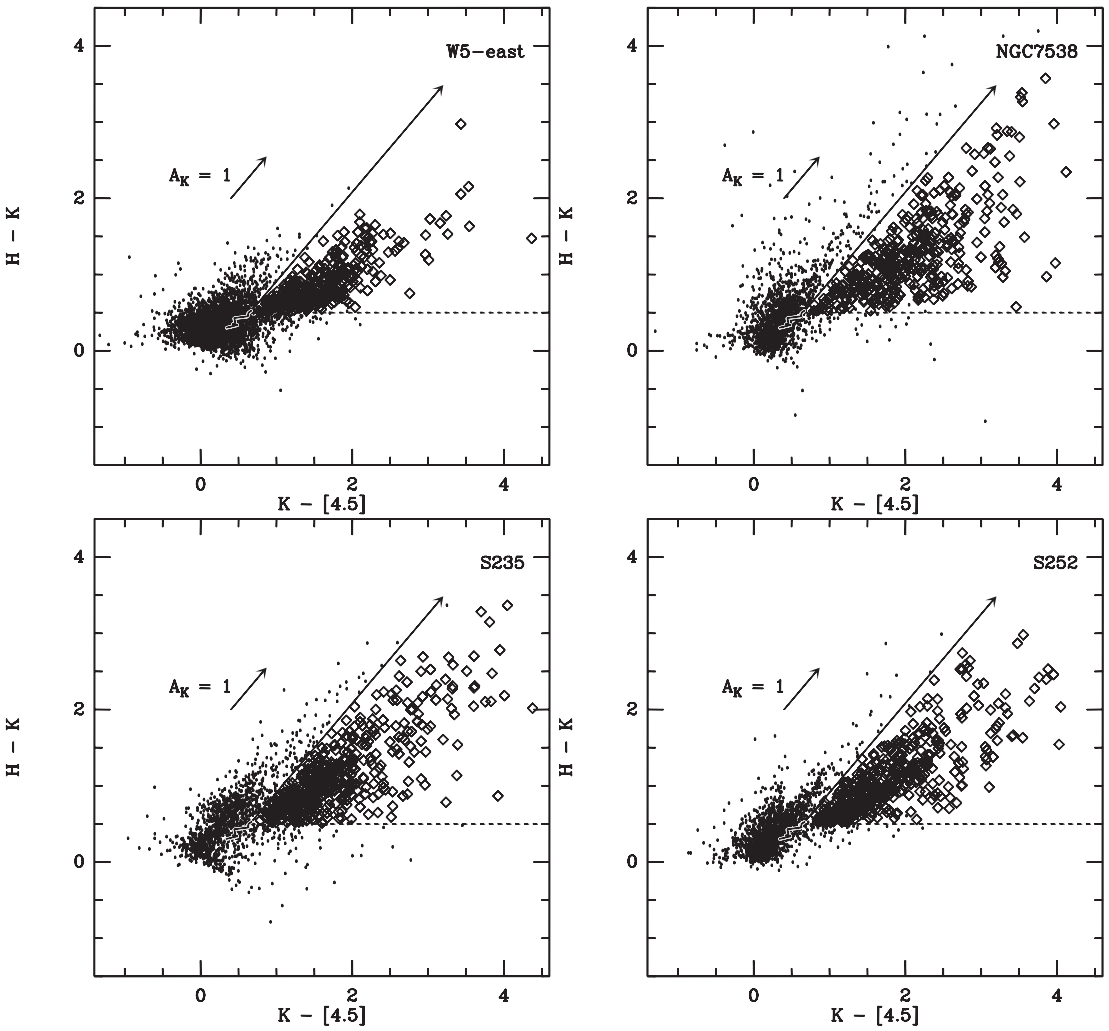

Figure 6. *H*, *K* and IRAC 4.5 μm band colour–colour diagrams for each region. The long arrow corresponds to the reddening vector (Flaherty et al. 2007). Stars with infrared excess, shown as diamonds, lay between the reddening vector and the dashed horizontal line. The black line near (0,0), close to the reddening vector base shows the main-sequence locus of a late type M dwarf star (Patten et al. 2006).

4.3.1 K-band extinction maps

The extinction maps were created from all sources detected in the NIR bands and located in the main-sequence region in the $[H-K]$ versus $[J-H]$ colour–colour diagram. Following the method from Chavarría et al. (2010), we derive the reddening law E_{J-H}/E_{H-K} towards each region (see Section 4.1.4) and estimate the sources *K*-band extinction (A_K) by dereddening them to their main-sequence colours. Then, the observed FOV is divided into a 3 arcsec grid and the average A_K values between the five NNs to the grid centre are assigned to each grid point. Since the studied regions are located at a few kpc from the Sun, foreground stars extinguished mostly by interstellar dust will 'artificially' decrease the average A_K value. This will lead to an underestimation of the extinction due to the molecular cloud. To correct this, we remove from each grid point the two sources with lower extinction if the grid A_K standard deviation is more than 0.5 mag (typical values are between 0.2 and 0.3 mag). After applying this correction, the extinction in the most embedded areas increased by 0.3–0.5 mag.

The clusters molecular mass were estimated from the extinction maps. We use the relations from Dickman (1978) and Cardelli, Clayton & Mathis (1989) to estimate the H_2 column density in each grid from the average A_K values using:

$$N(H_2) = 1.25 \times 10^{21} \left(\frac{A_K - A_{K,fg}}{0.114} \right) \mathrm{cm}^{-2}\,\mathrm{mag}^{-1}, \quad (3)$$

where $A_{K,fg} = 0.15 \times$ distance[kpc] is the interstellar contribution to the extinction (Indebetouw et al. 2005). Then, the cluster molecular mass is calculated by integrating the column density over each cluster convex hull area and multiplying by the H_2 molecule mass. Extinction maps are shown in Figs 12–16.

The properties of the molecular material associated with embedded clusters (meaning inside the convex hull area) are listed in Table 5: the mean and peak *K*-band extinction are listed in columns 2 and 3, respectively; molecular mass derived from the cluster extinction and in regions where $A_K > 0.8$ mag (see Section 5.3) are listed in columns 4 and 5, respectively; mean and peak ^{13}CO column densities are listed in columns 6 and 7, respectively (see Section 4.3.2); molecular mass derived from the cluster CO column density is listed in column 8; the ^{13}CO V_{LSR} and full width at half-maximum (FWHM) of the integrated ^{13}CO emission over the hull area are listed in columns 9 and 10, respectively.

Table 2. Identified YSOs per region[a].

ID	Region	RA J(2000)	Dec. J(2000)	*J* (mag)	*H* (mag)	*K* (mag)	[3.6] (mag)	[4.5] (mag)	[5.8] (mag)	[8.0] (mag)	Class[b]	α_{IRAC}	Method[c]
1	W5-east	45.090 149	60.301 571	–	13.86(0.01)	13.64(0.01)	13.55(0.04)	13.52(0.04)	13.42(0.06)	12.80(0.08)	II	−1.99	IRAC
2	W5-east	45.097 431	60.530 960	16.46(0.07)	15.36(0.04)	14.69(0.03)	13.56(0.04)	13.30(0.04)	12.94(0.05)	–	–	–	HK2
3	W5-east	45.100 380	60.543 320	15.88(0.02)	14.83(0.01)	14.23(0.01)	13.43(0.04)	13.02(0.03)	12.36(0.04)	–	–	–	HK2
4	W5-east	45.106 079	60.469 429	13.73(0.01)	–	–	11.57(0.02)	11.86(0.02)	10.87(0.02)	10.40(0.02)	II	−1.25	IRAC
5	W5-east	45.107 029	60.307 461	15.43(0.02)	14.57(0.02)	14.04(0.01)	13.72(0.04)	13.30(0.04)	13.06(0.05)	12.22(0.05)	II	−1.32	IRAC
6	W5-east	45.108 589	60.415 009	15.08(0.01)	14.27(0.01)	12.99(0.01)	–	–	–	–	–	–	NIR
7	W5-east	45.113 110	60.461 498	17.16(0.07)	16.89(0.11)	15.55(0.04)	–	–	–	–	–	–	NIR
8	W5-east	45.113 350	60.369 598	17.58(0.08)	16.85(0.09)	15.74(0.04)	–	–	–	–	–	–	NIR
9	W5-east	45.114 182	60.460 030	12.87(0.01)	12.17(0.01)	11.55(0.01)	10.89(0.01)	10.62(0.01)	9.92(0.01)	9.45(0.01)	II	−1.10	IRAC
10	W5-east	45.114 819	60.479 900	15.04(0.01)	14.07(0.01)	13.68(0.01)	13.49(0.04)	13.16(0.04)	13.08(0.05)	12.32(0.05)	II	−1.66	IRAC

[a] Table 2 is published in its entirety in the electronic version of this paper. A portion is shown here as a guidance. Values in parentheses by photometry signify error in magnitudes.
[b] YSOs are classified in Class I or II following the description in Section 4.1.
[c] This column shows the method used to identify the YSO: α_{IRAC} slope (IRAC, see Section 4.1.1), excess in the *H*, *K* and [4.5] bands (HK2, see Section 4.1.3), excess in the NIR bands (NIR, see Section 4.1.4) or detection in the MIPS 24 μm band (MIPS, see Section 4.1.2).

4.3.2 ^{12}CO and ^{13}CO emission

The CO observations trace the less dense gas which is more difficult to detect using the extinction maps. In addition, CO maps contain velocity information that allows us to disentangle different molecular components along the line of sight.

We created ^{13}CO column density maps for all regions by calculating the ^{13}CO column density at each CO grid point (see Section 3.3) using (Heyer et al. 1989):

$$N(^{13}\mathrm{CO}) = 2.42 \times 10^{14} \left(\frac{\Delta v T_{\mathrm{ex}} \tau}{1 - \mathrm{e}^{-5.29/T_{\mathrm{ex}}}} \right) \mathrm{cm}^{-2}, \quad (4)$$

where

$$T_{\mathrm{ex}} = \frac{5.53}{ln\left(1 + \frac{5.53}{T_A + 0.82}\right)}, \quad (5)$$

$$\tau = -ln\left(1 - \frac{T_{\mathrm{ex}}}{\frac{5.29}{e^{5.29/T_{\mathrm{ex}}} - 1} - 1}\right), \quad (6)$$

T_A is the observed peak temperature of the ^{12}CO, and Δv correspond to the FWHM of the ^{13}CO line (see the maps in Figs 12–16). The FWHM was obtained by fitting a Gaussian profile to each grid velocity spectrum. From the ^{13}CO column density maps, we estimate the gas mass by integrating the column density over the clusters convex hull area (assuming $^{12}\mathrm{CO}/^{13}\mathrm{CO} = 50$ and $^{12}\mathrm{CO}/\mathrm{H}_2 = 8.5 \times 10^{-5}$ from Israel et al. 2003; Frerking, Langer & Wilson 1982, respectively). The derived ^{13}CO column densities and gas masses are summarized in Table 5.

In order to distinguish different molecular components along the line of sight, we used the ^{13}CO first moment maps to identify the V_{LSR} of the peak emission at each grid point (Figs 12–16). In addition to this, we estimate the V_{LSR} of the molecular material associated with each embedded cluster by integrating the ^{13}CO spectra over the cluster convex hull area (assuming that ^{13}CO is more optically thin than ^{12}CO) and then adjusting Gaussians to the integrated line profile. For simplicity, we assume that the V_{LSR} of the molecular material associated with each cluster corresponds to the strongest Gaussian component. As an example, Fig. 17 shows the two Gaussian components used to fit the ^{13}CO integrated spectra for cluster AFGL416 (the integrated ^{13}CO spectrum for the other clusters are shown in Figs E1–E5). The V_{LSR} and FWHM of the Gaussian component associated with each cluster are also listed in Table 5.

5 DISCUSSION

5.1 Spatial distribution of identified YSOs

In general, YSOs are distributed in groups surrounded by a more scattered population (see Figs 7–11). The distribution parameter $\mathcal{Q}$ (see Section 4.2.3) for all YSOs in each region has values in the hierarchical range between 0.64 and 0.81 (see Table 3). This is not surprising, since each region contains several clusters resembling a fractal distribution. Synthetic cluster analysis from Maschberger et al. (2010) give also hierarchical $\mathcal{Q}$ values for a globally unbound 10^4 $\mathrm{M}_\odot$ cloud, which is comparable in mass and size to our regions.

5.1.1 *Scattered YSOs population*

The scattered or isolated population are YSOs not included in clusters. In our case, they range between 30 and 50 per cent of the

Table 3. Properties of detected YSOs per region[a].

	W5-east	S235	S252	S254-S258	NGC 7538
$N_{\rm YSO}$	478	690	679	512	662
Parameter $\mathcal{Q}_{\rm YSO}$	0.69	0.74	0.64	0.68	0.81
$N_{\rm I}$	62	113	169	87	108
$N_{\rm II}$	242	288	273	162	221
$\overline{A}_{K,\rm I}$ (mag)	0.98	1.13	0.77	1.19	1.15
$\overline{A}_{K,\rm II}$ (mag)	0.48	0.95	0.75	1.02	0.75
$\overline{\Sigma}_{\rm YSO,I}$ (pc^{-2})	58	82	35	52	26
$\overline{\Sigma}_{\rm YSO,II}$ (pc^{-2})	24	36	15	42	11
$\overline{\sigma}_{\rm CO,I}$ (10^{16} cm^{-2})	1.50	2.38	2.79	2.13	2.60
$\overline{\sigma}_{\rm CO,II}$	0.74	1.53	1.90	1.46	1.72
per cent of YSOs in clusters	0.52	0.61	0.63	0.73	0.69
per cent of I in clusters	0.66	0.73	0.76	0.71	0.79
per cent of II in clusters	0.33	0.46	0.42	0.62	0.51

[a] $\overline{A}_K$ values are corrected by interstellar extinction.

detected YSOs per region (see Table 3). This is in agreement with the 10 to 50 per cent expected depending on the adopted cluster definition (Koenig et al. 2008, suggests that there is at least 10 per cent of scattered YSOs in massive star-forming regions).

The presence of a scattered population can be due to several reasons: dynamical interaction between cluster members, small groups merging, cluster definition and isolated star formation. In the first case, close encounters between cluster members may expel some at high velocities and place them far from its former companions. In the second case, the merging between small YSOs groups (with less members than the required number to be considered as a cluster) may dissipate cluster members forming a scattered halo whose members have a longer critical distance and will not be considered as part of the cluster (numerical simulations on hierarchically formed star clusters gives sub-clusters merging time-scales of a few 10^5 years, which is comparable to the observed clusters age, Maschberger et al. 2010). Thirdly, the number of cluster members depends on the critical distance d_c. However, cluster dynamics and cluster merging creates a mixture of YSOs with different properties that may not be represented by a single d_c value (e.g. Maschberger et al. 2010, showed that the percentage of cluster members for a 10^3 M$_\odot$ simulation slowly decreases due to the merging of several small groups). In such a case, the used cluster finding algorithm may not identify properly all cluster members. Finally, the formation of isolated stars cannot be ruled out as another alternative. Bressert et al. (2012) and Oey et al. (2013) claim to have found several isolated young OB stars in 30 Doradus and the Small Magellanic Cloud, respectively.

When we compare the scattered population in the studied regions per class, we find that it is mainly composed by Class II sources (67 per cent, see Table 3). However, this difference decreases if we use a critical distance of $2 \times d_c$ (53 per cent of Class II). If we assume that Class II sources are more evolved than Class I, the former are expected to be more dynamically relaxed and therefore more spread than the Class I. As a consequence, some Class II sources located at the outer parts of the cluster are likely to have a projected distance larger than d_c to the nearest cluster companion and will not be considered as cluster members. Following this argument, we estimate that between 10 and 20 per cent of the scattered population in the observed regions may correspond to dynamically evolved cluster members.

5.1.2 Class I versus Class II sources

In average, Class I sources are located towards places with higher extinction (A_K) and column density ($\sigma_{\rm CO}$) than Class II sources (see Table 3 and Fig. 18). The histograms and cumulative distribution functions in Fig. 18 show that both Class I and Class II are distributed over a similar range of A_K, $\sigma_{\rm CO}$ and $\Sigma_{\rm YSO}$ values. However, Class I histograms are skewed towards higher values. This difference is better quantified by the K–S tests, which show that the probability that both classes have the same distribution is in general low ($P_{\rm (I-II)} \lesssim 10^{-2}$, see Fig 18). In addition, we find that Class I sources are more close to each other than Class II (see $\overline{\Sigma}_{\rm YSO,I}$ versus $\overline{\Sigma}_{\rm YSO,II}$ in Table 3). These properties agree well with previous findings in region W5 (Koenig et al. 2008; Deharveng et al. 2012) and with the assumed evolutionary stage of both classes: the youngest Class I sources are more clustered and associated with the most dense molecular material in which they were born while the Class II sources are moving away from their birthplace due to dynamic interactions. Also, as shown in Section 5.2.3, the embedded clusters in these regions are likely gravitationally unbound. This way, the more evolved cluster members will move further away due to the weak pulling from the cluster gravitational well.

5.2 Physical properties of embedded clusters

In this section, we investigate the physical properties of identified embedded clusters and look for trends that will help us to better understand their formation process. Clusters G192.54−0.15 in region S254-S258, G111.59+0.64 and G111.65+0.63 in region NGC 7538 and G189.79+0.29 in S252 will not be included in our analyses. The former is likely located at a much further distance than the rest of clusters in the region (Chavarría et al. 2008). The other three were not observed completely in the *K*-band, which reduces number of identified cluster members. We also assume that cluster S235C is part of cluster S235AB due to its location and similar $V_{\rm LSR}$.

The 36 remaining clusters are divided in two sets: MEC and LEC. MEC are clusters known to harbour massive stars while LEC show no evidence of containing massive stars, even though some of them may actually form massive stars in the future. Following this definition, 12 embedded clusters are MEC: AFGL4029, AFGL416,

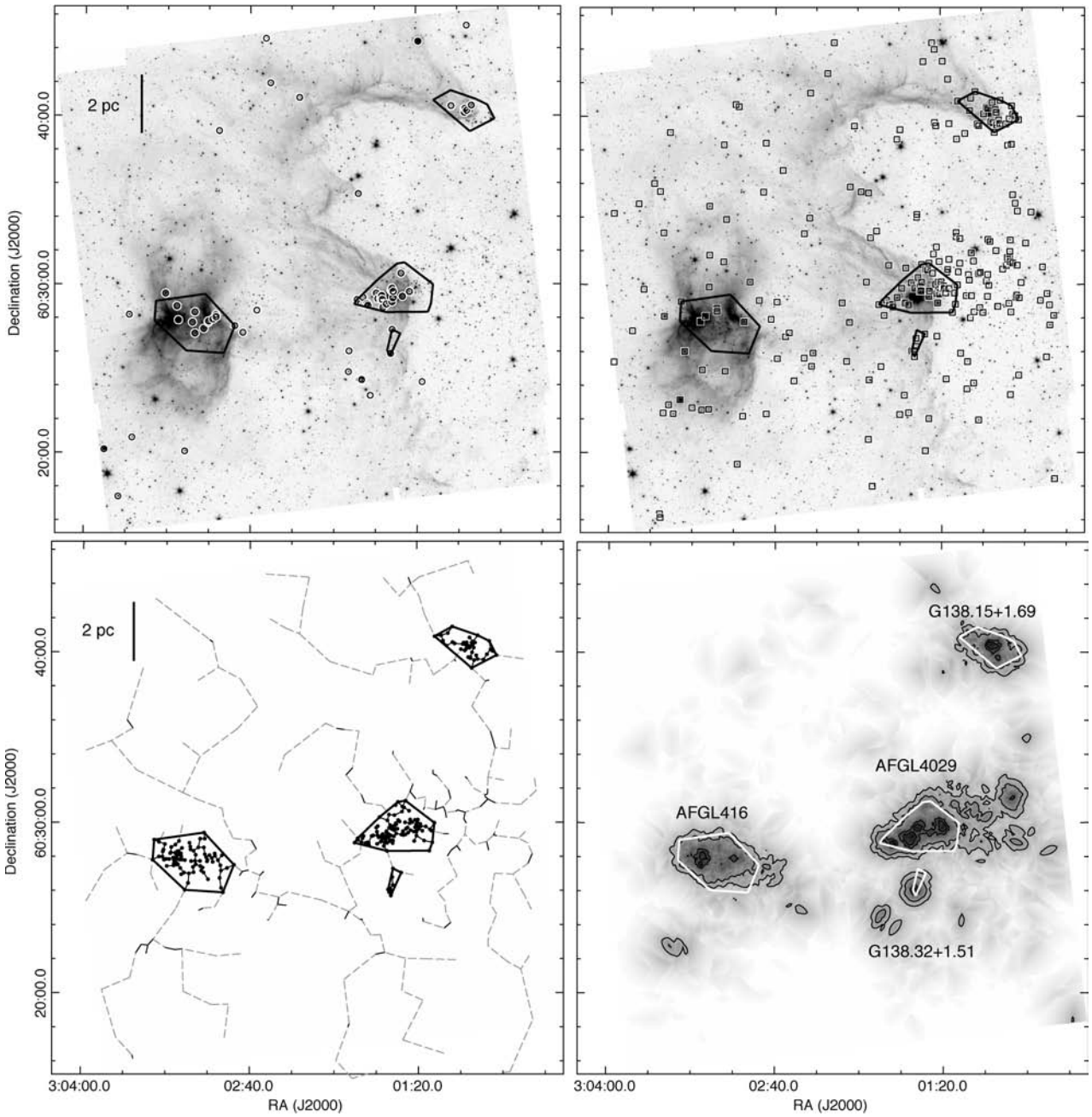

Figure 7. Spatial distribution of YSOs in W5-east. Polygons enclosing cluster members (convex hull) are shown in all panels. Upper: sources identified as Class I (left) and Class II (right) are shown over the IRAC 4.5 μm mosaic. Lower left: minimum spanning tree (MST) of YSOs positions (including Class I, Class II and IR-excess sources). Contiguous sources separated by a projected distance of less than the critical distance d_c (see Table 4) are joined by a continuous black line. Groups of YSOs that qualify as clusters are marked by black dots. Lower right: surface density of YSOs. Contour levels are at 5, 10, 50 and 100 stars pc^{-2}. Cluster names are labelled in this panel.

S235AB, S235, S252A, S252C, S252E, S256, S255-2, G192.63-0.00, S258 and NGC 7538. The other 24 clusters are LEC.

5.2.1 *Spatial distribution of cluster members*

We characterize the distribution of members in embedded clusters by their surface density ($\Sigma_{\rm YSO}$). The YSOs in our cluster sample have mean surface densities between 10 and 100 pc^{-2} with no significant difference between MEC and LEC (see Fig. 19). However, MEC show significantly higher peak surface densities than LEC. In addition, we find a weak trend between the peak surface density and the number of cluster members, with a correlation index $r = 0.75$ and exponent of 0.8 ± 0.2. Since the embedded clusters have their most agglomerated population concentrated in small areas compared to the cluster extent (which can also be seen in Figs 7–11, where the surface density contours of more than 50 stars pc^{-2} enclose small areas compared to the cluster size), this correlation suggest that clusters are better characterized by their peak YSOs surface density.

Fig. 19 also shows the surface densities derived by Gutermuth et al. (2009) in their low-mass EC sample. Their mean surface densities are somehow higher than our measurements. This may be

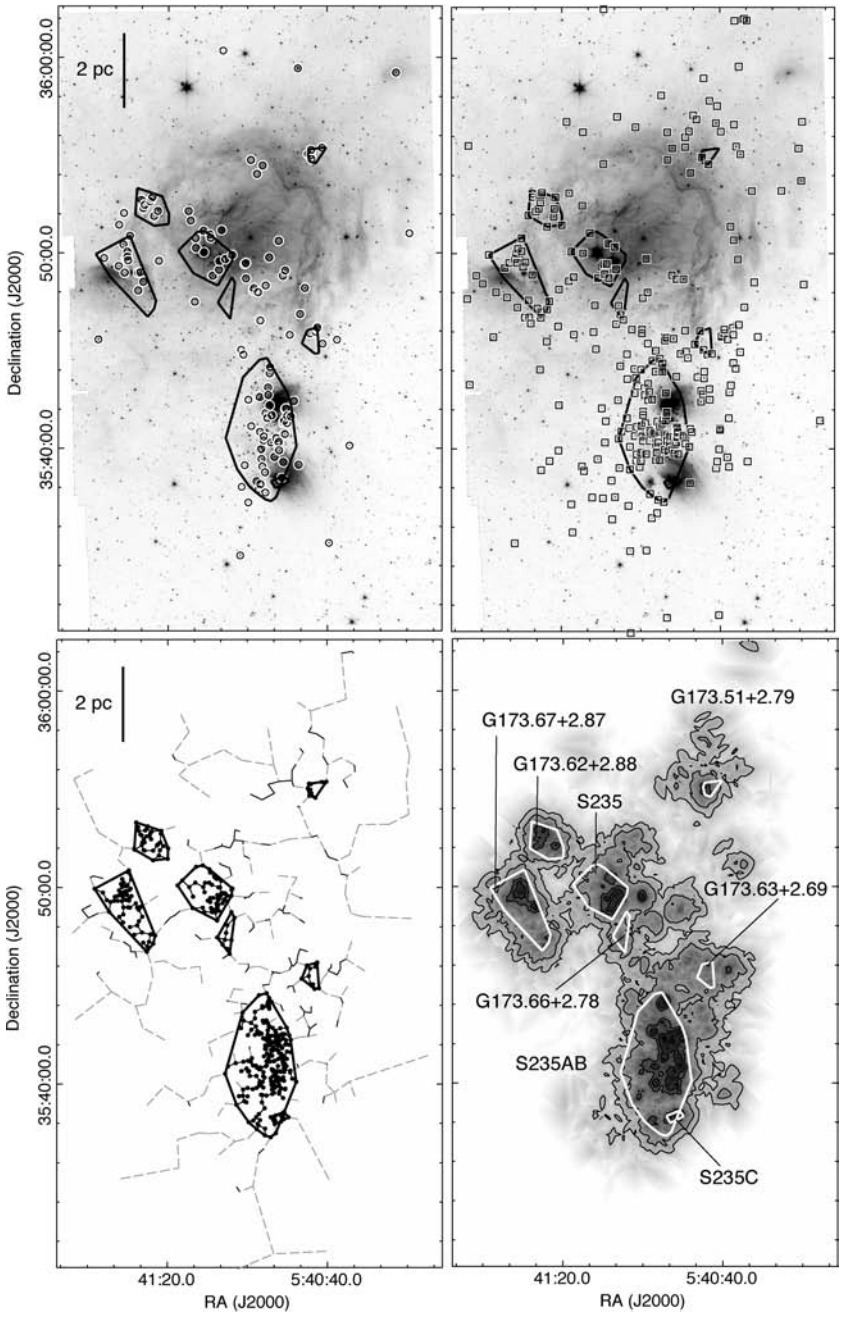

Figure 8. Same as in Fig. 7 for region S235.

a consequence of the shorter distance to their clusters, which allows the detection of fainter cluster members. We tested this hypothesis by estimating the mean $\Sigma_{\rm YSO}$ using the corrected number of cluster members $N^*_{\rm YSO}$ (see Section C and Table 4). Fig. 20 shows that the surface densities from Gutermuth et al. (2009) are in agreement with our corrected mean $\Sigma_{\rm YSO}$ values.

The spatial distribution of cluster members is also investigated by their structural $\mathcal{Q}$ parameter value. We find that embedded clusters have an average $\mathcal{Q}$ value of 0.85 ± 0.1, ranging from 0.64 to 1.11. If we compare the $\mathcal{Q}$ values between MEC and LEC, we find that MEC have slightly lower values than LEC (with average of 0.78 ± 0.1 and 0.88 ± 0.1, respectively). This is also seen as a weak trend in the distribution of $\mathcal{Q}$ values per members number (see Fig. 21, where $N_{\rm YSO} \propto {Q_{\rm YSO}}^{-0.17 \pm 0.05}$ and $r = -0.68$). Such a correlation may be due to a higher occurrence of sub-clusters merging in the most massive clusters caused by their higher potential well, which decreases the value of the $\mathcal{Q}$ parameter (see models from Maschberger et al. 2010).

We compare the $\mathcal{Q}$ parameter for Class I and Class II sources in the most numerous clusters in each region (see Table 6). With the exception of cluster AFGL4029, all clusters have the Class I sources distributed more hierarchically than the Class II ($\mathcal{Q}_{\rm I} < \mathcal{Q}_{\rm II}$). However, cluster AFGL4029 is a special case since it is located near a group of Class II sources that may be cluster members as well (see Section 5.7.1). If we re-calculate the $\mathcal{Q}$ parameter for those clusters but increasing d_c by 25 per cent (hence including the western Class II sources in AFGL4029), we find that the Class I sources are distributed more hierarchically than the Class II in the

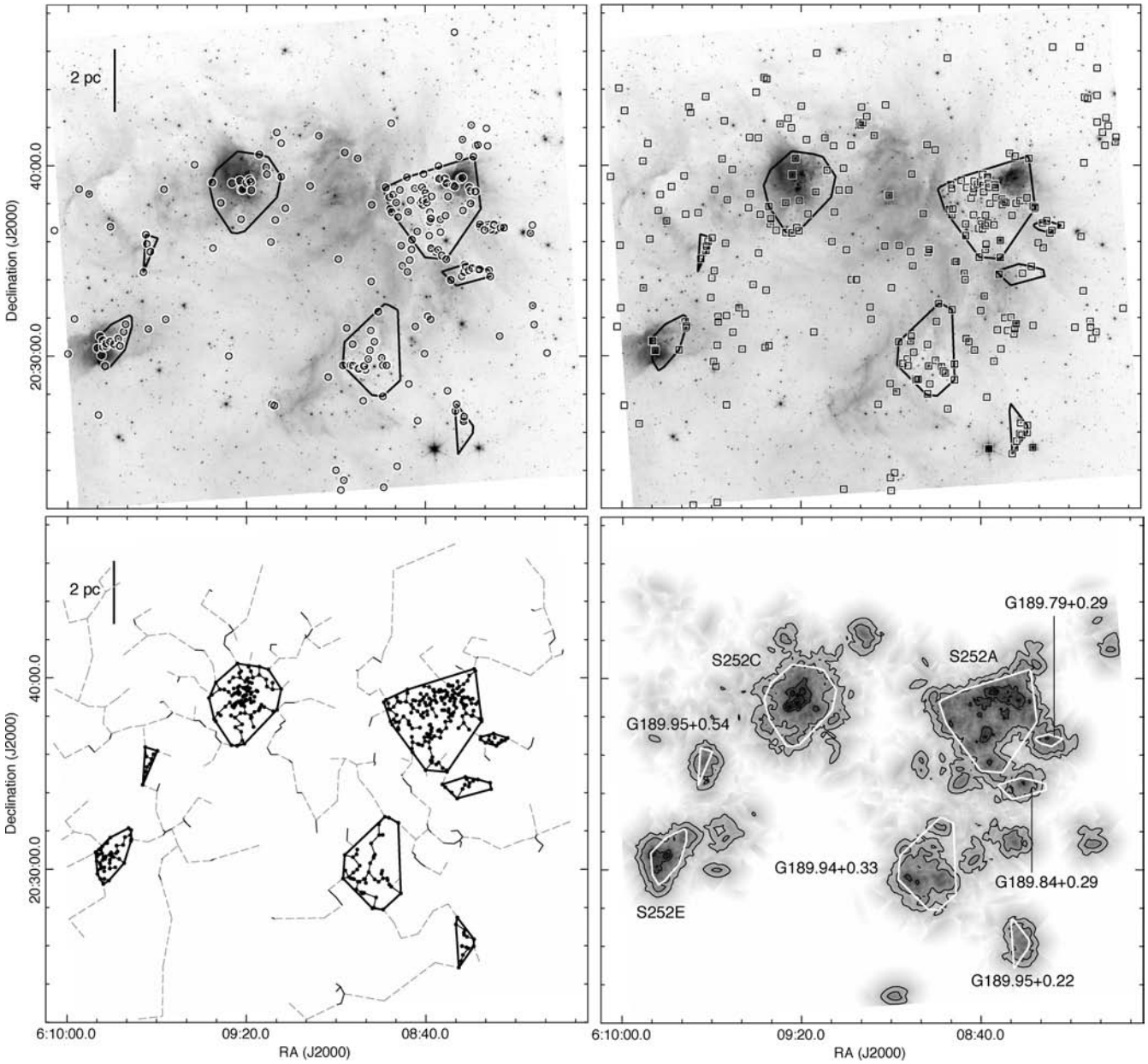

Figure 9. Same as in Fig. 7 for region S252.

five clusters. Similar results have also been found in low-mass star-forming regions (Schmeja et al. 2008) and is likely a consequence of the cluster dynamic relaxation.

5.2.2 *Clusters morphology*

We use the clusters convex hull radius ($R_{\rm H}$) and aspect ratio (AR) to investigate their morphology (see Fig. 22). Clusters $R_{\rm H}$ values range between 0.1 and 2.0 pc with a mean value of 0.67 ± 0.4 pc. We find that MEC are more extended than LEC, with mean $R_{\rm H}$ of 1.01 ± 0.4 pc and 0.50 ± 0.3 pc, respectively. This difference agrees with a low K–S similarity test value of $\sim 10^{-4}$. On the other hand, the clusters elongation, given by their aspect ratio (AR), range between 0.2 and 3.5 with a mean of 1.62 ± 0.6. This means that clusters are in general elongated (a circular cluster will have AR of unity). We find no significant differences between the LEC and MEC aspect ratios.

5.2.3 *Associated molecular material*

We find that clusters have mean A_K values (after subtracting the interstellar contribution) between 0.2 and 1.6 mag with an average of 1.1 ± 0.3 (see Fig. 23). We find no differences between the mean A_K for MEC and LEC. However, the peak A_K values are higher for MEC, with an average of $\sim$2.5 mag versus 1.5 mag for LEC. As for the peak YSOs surface density, there is a weak correlation between the peak A_K and the number of cluster members.

We integrate the ^{13}CO spectra over the clusters convex hull area to investigate the molecular cloud dynamics at a cluster-size scale. We use the ^{13}CO line since is more optically thin that the ^{12}CO line and traces more dense material as well as it may be less affected

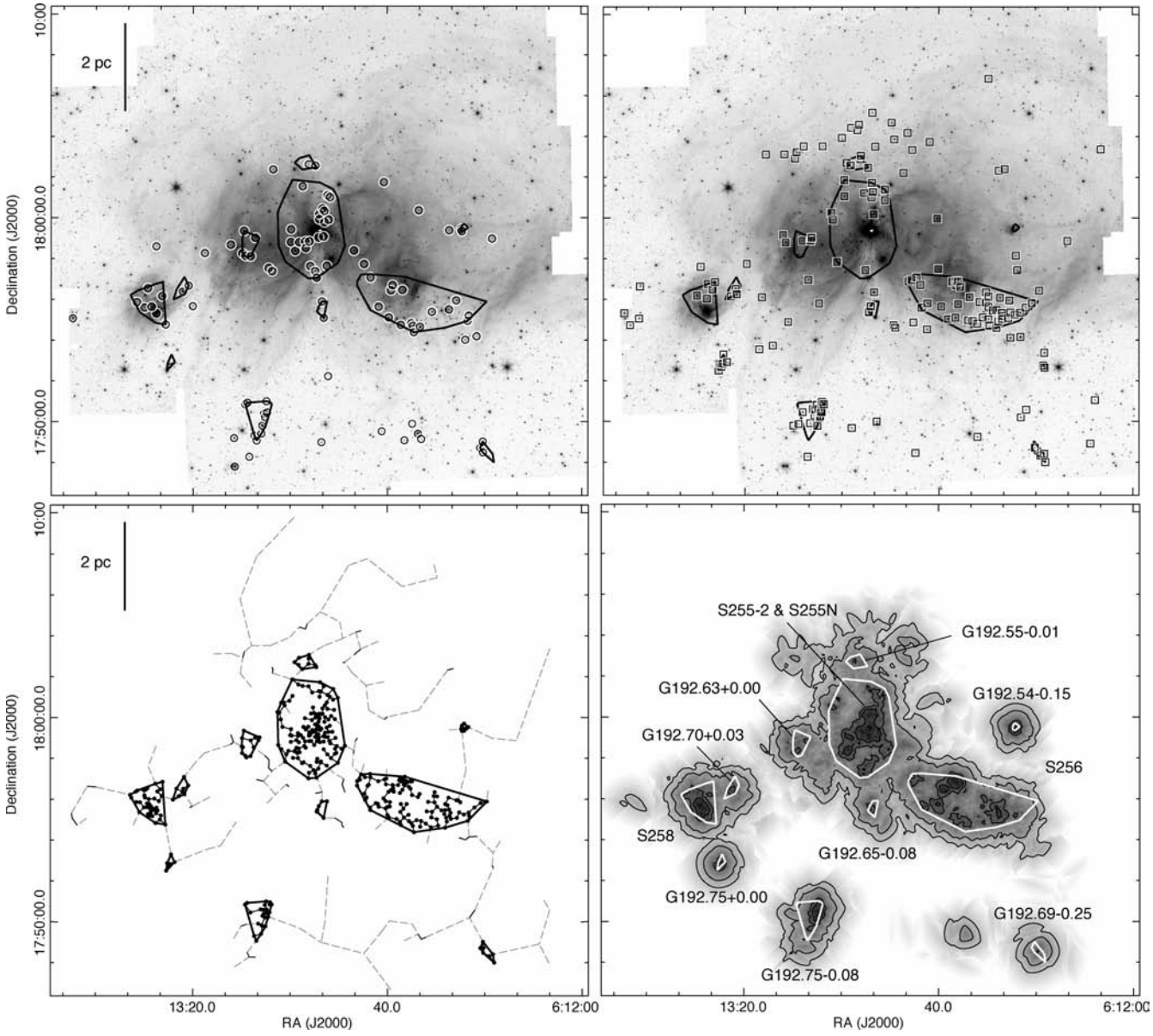

Figure 10. Same as in Fig. 7 for region S254-S258.

by outflow emission. We find no correlations between the ^{13}CO FWHM and cluster physical properties like YSOs surface density, size, extinction or cluster mass. However, all clusters have integrated ^{13}CO linewidths larger than the thermal line broadening for an assumed cloud temperature of 30 K (see Fig. 24):

$$\Delta v_{\rm th} = \left(\frac{8kT\ln(2)}{\mu_{\rm CO} m_{\rm H}}\right)^{1/2} \sim 0.3\,{\rm km\,s^{-1}}, \quad (7)$$

where k is the Boltzmann constant, $m_{\rm H}$ is the hydrogen mass and $\mu_{\rm CO}$ the ^{13}CO atomic weight. This implies that the main support against gravity in all clusters is mainly due to non-thermal processes like turbulence. We find no significant differences between the MEC and LEC ^{13}CO line broadenings.

The Virial parameter ($\alpha_{\rm vir}$) measures the ratio between the kinetic and gravitational energy and is defined by

$$\alpha_{\rm vir} = \frac{5\Delta {v_{\rm CO}}^2 R_{\rm H}}{G M_{\rm T}}, \quad (8)$$

where $\Delta v_{\rm CO} = {\rm FWHM}/(8\ln 2)^{1/2}$ is the dispersion velocity (of ^{13}CO), G is the gravitational constant, $M_{\rm T}$ is the total cluster mass (gas and stars) and $R_{\rm H}$ is the convex hull radius. The total mass is given by the maximum between M_{A_K} and $M_{\sigma_{\rm CO}}$ plus the corrected cluster members mass (assuming a mean mass of 0.5 $M_\odot$ per source, see Section C). We calculated $\alpha_{\rm vir}$ for all clusters and find values >1 in all cases. This suggest that clusters are mostly gravitationally unbound. However, the derivation of equation (8) has several assumption which may not apply for the observed clusters and this result must be treated carefully.

We also investigate the clusters fragmentation scale by calculating their Jeans length (λ_J). This is the minimum radius for the gravitational collapse of a homogeneous isothermal sphere, and is given by

$$\lambda_J = \left(\frac{15kT}{4\pi G m_{\rm H}\rho_o}\right)^{1/2}, \quad (9)$$

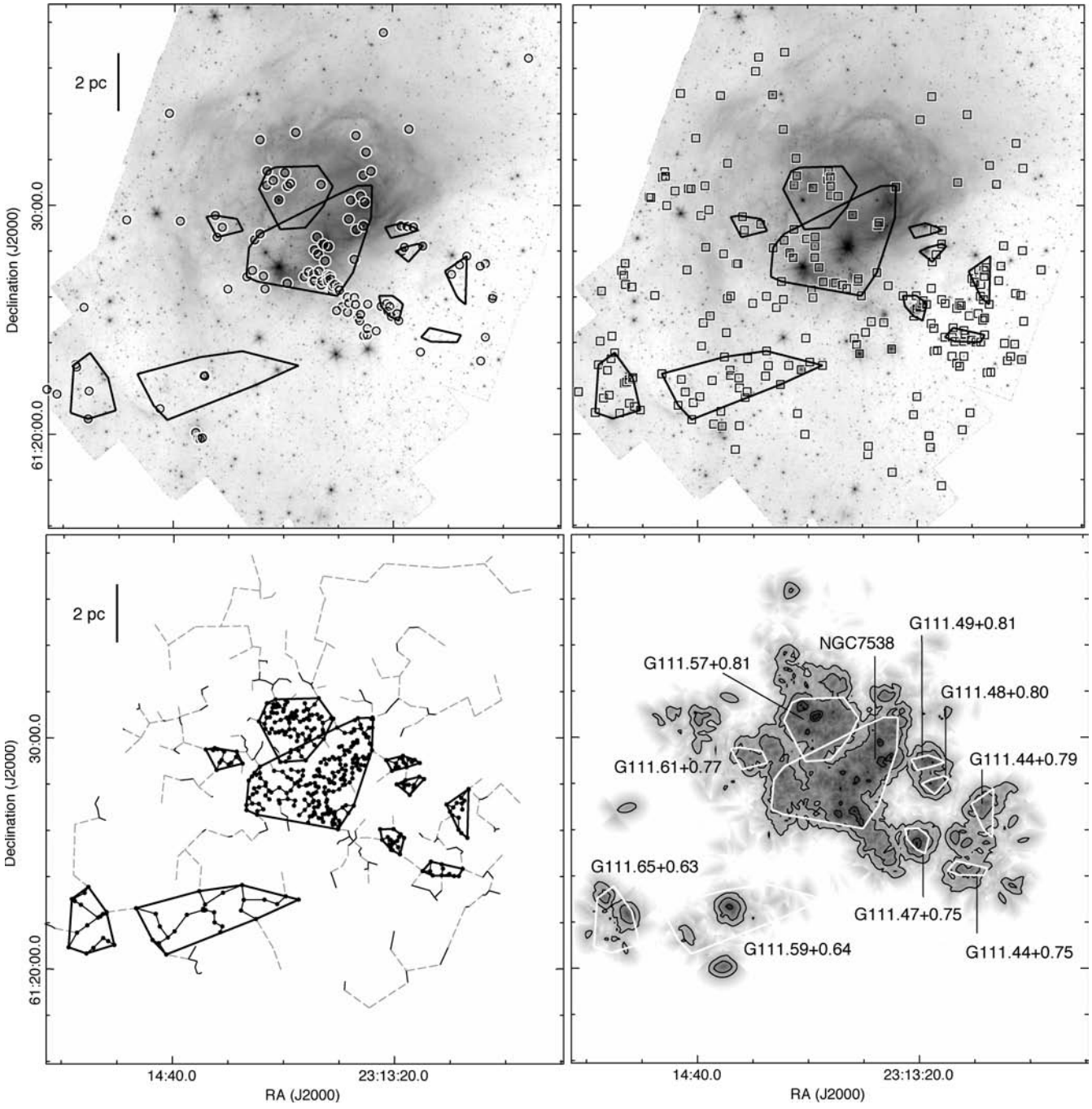

Figure 11. Same as in Fig. 7 for region NGC 7538.

where $m_{\rm H}$ is the hydrogen atom mass, T is the cloud temperature (30 K) and the mean density ρ_o is defined by

$$\rho_o = \frac{3}{4\pi}\frac{M_{\rm T}}{R_{\rm H}^3}. \quad (10)$$

We compare λ_J and the mean separation between cluster members ($s_{\rm YSO}$, Table 4) and find that the ratio $\lambda_J/s_{\rm YSO}$ has values higher than 1 for all clusters, with an average of 4.3 ± 1.5 (see Fig. 25). We also find that MEC have larger $\lambda_J/s_{\rm YSO}$ ratios than LEC. Those results agree with a non-thermal driven fragmentation since it occurs at scales smaller than the Jeans length. They also suggest that non-thermal processes may have a more pronounced effect in clusters harbouring massive stars. Alternatively, a higher potential well in MEC may keep the cluster members closer to each other and for a longer time. This can also explain the $\lambda_J/s_{\rm YSO}$ ratio differences between LEC and MEC.

We note that the $s_{\rm YSO}$ values are calculated from the projected distance between cluster members. This means that the real distance is on average longer by a factor of 2–3. However, this effect could be counteracted by including the non-detected cluster members, which decreases the $s_{\rm YSO}$ values by a similar factor.

Magnetic fields may also play an important role on the clusters fragmentation level. Observations show that magnetic fields become stronger at higher densities in star-forming regions, which means

Table 4. Properties of embedded clusters[a].

Name	RA (J2000)	Dec. (J2000)	N_{YSO}	I	II	I/II	$\widetilde{\alpha}_{IRAC}$	R_C (pc)	R_H (pc)	AR	$\overline{\Sigma}_{YSO}$ (pc^{-2})	Σ_{YSO} (pc^{-2})	d_c (pc)	s_{YSO} (pc)	$\mathcal{Q}_{YSO}$	$\delta\mathcal{Q}$	N^*_{YSO}
G138.15+1.69	03:00:55.5	60:40:18.7	42	6	23	0.26	−0.98	1.10	0.76	2.12	28	591	0.30	0.140	0.82	0.01	71
AFGL4029	03:01:30.6	60:29:31.4	114	23	43	0.53	−0.67	1.40	1.03	1.86	42	708	0.32	0.104	0.73	0.05	206
G138.32+1.51	03:01:32.3	60:26:20.2	11	2	6	0.33	−1.04	0.46	0.33	1.97	33	335	0.32	0.142	0.86	0.01	17
AFGL416	03:03:08.8	60:27:52.6	83	12	6	2.00	0.05	1.48	1.17	1.60	22	220	0.35	0.146	0.82	0.02	143
G173.51+2.79	05:40:43.5	35:55:02.3	10	5	2	2.50	0.57	0.29	0.31	0.88	45	150	0.37	0.116	0.91	0.13	24
G173.63+2.69	05:40:43.7	35:45:32.7	11	4	5	0.80	−0.35	0.35	0.35	1.01	25	72	0.23	0.133	0.77	0.03	11
S235C	05:40:52.0	35:38:18.1	14	1	0	0.00	0.00	0.23	0.20	1.41	114	327	0.23	0.077	0.91	0.20	44
S235AB	05:40:54.6	35:41:16.6	226	41	79	0.52	−0.44	1.92	1.27	2.28	58	950	0.23	0.086	0.70	0.02	373
G173.66+2.78	05:41:05.0	35:47:34.8	10	0	3	0.00	−1.15	0.56	0.39	2.02	18	40	0.28	0.187	0.86	0.07	12
S235	05:41:08.6	35:49:40.2	58	10	12	0.83	−0.62	0.77	0.69	1.24	45	330	0.28	0.108	0.85	0.06	104
G173.62+2.88 (E2)	05:41:24.5	35:52:23.1	33	9	14	0.64	−0.52	0.64	0.51	1.56	44	150	0.28	0.115	0.85	0.02	57
G173.67+2.87 (E1)	05:41:29.8	35:49:06.2	61	14	18	0.78	−0.45	1.11	0.74	2.24	40	444	0.28	0.116	0.71	0.05	83
G189.79+0.29	06:08:24.9	20:36:46.5	13	8	5	1.60	0.03	0.43	0.37	1.37	33	70	0.31	0.132	0.85	0.01	–
G189.84+0.29	06:08:31.0	20:34:23.4	14	9	3	3.00	0.18	0.78	0.63	1.51	15	72	0.41	0.195	0.78	0.01	21
G189.95+0.22	06:08:31.4	20:26:12.4	14	5	8	0.63	−0.83	0.80	0.50	2.61	16	37	0.41	0.184	0.76	0.01	24
S252A	06:08:37.5	20:38:13.0	174	54	52	1.04	−0.10	1.79	1.50	1.43	29	535	0.31	0.122	0.72	0.02	700
G189.94+0.33	06:08:52.5	20:29:56.3	48	18	22	0.82	−0.23	1.48	1.21	1.50	12	113	0.41	0.186	0.71	0.05	49
S252C	06:09:20.6	20:38:55.2	109	14	12	1.17	0.04	1.35	1.20	1.25	30	508	0.32	0.128	0.91	0.01	173
G189.95+0.54	06:09:41.8	20:35:32.5	11	4	5	0.80	−0.05	0.60	0.32	3.53	21	51	0.32	0.171	0.80	0.01	16
S252E	06:09:50.6	20:30:40.0	50	14	7	2.00	0.36	0.98	0.69	2.03	37	278	0.32	0.118	0.77	0.02	78
G192.69–0.25	06:12:19.6	17:48:31.5	7	3	4	0.75	−0.30	0.27	0.25	1.17	30	61	0.25	0.127	1.01	0.01	6
G192.54–0.15	06:12:24.3	17:59:29.4	8	2	0	0.00	1.61	0.49	0.80	0.37	126	334	1.18	0.056	1.10	0.01	19
S256	06:12:35.0	17:56:54.8	111	16	42	0.38	−0.88	1.49	0.92	2.64	46	530	0.21	0.103	0.64	0.02	154
G192.65–0.08	06:12:53.8	17:55:33.0	7	2	3	0.67	−0.37	0.19	0.34	0.33	44	87	0.21	0.099	0.95	0.02	11
S255-2 and S255N	06:12:55.3	17:59:20.1	142	21	13	1.62	0.26	1.15	0.97	1.42	57	664	0.21	0.093	0.81	0.02	251
G192.55−0.01	06:12:56.8	18:02:41.8	9	2	5	0.40	−0.77	0.25	0.30	0.69	35	61	0.21	0.122	0.97	0.01	16
G192.75−0.08	06:13:05.7	17:50:20.0	27	10	15	0.67	−0.51	0.47	0.40	1.35	66	538	0.25	0.089	0.80	0.03	50
G192.63+0.00	06:13:08.6	17:58:39.1	13	5	3	1.67	0.56	0.31	0.36	0.72	38	88	0.21	0.119	0.90	0.01	16
G192.70+0.03	06:13:22.4	17:56:26.9	8	2	3	0.67	−0.10	0.29	0.26	1.23	44	129	0.22	0.133	1.10	0.01	8
G192.75+0.00	06:13:24.8	17:52:51.1	7	0	7	0.00	−0.91	0.19	0.15	1.71	93	177	0.22	0.090	1.11	0.01	7
S258	06:13:28.9	17:55:47.2	43	7	5	1.40	−0.13	0.52	0.44	1.42	80	646	0.22	0.088	0.88	0.01	83
G111.44+0.79	23:12:56.5	61:26:56.1	14	3	10	0.30	−0.53	0.83	0.62	1.80	12	41	0.43	0.221	0.83	0.17	28
G111.44+0.75	23:13:02.5	61:24:15.4	11	0	11	0.00	−1.05	0.68	0.50	1.83	10	24	0.43	0.232	0.89	0.23	28
G111.48+0.80	23:13:14.4	61:28:00.4	11	3	3	1.00	−0.59	0.47	0.40	1.41	23	71	0.43	0.177	1.01	0.16	73
G111.49+0.81	23:13:16.7	61:28:57.4	15	3	2	1.50	0.76	0.57	0.42	1.88	27	64	0.43	0.157	0.94	0.09	100
G111.47+0.75	23:13:20.1	61:25:30.6	18	7	11	0.64	−0.17	0.54	0.48	1.29	32	131	0.43	0.153	1.03	0.37	53
NGC 7538	23:13:45.9	61:28:12.9	192	36	22	1.64	0.57	2.65	1.88	1.99	21	166	0.32	0.146	0.64	0.08	1227
G111.57+0.81	23:13:56.3	61:30:39.1	90	8	9	0.89	−0.23	1.27	1.18	1.16	24	192	0.32	0.149	0.82	0.10	321
G111.61+0.77	23:14:21.7	61:29:12.2	13	3	3	1.00	−0.06	0.67	0.55	1.47	12	28	0.32	0.209	0.87	0.15	27
G111.59+0.64	23:14:26.5	61:22:23.2	28	4	24	0.17	−1.11	2.77	1.67	2.73	4	135	0.76	0.411	0.70	0.03	–
G111.65+0.63	23:15:09.6	61:22:06.7	19	4	15	0.27	−0.77	1.13	1.06	1.14	6	49	0.76	0.304	0.77	0.01	–

[a] See a detailed description of the columns at the end of Section 4.2.1.

that they stay coupled to the gas for a wide range of densities (Vlemmings 2008). Troland & Crutcher (2008) showed that the ratio between turbulent and magnetic energy in dark clouds cores is roughly a factor of 2. Vlemmings (2008) showed that the energy due to turbulence and magnetic fields may be also of the same order for more evolved and massive cores. More detailed studies on the density structure and magnetic field measurements at similar scales are needed to disentangle which force dominates the support against gravitation.

5.3 N_{YSO} versus dense cloud mass

Lada, Lombardi & Alves (2010) found that the number of YSOs in a cluster is directly proportional to the dense cloud mass ($M_{0.8}$), which is the mass above a column density equivalent to $A_K \sim 0.8$ mag. This has important implications since establishes an empirical relation between the young stellar content and its parent molecular cloud.

To investigate this correlation among our cluster sample, we calculate $M_{0.8}$ by integrating the gas mass over the region inside the cluster convex hull area where $A_K > 0.8$ mag. For the number of YSOs, we use the total number of cluster members N_{YSO}. If well several cluster members are located in areas where $A_K < 0.8$, we assume that some of the young stars will move away from their birthplace (Lada et al. 2010, used all YSOs for which $A_K > 0.1$ mag and showed that their correlation does not depend on the cutoff magnitude for the accounted members). We find that the number of YSOs is proportional to ${M_{0.8}}^{\alpha}$ with an index $\alpha = 0.6$. This is different from the index $\alpha = 1$ obtained by Lada et al. (2010). However, since our clusters are located further than the sample from Lada et al. (2010), our results may be biased by the non-detection of faint cluster members. To correct this, we estimate the percentage of non-detected YSOs in each cluster (for a detailed explanation see Appendix C), and corrected the number of cluster members. After this, we find a correlation between the dense gas mass and the corrected number of cluster members (N^*_{YSO}) is given by (see Fig. 26)

$$N^*_{YSO} \propto {M_{0.8}}^{0.89\pm0.15}, \quad (11)$$

in better agreement with Lada et al. (2010). The linear relation between the dense cloud mass and the number of stars suggest that

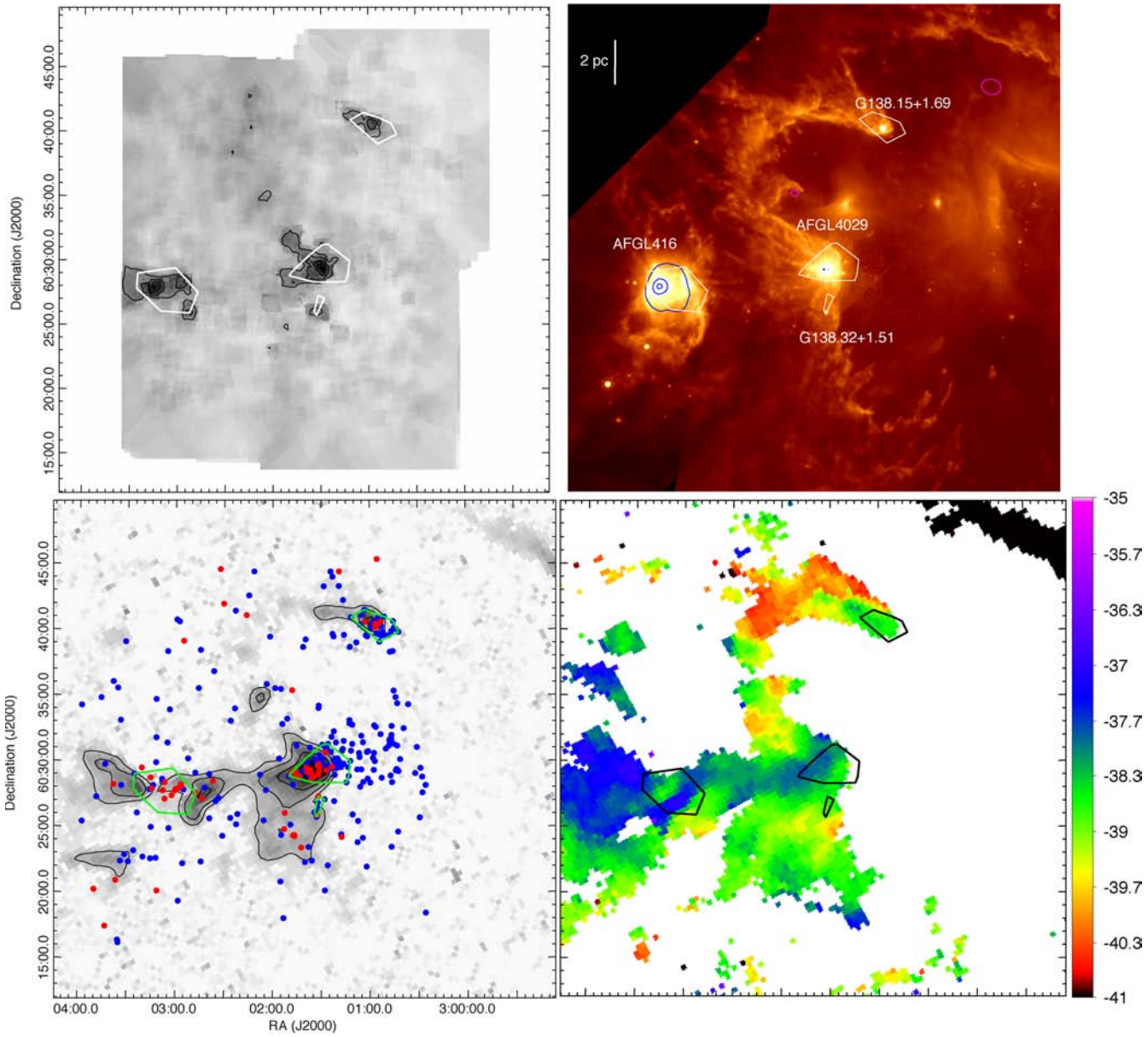

Figure 12. Molecular material distribution in region W5-east. Clusters convex hull area are shown as white, green or black polygons. Upper left: *K*-band extinction map. Black contours begin at $A_K = 0.4$ (mag) and increase by 0.2. Upper right: MIPS map at 24 μm. Blue and magenta contours correspond to the 1.4 GHz radio emission from VLA observations. Contours are at 10, 50 and 90 per cent of peak intensity. Blue contours correspond to emission associated with W5-east. Magenta contours show likely background emission. Cluster names are labelled in this panel. Lower left: ^{13}CO column density map. Black contours are spaced logarithmically between 0.6 and 4×10^{16} cm^{-2}. Red dots correspond to Class I sources, blue dots correspond to Class II. Lower right: ^{13}CO first momentum map. The bar at the left shows the colour-coded $V_{\rm LSR}$ in km s^{-1}.

the SFR depends also linearly from the dense mass (Lada et al. 2010).

5.4 Star formation efficiency

The star formation efficiency (SFE), is defined as the percentage of gas and dust mass converted into stars:

$$\mathrm{SFE} = \frac{M_{\rm stars}}{M_{g+d} + M_{\rm stars}}. \quad (12)$$

In many cases, the SFR is calculated over regions which are still forming stars. This means that there is still some molecular mass (M_{g+d}) that will be converted into star mass ($M_{\rm stars}$). Since this is the case in the studied regions, the SFE will be then considered as a lower limit.

We use the cluster properties to estimate the SFE on a cluster size-scale. In this case, most of the low-density gas that surrounds the clusters is not taken into account since we calculate the cloud mass only from inside the cluster convex hull area. We use $M_{g+d} = \max(M_{A_K}, M_{\sigma_{\rm CO}})$, $M_{\rm stars} = 0.5 \times N^*_{\rm YSO}$ and find SFEs between 3 and 45 per cent with an average of ∼20 per cent. Those values are in agreement with the efficiencies needed to go from the core mass function to the initial mass function (IMF) (e.g. 30 per cent in the Pipe nebula and 40 per cent in Aquila, from Alves, Lada & Lombardi 2007; André et al. 2010, respectively). Also, similar values are obtained in cluster formation simulations

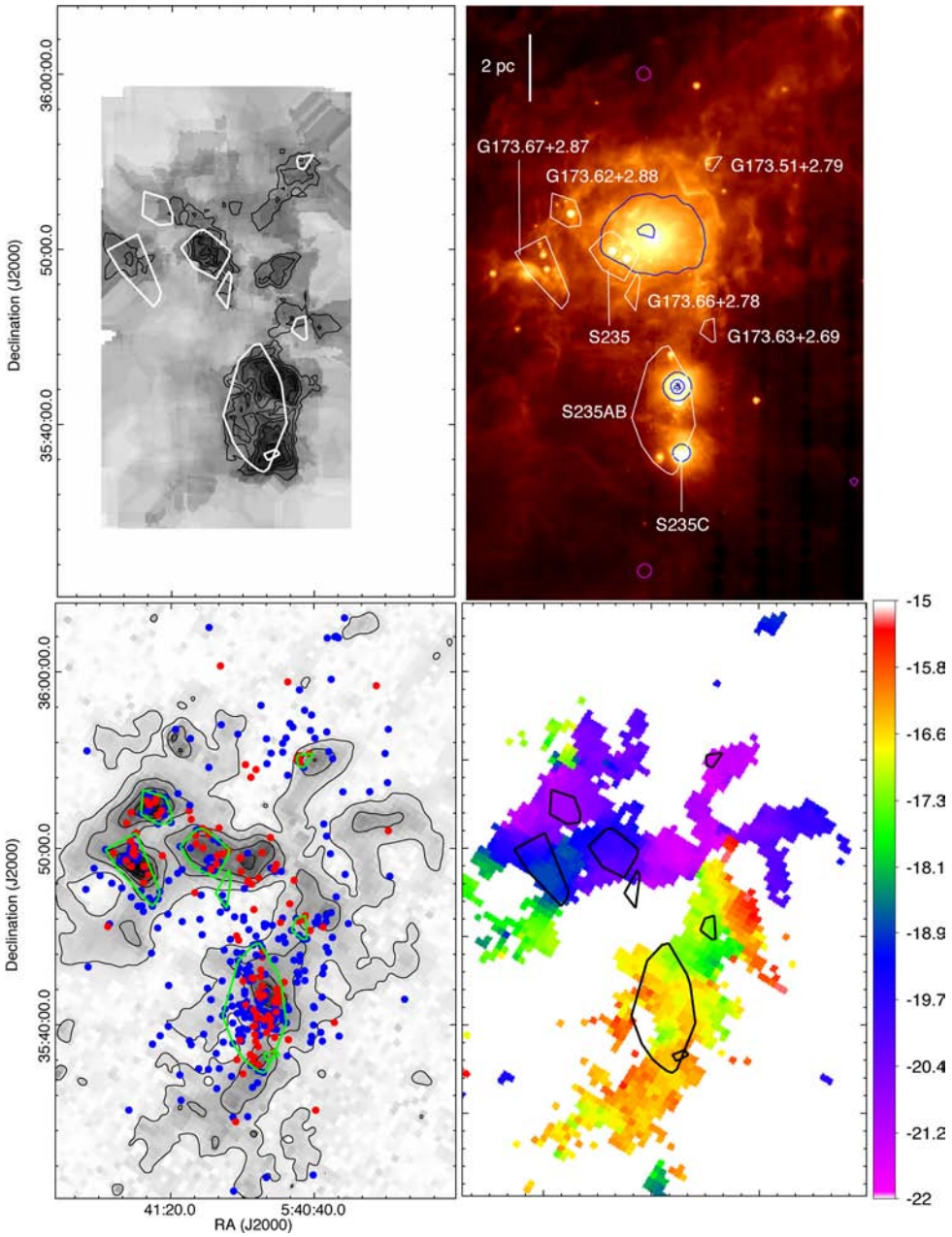

Figure 13. Same as Fig. 12 for region S235. Contours in the *K*-band extinction map (upper right) begin at $A_K = 0.8$ (mag) and increase by 0.2. ^{13}CO column density contours (lower left) are logarithmically spaced between 0.6 and 4.2 $\times$ 10^{16} cm^{-2}.

from Maschberger et al. (2010) for a 10^3 M$_\odot$ cloud. The SFE distribution over the corrected number of cluster members is shown in Fig. 27. There is no correlation between the SFEs and number of cluster members. This suggest that feedback processes, which may be particularly important in MEC, may start having an impact only in the later stages of the cluster evolution.

On a bigger spatial scale, we estimate the SFE for the whole regions, which is a better estimate of the SFE in molecular clouds. In this case, we calculate the cloud mass inside the field covered by all the IRAC and NIR bands using our ^{13}CO density maps since those are more sensitive to the less dense gas than the extinction maps. For the number of stars, we use $N_{\rm YSO}$ from Table 3 corrected by a factor of 2 to roughly include the undetected YSOs. We find that the SFE in each region is between 4 and 8 per cent.

5.5 The initial mass function (IMF) high-mass end

Once formed, the evolution of stars is determined by their mass. This is why a complete understanding of the IMF is fundamental to understand the evolution of star clusters as well as to constrain star formation theories. The IMF is defined as a power law of the form:

$$\frac{{\rm d}N}{{\rm d}\log m} = m^{\Gamma}, \tag{13}$$

where N is the number of stars of mass m and Γ is the mass function index or slope.

We use synthetic cluster models to estimate the high-mass index of the IMF ($\Gamma_{\rm HM}$) for the most numerous cluster in each region: AFGL4029, S235ABC, S252A, S255-2 and NGC 7538. To derive $\Gamma_{\rm HM}$, we compare the observed cluster *K*-band luminosity function

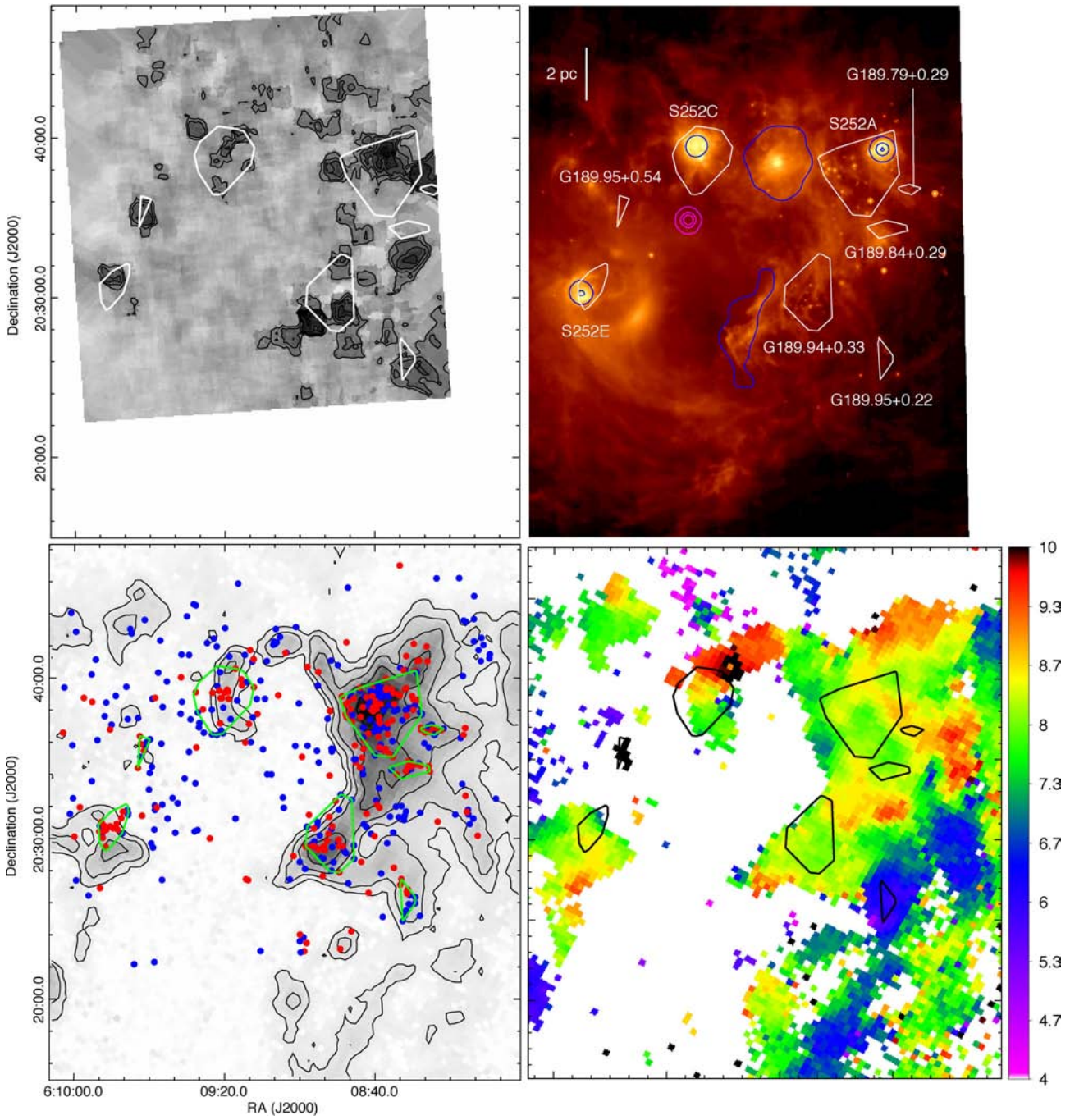

Figure 14. Same as Fig. 12 for region S252. Contours in the *K*-band extinction map (upper right) begin at $A_K = 0.6$ (mag) and increase by 0.2. ^{13}CO column density contours (lower left) are logarithmically spaced between 0.6 and 6×10^{16} cm^{-2}.

(KLF, derived from all the detections in the *K*-band, see Section D) with the KLF drawn from synthetic clusters having as variables the IMF slope and the range of ages. This way we constrain both the cluster age and the IMF index. Clusters IMF is in general described by several Γ slopes for different mass ranges. Muench et al. (2002) found that the IMF for the Orion-Trapezium cluster is fitted by five slopes: $\Gamma_1 = -1.21$ for $m(\mathrm{M}_\odot) > 0.6$, $\Gamma_2 = -0.15$ for $0.6 > m(\mathrm{M}_\odot) > 0.12$, $\Gamma_3 = 0.73$ for $0.12 > m(\mathrm{M}_\odot) > 0.025$, $\Gamma_4 = -0.5$ for $0.025 > m(\mathrm{M}_\odot) > 0.027$ and $\Gamma_5 = 0$ for $m(\mathrm{M}_\odot) < 0.017$. Since our observations detect sources of more than around 0.5–1 $\mathrm{M}_\odot$, we vary the IMF slope only in the range $m > 0.6$ $\mathrm{M}_\odot$ and assume a Trapezium IMF for lower masses. Synthetic clusters are created using the Monte Carlo algorithm from Muench, Lada & Lada (2000), which creates 100 000 stars with two main parameters as input: the IMF and age range. From those stars, we randomly select 2000 to create a synthetic cluster. This cluster is later convolved with the observed clusters individual properties like extinction and IR-excess (for details see Appendix D). Finally, we use a least-squares fitting algorithm to compare the observed and synthetic KLF for different $\Gamma_{\mathrm{HM}}$and age range values (see Fig. 28).

We find that the five clusters have luminosity functions better fitted by Γ_{HM}values between -1.00 and -1.60 and cluster ages

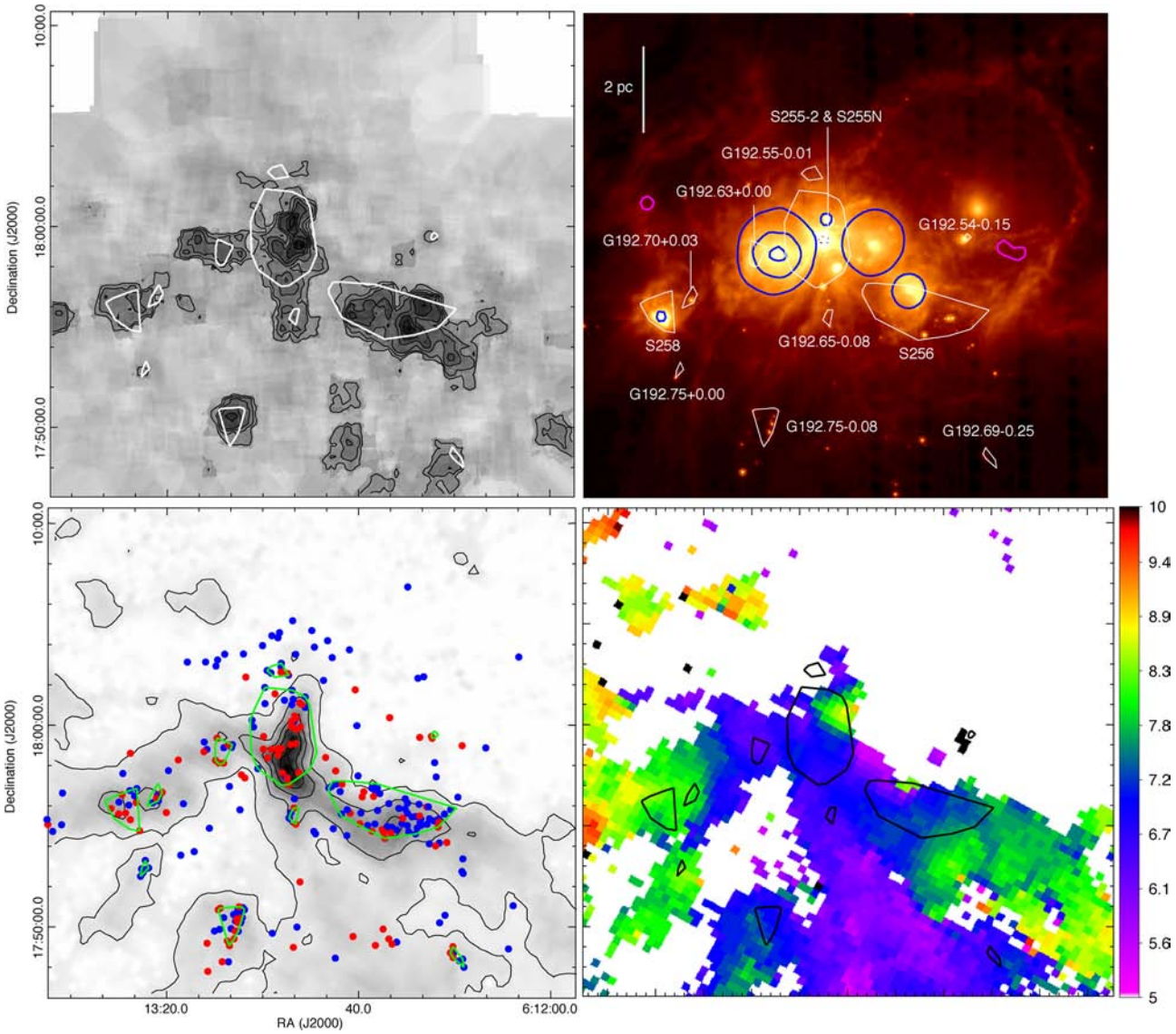

Figure 15. Same as Fig. 12 for region S254-S258. Contours in the *K*-band extinction map (upper right) begin at $A_K = 1.0$ (mag) and increase by 0.2. ^{13}CO column density contours (lower left) are logarithmically spaced between 0.6 and 6×10^{16} cm^{-2}.

of less than ∼1.5 Myr (see Table 7). Clusters AFGL4029 and S235ABC have a top-heavy IMF. This may be of particular interest for AFGL4029 since this cluster is a good candidate for triggered star formation and some numerical models propose that triggering stars may trigger the formation of a more massive second generation of stars with a top-heavy IMF (Hosokawa & Inutsuka 2006). On the other hand, cluster S255-2, which is also a candidate for triggered star formation, has a steeper IMF. We conclude that if well the KLF is a useful method to estimate the IMF, it should be complemented with other observations like stars spectra to better constrain the cluster members ages and masses.

5.6 About triggered star formation

We searched for trends among the physical properties of triggered cluster candidates and find no clear differences between them and clusters formed by 'simple' gravitational collapse. Unless there is a clear age sequence like in cluster AFGL4029, it is difficult to estimate whether star formation around H II regions has been triggered or if it started before the arrival of the expansion wave. Even if this is the case, our observations show that once the clump becomes gravitationally unstable, the fragmentation processes and cluster dynamics rapidly erase the imprints of the triggering mechanism. This has already been suggested by the simulations from Dale, Ercolano & Bonnell (2013). Finally, one of the key parameters missing in current studies about triggered star formation is the investigation of time causality between the expansion of the H II region and the age of the newly formed stars.

5.7 Properties of individual embedded clusters

In this section, we summarize the most important characteristics of the identified embedded clusters. We also search for individual sources of interest (not necessary inside clusters). The properties of clusters in region S254-S258 are explained in detail in Chavarría et al. (2008). All given coordinates are in the J2000 system.

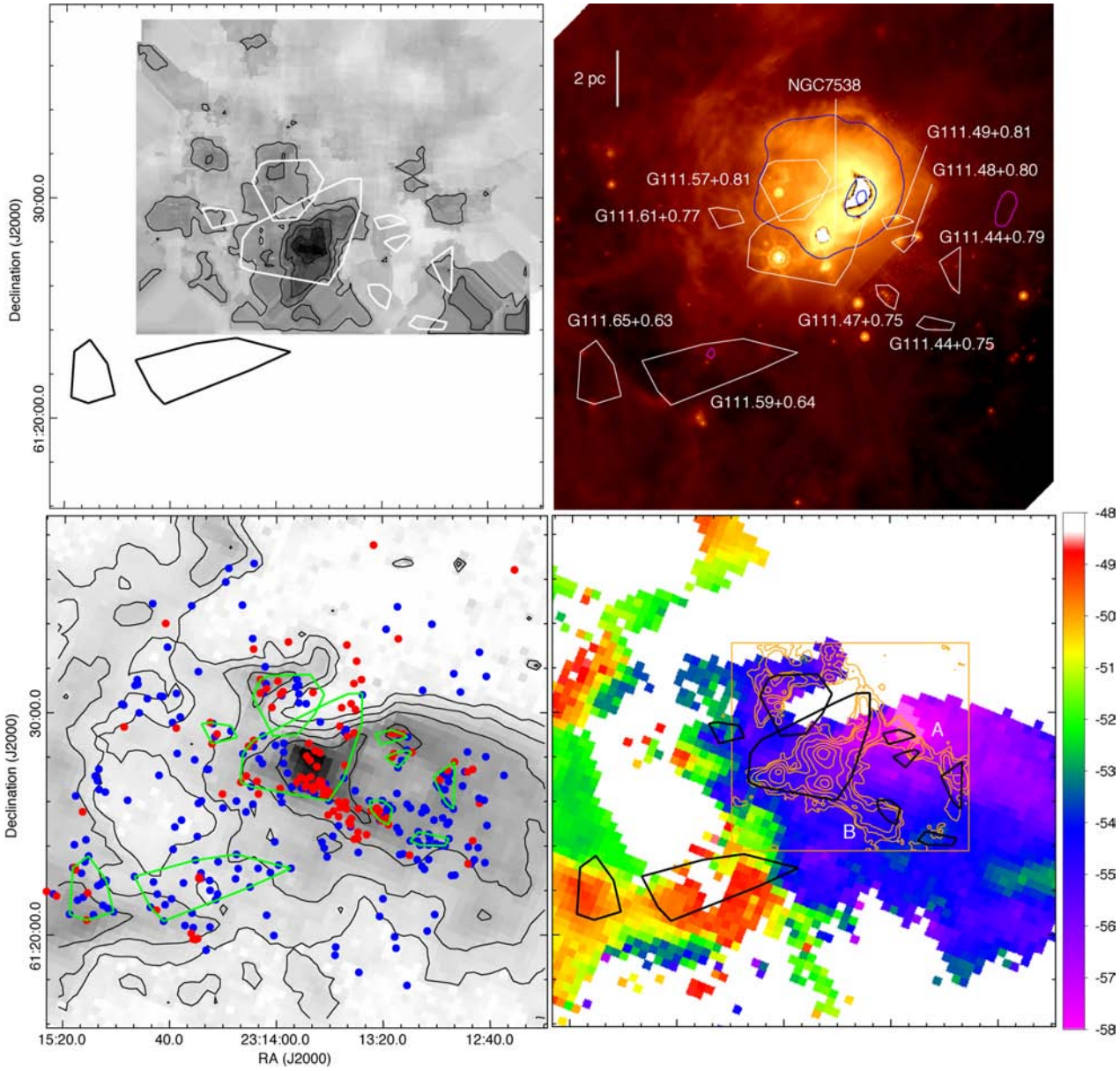

Figure 16. Same as Fig. 12 for region NGC 7538. The contours in the *K*-band extinction map (upper right) begin at $A_K = 0.6$ (mag) and increase by 0.2. ^{13}CO column density contours (lower left) are logarithmically spaced between 0.6 and 6×10^{16} cm^{-2}. The ^{13}CO first momentum map (lower right) includes the SCUBA 850 μm emission in orange contours (spaced logarithmically between 0.063 and 18 Jy). The orange box shows the SCUBA observed FOV. Filaments A and B are also labelled (see NGC 7538 in Section 5.7.5).

5.7.1 W5-east

5.7.1.1 G138.15+1.69. This cluster is a good candidate for triggered star formation since it is located at the extreme of a pillar-like structure pointing in the direction of the ionizing star HD 18326 (Deharveng et al. 2012). Our dense-gas mass estimation ($M_{0.8}$) is a factor of 2 higher than previous estimations from Submillimetre Common-User Bolometer Array (SCUBA) data enclosing a similar area (Morgan et al. 2008). The estimated $V_{\rm LSR}$ is in agreement with previous ^{13}CO observations from Niwa et al. (2009).

Ginsburg, Bally & Williams (2011) reported several outflows in the W5 region using CO 3–2 observations. We searched for their powering sources and find one IR-excess source located between the outflow #25 red and blue lobes which is a good candidate (coordinates $3^{\rm h}00^{\rm m}56\overset{\rm s}{.}16$ $+60^\circ40'34\overset{''}{.}5$). However, this source is not identified as a *Herschel* point source by Deharveng et al. (2012) neither as a T-Tauri star by Ogura et al. (2002).

Deharveng et al. (2012) identified several YSOs towards G138.15+1.69. Of the five point-like sources detected in their study, four are classified as Class I following our classification scheme. Their fifth source is located inside the cluster convex hull area but is not detected in our observations.

We also identified a Class I source located at the top of a pillar-like structure between clusters G138.15+1.69 and AFGL4029, with coordinates: $3^{\rm h}01^{\rm m}47\overset{\rm s}{.}04$ $+60^\circ35'23\overset{''}{.}9$.

5.7.1.2 AFGL4029. The spatial distribution of Class I and Class II sources around AFGL4029 suggest a sequential star formation scenario. Figs 7 and 12 show that Class I sources are located preferentially east from the PAH ridge and are associated with the ^{13}CO

Table 5. Properties of molecular material associated with embedded clusters.

Name	$\overline{A}_K$ (mag)	A_K (mag)	M_{A_K} ($M_\odot$)	$M_{0.8}$ ($M_\odot$)	$\overline{\sigma}_{CO}$ (10^{16} cm^{-2})	σ_{CO} (10^{16} cm^{-2})	$M_{\sigma_{CO}}$ ($M_\odot$)	V_{LSR} (km s^{-1})	$^{13}CO_{FWHM}$ (km s^{-1})
G138.15+1.69	0.71	2.33	209	46	0.95	2.24	131	−38.5	1.65
AFGL4029	0.83	2.61	477	143	1.60	3.75	435	−38.3	1.81
G138.32+1.51	0.52	1.38	19	2	0.62	1.06	11	−38.7	1.74
AFGL416	0.88	2.24	665	155	0.68	1.78	242	−37.7	2.26
G173.51+2.79	1.33	1.51	32	13	1.68	2.29	19	−21.4	1.39
G173.63+2.69	1.01	1.69	42	13	1.53	1.97	30	−17.4	1.06
S235C	1.99	2.16	31	18	2.13	2.53	15	−16.2	1.72
S235AB	1.43	2.37	1305	580	1.79	4.03	770	−16.7	1.76
G173.66+2.78	1.15	1.67	54	17	0.72	2.07	16	−19.7	4.14
S235	1.54	3.40	375	192	2.60	3.46	297	−19.8	1.74
G173.62+2.88 (E2)	0.75	1.38	94	17	3.29	4.89	194	−20.7	1.36
G173.67+2.87 (E1)	0.98	1.63	294	85	2.77	5.12	391	−18.8	1.69
G189.79+0.29	–	–	–	–	2.80	3.26	53	10.0	1.93
G189.84+0.29	0.51	1.22	63	10	3.29	4.03	191	8.8	2.82
G189.95+0.22	1.25	1.79	122	46	1.56	1.84	71	5.8	1.88
S252A	0.87	2.15	1121	318	4.60	7.34	2788	8.3	3.57
G189.94+0.33	0.94	2.37	623	222	2.49	4.82	778	8.7	2.16
S252C	0.66	2.39	540	83	0.89	2.37	338	8.7	2.94
G189.95+0.54	1.00	1.64	47	17	0.43	0.61	10	9.5	1.27
S252E	0.78	1.86	183	36	1.01	2.45	112	8.5	1.72
G192.69−0.25	1.16	1.62	13	4	1.03	1.12	5	6.7	2.44
G192.54−0.15	0.88	0.97	3	1	0.34	0.34	1	24.1	1.67
S256	1.35	2.48	626	269	2.00	3.56	435	7.3	2.16
G192.65−0.08	1.59	1.74	16	8	2.28	2.39	11	6.5	1.90
S255-2 and S255N	1.24	3.08	640	264	2.79	7.02	677	6.6	2.14
G192.55−0.01	0.68	0.85	13	1	0.19	0.40	2	5.9	2.19
G192.75−0.08	1.57	2.10	97	49	1.47	1.87	42	7.1	1.97
G192.63+0.00	1.43	1.61	44	19	1.73	1.95	25	6.5	1.57
G192.70+0.03	0.83	1.71	13	1	1.96	2.11	14	8.0	1.39
G192.75+0.00	0.91	1.92	5	1	0.81	0.92	2	8.2	1.81
S258	1.14	1.92	108	35	1.82	2.50	81	8.2	1.72
G111.44+0.79	1.12	1.72	150	46	2.83	3.16	177	−57.8	5.05
G111.44+0.75	0.87	1.29	73	27	2.34	2.77	92	−54.5	6.53
G111.48+0.80	0.83	1.06	43	4	2.59	2.98	63	−57.6	5.03
G111.49+0.81	0.18	1.06	12	1	3.42	3.96	102	−58.9	3.92
G111.47+0.75	0.83	0.89	71	6	2.87	3.08	115	−53.2	2.14
NGC 7538	1.27	2.97	2512	1211	2.82	6.82	2616	−56.4	4.00
G111.57+0.81	0.76	1.45	577	90	1.20	3.10	430	−55.5	4.16
G111.61+0.77	0.84	1.49	94	25	0.91	1.17	48	−52.8	2.23
G111.59+0.64	–	–	–	–	1.61	0.89	615	−48.8	2.02
G111.65+0.63	–	–	–	–	1.90	2.77	535	−50.5	3.60

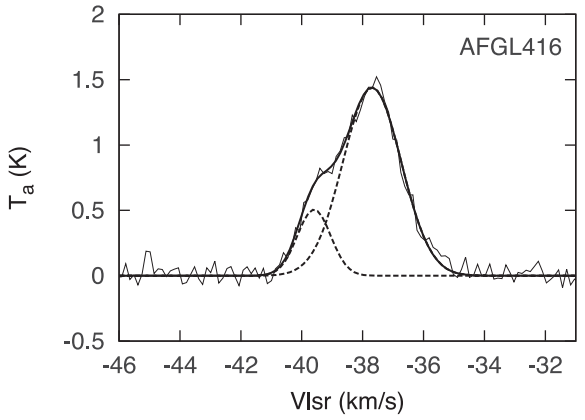

Figure 17. Integrated ^{13}CO spectra over the convex hull area for cluster AFGL416. The spectra is fitted (continuous black line) by combining Gaussian components (dashed lines). We assume that AFGL416 is associated with the component at $V_{LSR} = -37.7$ km s^{-1}.

column density peak (most dense material), while Class II are located mainly to the west of the PAH front, closer to the ionizing star HD18326 and where the molecular material has already been removed. Assuming that Class I sources are younger than Class II, the observed star formation sequence makes this region a good candidate for triggered star formation; the ionizing front compresses and accumulates the molecular material as it expands, until exceeds its critical mass and collapses forming stars. Such a process will have the youngest stars located close to the expansion front, immersed in a pillar-like structure and preceded by older stars closer to the ionizing star. This is what we observe in AFGL4029 and in a minor scale in G138.15+1.69. However, the question if the expansion front compressed an already existing clump is still open.

We hypothesize that the Class II sources located up to ∼4 arcmin west from the cluster represent a previous period of star formation and are dynamically more evolved. However, this does not rule out the possibility that those sources are also part of cluster AFGL4029. Since they are outnumbered by the newly formed stars, the critical distance d_c around the cluster is overweighted by the most embedded

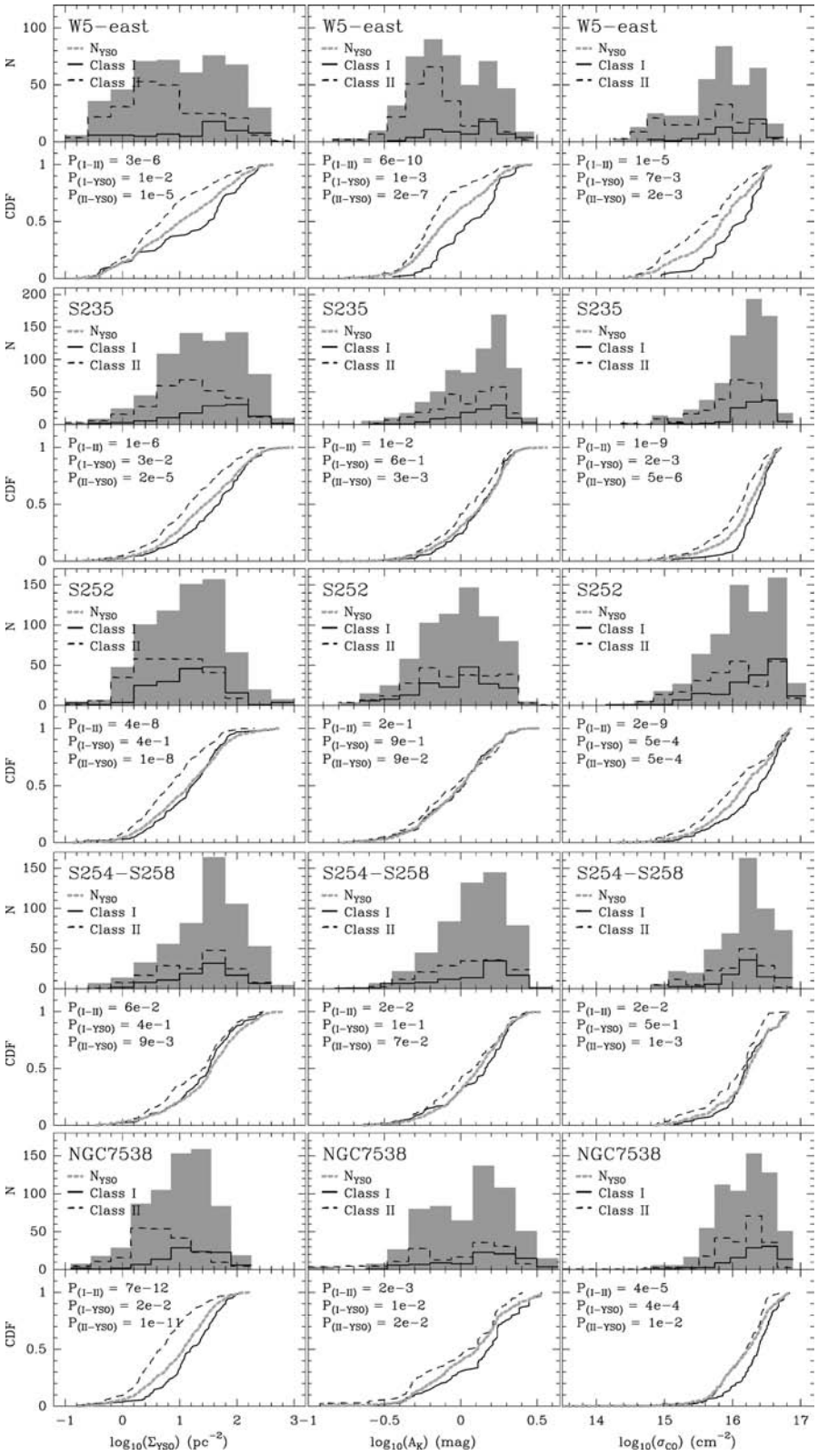

Figure 18. YSOs distribution over molecular material per region. Each panel shows the distribution histogram and cumulative distribution function (CDF) of all detected YSOs ($N_{\rm YSO}$, grey histogram and grey dashed line), Class I (continuous black line) and Class II (dashed black line) according to the surface density of YSOs (left), *K*-band extinction (centre) and ^{13}CO surface density (right). K–S similarity probabilities are shown for two data sets comparison.

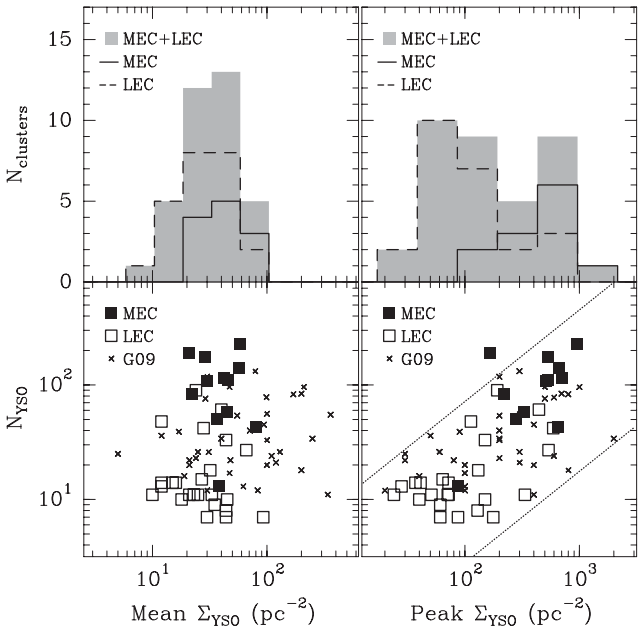

Figure 19. Clusters mean (left) and peak (right) YSOs surface density. The histograms (upper charts) show the $\Sigma_{\rm YSO}$ distribution for all clusters (grey), MEC (continuous line) and LEC (segmented line). The YSOs surface densities versus the number of cluster members are shown in the lower charts. The crosses show the values obtained by Gutermuth et al. (2009) in LEC. Dotted lines in the lower-right chart enclose most clusters (except cluster NGC 7538) and show a weak correlation with index 0.8 ± 0.2 and correlation index $r = 0.75$.

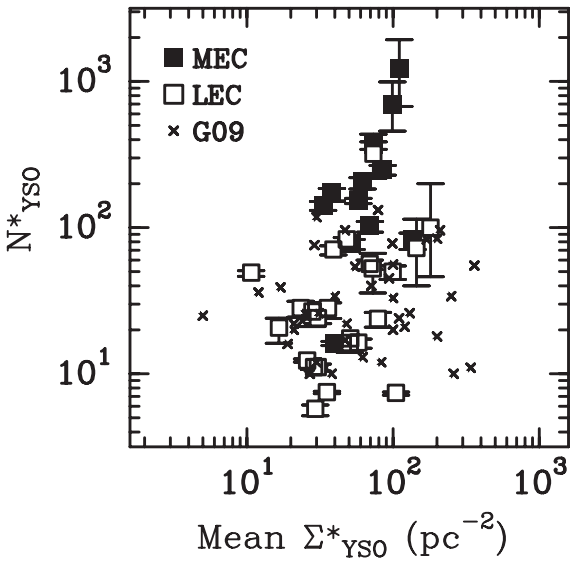

Figure 20. Mean YSOs surface density per number of members derived from the corrected number of cluster members $N^*_{\rm YSO}$. Crosses show the mean surface densities from Gutermuth et al. (2009).

sources. Consequently, the western Class II sources will not be consider part of the cluster with our current cluster definition.

The dense-gas mass estimated for AFGL4029 is a factor of 3 higher than previous calculations from Morgan et al. (2008). However, our estimations involve a larger area. As for cluster G138.15+1.69, we found a Class I source ($3^{\rm h}01^{\rm m}31\overset{\rm s}{.}30$ $+60^\circ29'13\overset{''}{.}0$) associated with the outflow #26 from

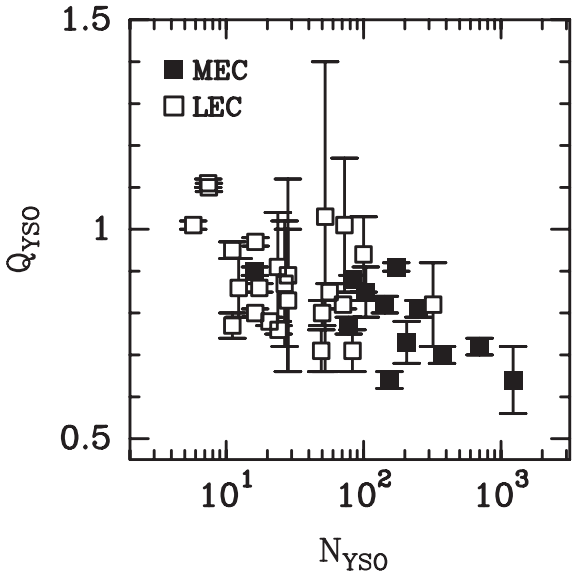

Figure 21. Structural $\mathcal{Q}$ parameter ($\mathcal{Q}_{\rm YSO}$) for cluster members. Black squares correspond to MEC while empty squares correspond to LEC.

Table 6. Structural $\mathcal{Q}$ parameter for different YSOs classes in selected clusters.

Cluster name	$\mathcal{Q}_{\rm I}$	$\mathcal{Q}_{\rm II}$
AFGL4029	0.79	0.74
S235AB	0.67	0.74
S252A	0.75	0.79
S255-2 and S255N	0.76	0.94
NGC 7538	0.68	0.76

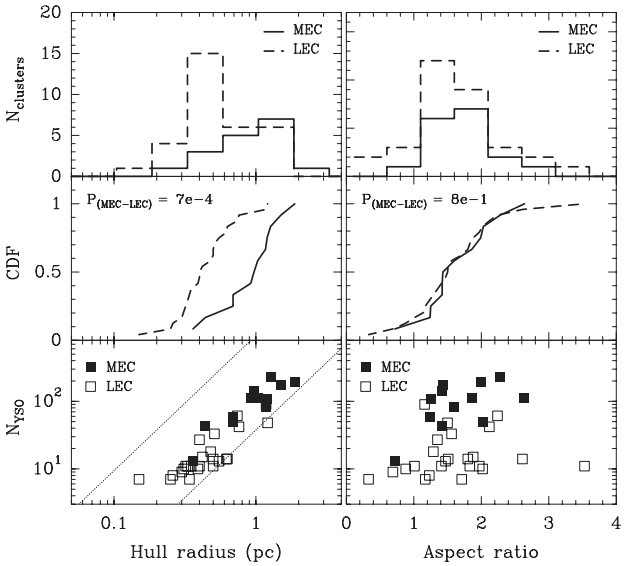

Figure 22. Morphology of embedded clusters. Left: hull radius distribution. The upper chart shows the $R_{\rm H}$ distribution for MEC and LEC. The middle chart shows the $R_{\rm H}$ cumulative distribution function for MEC (continuous line) and LEC (dashed line). The K–S similarity test value is also shown. The lower chart shows the $R_{\rm H}$ versus the number of cluster members. The dot–dashed lines shows constant surface densities at 12 and 300 $\rm pc^{-2}$. Those correspond to the range spanned by the embedded clusters from Gutermuth et al. (2009). Right: the same for the aspect ratio AR distribution.

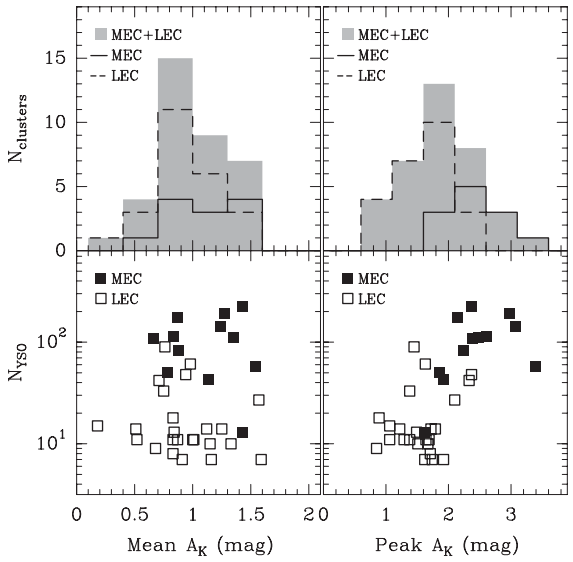

Figure 23. Embedded clusters *K*-band extinction. Left: the upper charts shows the mean A_K for LEC and MEC. The lower chart show the mean A_K versus the number of cluster members. Right: the same for the peak A_K values.

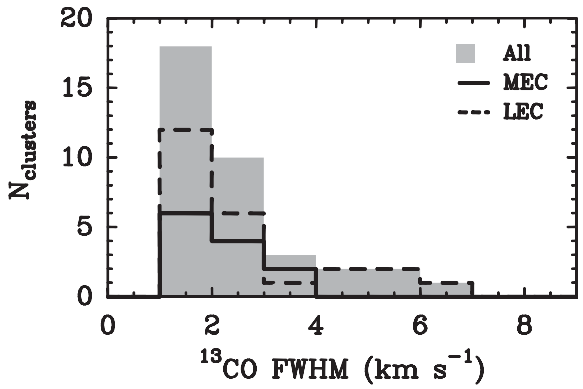

Figure 24. Distribution of clusters ^{13}CO linewidth (FWHM). The thermal broadening for a 30 K cloud is $\sim$0.3 km s^{-1}.

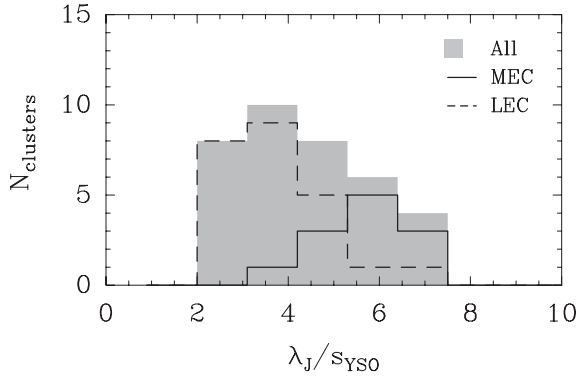

Figure 25. Ratio between the cluster Jeans length (λ_J) and the mean projected distance between cluster members (s_{YSO}). Massive embedded clusters are shown by the continuous line. Low-mass EC are shown by the segmented line.

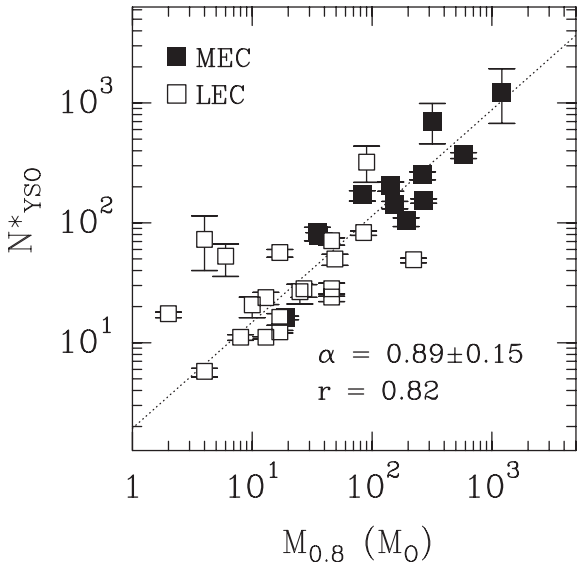

Figure 26. Corrected number of cluster members and molecular mass above $A_K = 0.8$ mag ($M_{0.8}$). The dotted line shows the best fit to the data using orthogonal regression. The error-bars are derived from the N^*_{YSO} values obtained with Γ_{HM} in the range -1.21 ± 0.5 (see Appendix C).

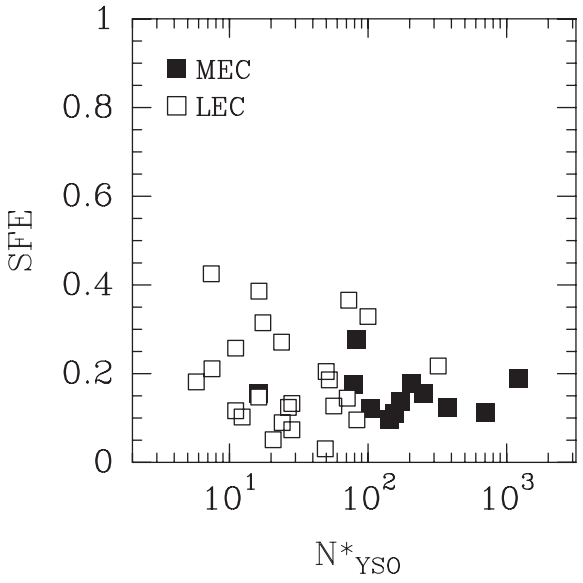

Figure 27. Star formation efficiency (SFE) in embedded clusters. The chart shows the SFE for the corrected number of cluster members (N^*_{YSO}).

Ginsburg et al. (2011). This source is also identified as the outflow counterpart by Deharveng et al. (2012), corresponding to their *Herschel* point source BRC14-c1.

Deharveng et al. (2012) identified several YSOs inside AFGL4029 convex hull area. Of the 13 point-like sources detected in their study, 4 are classified as Class II, 4 as Class I, 1 as YSO and 3 are not detected in our observations.

5.7.1.3 G138.32+1.51. This cluster is located $\sim$3 arcmin south of AFGL4029 and it is associated with continuum emission at 850 μm (MacKenzie et al. 2011). We found a Class I source ($3^h01^m47\overset{s}{.}44$ $+60^\circ24'20\overset{''}{.}4$) located south-east of G138.32+1.51 which is associated with the outflow #33 from Ginsburg et al. (2011). This source is also identified as the outflow counterpart

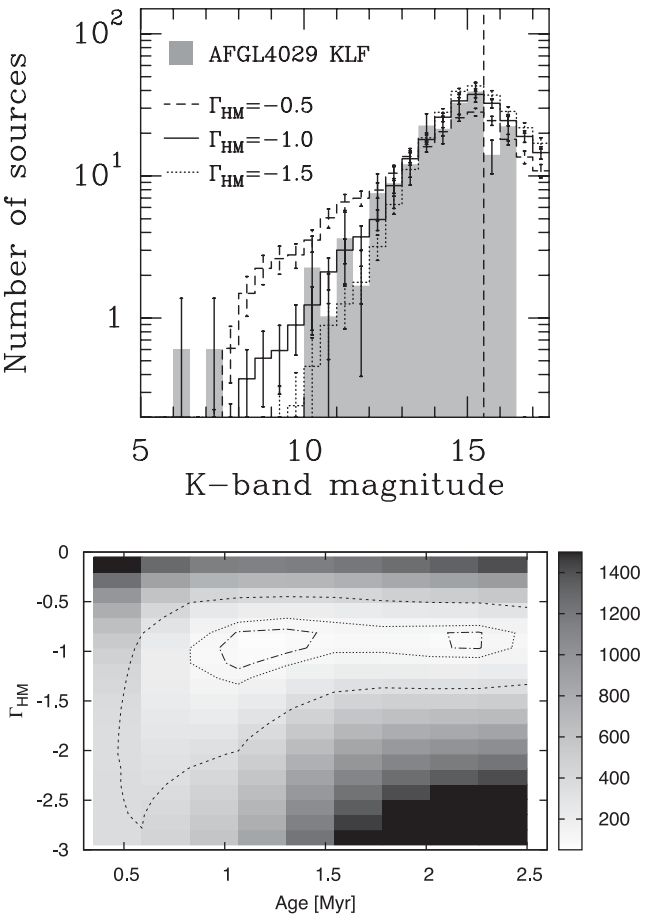

Figure 28. Upper: field-substracted *K*-band luminosity function for cluster AFGL4029 (in grey) and for synthetic clusters (lines); the continuous line shows the synthetic cluster that better fits our data. All models have ages between 0.5 and 1.5 Myr. Lower: $\Gamma_{HM}$and cluster-age parameter space for cluster AFGL4029. The contours are at 1, 3 and 5σ. The χ^2 values are shown in the grey-scale.

Table 7. Best-fitted $\Gamma_{HM}$and ages for selected clusters.

Cluster name	Γ_{HM}	Age (Myr)
AFGL4029	-1.0 ± 0.3	1.2 ± 0.5
S235ABC	-1.0 ± 0.4	1.0 ± 0.5
S252A	-1.2 ± 0.5	0.6 ± 0.5
S255-2 and S255N	-1.6 ± 0.3	1.3 ± 0.6
NGC 7538	-1.3 ± 0.5	0.4 ± 0.3

by Deharveng et al. (2012), corresponding to their *Herschel* point source BRC14-a3.

Deharveng et al. (2012) identified one YSO inside the cluster convex hull that we classify as Class I source.

5.7.1.4 AFGL416. The bipolar PAH structure observed in this cluster is due to the interaction between the stellar winds from massive stars and the ISM (Felli, Hjellming & Cesaroni 1987). Fig. 1 shows that the radio emission peaks between the lobes. On the other hand, Fig. 12 shows that the ^{13}CO column density between the lobes increases from the centre to the east and west sides. All of this implies that massive stars in AFGL416 are located along the east–west filament. Those stars generate strong stellar winds that sweep out the molecular material more efficiently along the north–south direction due to the lower column density, creating the observed bipolar morphology.

Our observations suggests that the interaction between molecular material associated with cluster AFGL416 and the ionizing front from star HD18326 is weak.

Ginsburg et al. (2011) found two outflows east of AFGL416 for which we have not detected powering source candidates.

Deharveng et al. (2012) identified several YSOs inside the cluster convex hull area. Of the five point-like sources detected in their study, two are classified as Class I, one as YSO and two are not detected in our observations.

5.7.2 S235

5.7.2.1 G173.51+2.79. This cluster is located north-west from the H II region S235. It is also called S235 North-West (Kirsanova et al. 2008), FSR 784 (Froebich, Scholz & Raftery 2007) and Koposov 7 (Koposov, Glushkova & Zolotukhin 2008). Dewangan & Anandarao (2011) found 3 YSOs in this cluster, all of them are also identified as YSOs in our observations.

Cluster G173.51+2.79, as well as clusters G173.62+2.88 and G173.67+2.87, are located in the outer parts of the H II region S235 and have been proposed as triggered star formation candidates. However, unlike clusters AFGL4029 and G138.15+1.69, these do not seem to be located in pillar-like structures. Kirsanova et al. (2008) observed ^{13}CO and CS emission towards the region S235 and found good agreement between the distribution of molecular material associated with these three clusters and what is expected in a triggered or induced star formation scenario.

5.7.2.2 G173.63+2.69. Cluster located south-west of the H II region S235 and together with S235AB and S235C, is associated with molecular material at a different V_{LSR} than the rest of the clusters surrounding the H II region S235 (see Table 5).

5.7.2.3 S235C. This cluster is probably part of cluster S235AB since it is located inside of it and it has a similar V_{LSR}. We identify a Class I source candidate to be the ionizing source of the UC H II region ($05^h40^m51^s.41$ $+35^\circ38'30''.0$, see Fig. 2). There is another Class I source at about 1 arcmin east from S235C ($05^h40^m57^s.7+35^\circ38'19''.1$) which shows an arc-shaped PAH structure pointing towards S235C.

5.7.2.4 S235AB. This cluster has the highest YSOs surface density in our sample (10^3 stars per pc^{-2}) and it has a different V_{LSR} than the rest of clusters located around the H II region S235 (see Fig. 13 and Table 5). We identified two Class I sources as candidates for powering the H II regions S235A and S235B with coordinates $05^h40^m52^s.58+35^\circ42'18''.6$ and $05^h40^m52^s.39+35^\circ41'29''.4$, respectively. We also identified 75 out of the 77 YSOs found by Dewangan & Anandarao (2011) as cluster members. Clusters BDSB71 to BDSB73 (Bica et al. 2003b) are included in S235AB.

5.7.2.5 G173.66+2.78. This cluster may be associated with cluster S235 due to its location and similar V_{LSR}. We identified all the YSOs found by Dewangan & Anandarao (2011) as cluster members.

5.7.2.6 S235. Also called S235 Central (Kirsanova et al. 2008), this cluster is apparently located inside the H II region S235. The projected distance from the cluster to the ionizing star BD+351201

is ∼1 pc. At this distance, we may expect to observe signatures of interaction between the ionizing front and the molecular material associated with the cluster (like in AFGL4029). The absence of such signatures suggest that the cluster is either located in another plane compared to the ionizing star, or not associated directly with the molecular material at this location. Since the ionizing star has a V_{LSR} of $-18\,km\,s^{-1}$(Kirsanova et al. 2008) and the ^{13}CO material associated with S235 has a V_{LSR} of $-19.8\,km\,s^{-1}$, we hypothesize that the cluster is located in a different plane than the H II region (Kirsanova et al. 2008).

The sources IRS1 and IRS2 are identified as cluster members and classified as Class I. The nine YSOs found by Dewangan & Anandarao (2011) are also identified as cluster members. The cluster CBB2 (Camargo et al. 2011) is part of cluster S235.

5.7.2.7 G173.62+2.88 (East 2). Located north-east from cluster S235. We identify 18 out of the 19 YSOs found by Dewangan & Anandarao (2011) as cluster members. Dewangan & Anandarao (2011) found 3 *Spitzer*-IRAC sources without NIR counterpart: e2s1, e2s2 and e2s3. Of those, we detect e2s2 and e2s3 in the *K*-band. All of them are classified as Class I.

5.7.2.8 G173.67+2.87 (East 1). We identify all the YSOs found by Dewangan & Anandarao (2011) as cluster members. Dewangan & Anandarao (2011) identified 4 *Spitzer*-IRAC sources without NIR counterparts: e1s1, e1s2, e1s3 and e1s4. We detect e1s2 and e1s3 in the *H* and *K* bands and e1s4 in the *J*, *H* and *K* bands. All of them (e1s1 to e1s4) are classified as Class I. The source e1s1 is also associated with extended emission seen in the *K*-band and at 4.5 μm, signs of outflow activity.

5.7.3 S252

5.7.3.1 G189.79+0.29. Located to the west of cluster S252A, it is associated with molecular material at V_{LSR} =10 $km\,s^{-1}$(S252A has V_{LSR} of $8.3\,km\,s^{-1}$).

5.7.3.2 G189.84+0.29. It may be associated with cluster S252A due to its similar V_{LSR} (see Table 5).

5.7.3.3 G189.95+0.22. This cluster has a V_{LSR} of $5.8\,km\,s^{-1}$, suggesting that it is located in a different plane in the sky than the rest of the clusters in the region.

5.7.3.4 S252A. It harbours the compact-H II region S252A on its north-west side. The ionizing source of this compact-H II region is classified as Class I ($06^h08^m32^s.05$ $+20^\circ39'18''.3$). The extension of this cluster agrees well with the sub-millimetre emission detected by Tej et al. (2006). The derived IMF slope is also in agreement with estimations by Jose et al. (2012) of −1.33. However, they cover a different mass range (from 0.3 to 2.5 $M_\odot$) and use a different methodology.

Both the ^{13}CO column density map and the extinction map show that the molecular material decreases towards the centre of the compact-H II region. This was also noticed by Tej et al. (2006) in their sub-millimetre observations. Moreover, they find that the peak emission at 850 μm lays ∼1 arcmin east from the centre of H II region, close to its edge. The sub-millimetre peak coincides with methanol and water maser emission and also with a Class I source ($06^h08^m35^s.33$ $+20^\circ39'06''.9$s) detected in the IRAC bands 3 and 4 and in the MIPS map. This source also seems to be associated with emission from shocked gas seen as a green blob in Fig. 3. Whether the formation of new stars at the sub-millimetre peak was triggered or not by the expansion of the contiguous H II region is still an open question.

5.7.3.5 G189.94+0.33. Cluster located on the PAH ridge, south-east of S252A and G189.84+0.29. Fig. 3 shows some extended radio emission at the east side of the PAH ridge. This emission may be due to the interaction of the ionizing front with the material along the ridge. The 24 μm emission observed right at the west side of the radio emission and the rapid increase of the column density evidence how the molecular cloud is being compressed by the expansion front. G189.94+0.33 may be also a candidate for triggered star formation.

5.7.3.6 S252C. This cluster has a ∼2 pc size lobe on the east side as shown in Fig. 3 and in the MIPS map. This lobe is probably created by strong stellar winds emitted by the ionizing source located close to the cluster centre. The ionizing source of the compact-H II region is classified as Class II.

5.7.3.7 G189.95+0.54. Cluster located ∼5 arcmin south-east from S252C. Its V_{LSR} suggest that it may be placed in a different plane in the sky than other clusters in the region.

5.7.3.8 S252E. Similarly to cluster S252C, the dust emission around cluster S252E shows lobes on its east and west side. The east lobe is not completely covered by our IRAC map. The candidate ionizing source of the compact-H II region S252E is classified as Class II.

S252E is located at a projected distance of about 3 pc from the ionizing star of the H II region NGC 2175. At this distance, we would expect to observe signs of interaction between the molecular material and the ionization front like a pillar. The absence of such a morphology suggest that this cluster is located in another plane compared to the ionizing star.

The H II region S252B is located between clusters S252A and S252C. We find no cluster around the ionizing source (classified as Class II) and propose three alternatives to explain these observations: isolated star formation, runaway star or non-detected cluster members. The ^{13}CO column-density map shows density enhancements on the west side of S252C and east side of S252A which are likely produced by the expansion of the H II region S252B. This means that the ionizing star of S252B is not a foreground or background source and that it could have been expelled from nearby embedded clusters due to dynamical interaction with other YSOs. Alternatively, this case may be similar to H II regions S255 and S257 in the S254-S258 complex. Those H II regions have a similar size than S252B and are powered by stars with similar spectral type. However, Chavarría et al. (2008, and this paper) identified some YSOs around the ionizing star in the H II region S257 suggesting that there is likely a cluster associated with that source. Those cluster members are difficult to detect due to the brightness of the extended emission inside the H II region, which may be the case for S252B. Furthermore, early disc evaporation due to a nearby ionizing star will make those cluster members difficult to detect at mid-IR wavelengths. Finally, we cannot leave aside the possibility that the ionizing source of H II region S252B was formed in isolation.

5.7.4 S254-S258

We investigate this region using the YSOs identified by Chavarría et al. (2008) and adding the sources detected from MIPS observations. In general, the number of cluster members as well as cluster sizes are very similar to the previous results from Chavarría et al. (2008). We find two main differences: former cluster S258 is now divided in clusters S258 and G192.70+0.03, and the number of identified members for cluster G192.63+0.00 decreased by a factor of 2. We believe that the last is due to the proximity of cluster S255-2; the YSOs in G192.63+0.00 are outnumbered by the dense population of S255-2 which decreases the value of the critical distance d_c used for both clusters. This way, the most dispersed population of G192.63+0.00 is not included as part of the cluster by our MST algorithm.

The new distance to this region (1.6 kpc, Rygl et al. 2010) has an effect in the derived YSOs surface densities and clusters molecular mass, changing the values given in Chavarría et al. (2008) by a factor of 2 or 3. However, this does not change the conclusions drawn from cluster to cluster comparison in the region since those factors are the same for all clusters.

5.7.5 NGC 7538

5.7.5.1 G111.44+0.79. This cluster is located west of the H II region NGC 7538, along a dusty filament connecting also clusters G111.48+0.80, G111.49+0.81 and NGC 7538. It is also associated with several clumps detected at 850 μm (Reid & Wilson 2005).

5.7.5.2 G111.44+0.75. Cluster located ∼1 arcmin south of G111.44+0.79 and associated with several sub-millimetre clumps (Reid & Wilson 2005). This cluster is located along a different filament than G111.44+0.79 that is also connected with cluster NGC 7538.

5.7.5.3 G111.48+0.80 and G111.49+0.81. These clusters are located between the H II region NGC 7538 and cluster G111.44+0.79. They are associated with the same dust filament as G111.44+0.79 as well as with sub-millimetre clumps (Reid & Wilson 2005).

5.7.5.4 G111.47+0.75. Cluster located along the same dusty filament as G111.44+0.75. It is associated with sub-millimetre emission (Reid & Wilson 2005) and with a water maser (H1, from Kameya et al. 1990). This maser is located less than 1 arcsec from a Class I source with coordinates $23^h13^m22\overset{s}{.}2$ $+61^\circ25'44\overset{''}{.}1$.

5.7.5.5 NGC 7538. Located along the south side of H II region NGC 7538 (cluster NGC 7538S is part of cluster NGC 7538). NGC 7538 is the largest and most massive of all the identified clusters in the five studied regions. However, since this region is also the furthest one, the cluster could be composed of several substructures that are not clearly separated by the MST algorithm. This could explain the cluster difference in the peak $\Sigma_{\rm YSO}$ value with respect to the other clusters trend in Fig. 19.

Sources IRS1 to 4, IRS9 and IRS11 are classified as Class I. Some sources with spectral type of early B-star (Puga et al. 2010) and the far-infrared source NGC 7538S are members of cluster NGC 7538. The ^{13}CO density map shows that the column density increases rapidly from the centre to the south and north sides of the H II region. The densest area in the region is located around sources IRS1-IRS3. It is a very active site of star formation and it seems to be the joint point of two dusty filaments (A and B, see Fig. 16).

5.7.5.6 G111.57+0.81. This cluster is located east from the H II region NGC 7538. It contains two Class I sources located at the tip of pillar-like structures and associated with the sub-millimetre clumps SMM 61 and SMM 57 from Reid & Wilson (2005) (their coordinates are $23^h14^m01\overset{s}{.}7$ $+61^\circ30'15\overset{''}{.}7$ and $23^h13^m58\overset{s}{.}5$ $+61^\circ30'51\overset{''}{.}5$, respectively).

5.7.5.7 G111.59+0.64 and G111.65+0.63. These clusters are located 3–5 arcmin south-east from the H II region NGC 7538. They are associated with molecular material at $V_{\rm LSR}$ of -51 km s^{-1}, suggesting that they belong to a different plane in the sky than the rest of the clusters in the region (see Fig. 16 and Table 5).

6 CONCLUSIONS

We performed a multiwavelength study on five regions of massive star formation: W5-east, S235, S252, S254-S258 and NGC 7538. *Spitzer*-IRAC/MIPS and NIR observations were used to classify the stellar population. While the molecular content was studied using ^{12}CO, ^{13}CO observations and extinction maps.

We found in total 3021 YSOs, including 539 Class I and 1186 Class II. An MST algorithm was used to identify YSO clusters based on the characteristic separation of their members. A total of 41 embedded clusters were found, 15 of which have not being identified before.

The Class I sources are spatially correlated with the most dense molecular material. They are also located in regions with higher YSOs surface density and are distributed more hierarchically than Class II. All this agrees well with the picture where stars are formed in dense and fractally arranged dust filaments. Then, dynamical interactions rearrange the YSOs in a more centrally condensed distribution.

We find that the mean separation between cluster members is smaller than the cluster Jeans length in most cases. This difference is more evident in the case of MEC. In addition, the ^{13}CO linewidth of clusters associated molecular material shows that the clusters are turbulent. This agrees with a scenario in which fragmentation is likely driven by turbulence. Though magnetic fields may also play an important role as a support against gravity.

Between 30 and 50 per cent of the total number of YSOs are not included in clusters. This percentage of scattered population depends on the used cluster finding algorithm. We propose that between 10 and 20 per cent of the scattered population in the studied regions corresponds, indeed, to cluster members.

We compared the physical properties of embedded clusters associated with high-mass stars and clusters with no evidence of harbouring massive stars. We find no systematical differences in the correlations derived for both samples. In all cases, the MEC seem to be an extrapolation of the LEC.

The correlation between the clusters dense mass and the number of cluster members is investigated. We find that this correlation is close to linear, in agreement with previous findings from Lada et al. (2010). We also find that the SFE is rather constant along the clusters mass range. In average, the estimated SFE agrees well with the mass factor between the cores mass function and the IMF.

The spatial distribution of Class I and Class II sources in clusters AFGL4029 and G138.15+1.69 suggests a sequential star formation which moves in the same direction as the ionization front. This is in agreement with previous estimation of the YSOs ages around these

clusters and supports the hypothesis that both clusters were created in a triggered star formation scenario (see Chauhan et al. 2011, and references there in).

We classify a total of 24 OB-type stars as either Class I or Class II sources. The IR-excess emitted by those sources may be due to a dusty structure around them. Those sources are good candidates for future high-resolution studies aimed to search for discs in high-mass stars. The presence of a rotating structure around massive stars suggest that they are formed via accretion.

ACKNOWLEDGEMENTS

This work is based in part on observations made with the *Spitzer* Space Telescope, which is operated by the Jet Propulsion Laboratory, Caltech, under a contract with NASA. Support for this work was provided by NASA through a contract issued by JPL/Caltech. We also thank NOAO for their student thesis support. The Five College Radio Astronomy Observatory was supported by NSF grant AST 0540852. CB is supported by an RCUK Fellowship at the University of Exeter, UK. This work is based in part on the IRAC post-BCD processing software 'IRACPROC' developed by Mike Schuster, Massimo Marengo and Brian Patten at the Smithsonian Astrophysical Observatory. This research used the facilities of the Canadian Astronomy Data Centre operated by the National Research Council of Canada with the support of the Canadian Space Agency. This research has made use of the NASA/ IPAC Infrared Science Archive, which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. We thank the Spanish MINECO for funding support from grants CSD2009-00038, AYA2009-07304 and AYA2012-32032.

REFERENCES

Allen L. et al., 2004, ApJS, 154, 363
Alves J., Lada C., Lombardi M., 2007, A&A, 462, 17
André Ph. et al., 2010, A&A, 518, 102
Balog Z., Kenyon S., Lada E., Barsony M., Vinkó J., Gáspár A., 2004, AJ, 128, 2942
Barriault L., Joncas G., 2007, ApJ, 667, 257
Battinelli P., 1991, A&A, 244, 69
Becker W., Fenkart R., 1971, A&AS, 4, 241
Beichman C. A., 1979, PhD thesis, Hawaii Univ.
Bica E., Dutra C., Barbuy B., 2003a, A&A, 397, 177
Bica E., Dutra C., Soares J., Barbuy B., 2003b, A&A, 404, 223
Bonnell I. A., Vine S. R., Bate M. R., 2004, MNRAS, 349, 735
Bonnell I. A., Clarke C. J., Bate M. R., 2008, MNRAS, 389, 1556
Bressert E. et al., 2012, A&A, 542, 49
Brunt C., 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. 79
Camargo D., Bonatto C., Bica E., 2011, MNRAS, 416, 1522
Cardelli J., Clayton C., Mathis J., 1989, ApJ, 345, 245
Carpenter J., Snell R., Schloerb F., Skrutskie M., 1993, ApJ, 407, 657
Carpenter J., Snell R., Schloerb F., 1995, ApJ, 450, 201
Carpenter J., Heyer M., Snell R., 2000, ApJS, 130, 381
Cartwright A., Whitworth A., 2004, MNRAS, 348, 589
Cesaroni R., Neri R., Olmi L., Testi L., Walmsley C. M., Hofner P., 2005, A&A, 434, 1039
Chauhan N., Pandey A., Ogura K., Jose J., Ojha D., Samal M., Mito H., 2011, MNRAS, 415, 1202
Chavarría L., Allen L., Hora J., Brunt C., Fazio G., 2008, ApJ, 682, 445
Chavarría L., Mardones D., Garay G., Escala A., Bronfman L., Lizano S., 2010, ApJ, 710, 583
Chavarría-K C., Leitherer C., de Lara E., Sánchez O., Zickgraf F., 1989, A&A, 215, 51
Churchwell E., Walmsley C., Cesaroni R., 1990, A&AS, 83, 119
Condon J., Cotton W., Greisen E., Yin Q., Perley R., Taylor G., Broderick J., 1998, AJ, 115, 1693
Crampton D., Georgelin Y. M., Georgelin Y. P., 1978, A&A, 66, 1
Cyganowski C. et al., 2008, AJ, 136, 2391
Dale J., Ercolano B., Bonnell I., 2013, MNRAS, 431, 1062
Deharveng L., Zavagno A., Cruz-González I., Salas L., Carrasco L., 1997, A&A, 317, 459
Deharveng L., Lefloch B., Kurtz S., Nadeau D., Pomarès M., Caplan J., Zavagno A., 2008, A&A, 482, 585
Deharveng L. et al., 2012, A&A, 546, 74
Dewangan L., Anandarao B., 2011, MNRAS, 414, 1526
Dickman R., 1978, ApJS, 37, 407
Eisenhardt P. et al., 2004, ApJS, 154, 48
Erickson N., Grosslein R., Erickson R., Weinreb S., 1999, IEEE Trans. Microw. Theory Tech., 47, 2212
Evans N., Blair G., 1981, ApJ, 246, 394
Fazio G. G. et al., 2004, ApJS, 154, 10
Felli M., Habing H., Israel F., 1977, A&A, 59, 43
Felli M., Hjellming R., Cesaroni R., 1987, A&A, 182, 313
Felli M., Massi F., Robberto M., Cesaroni R., 2006, A&A, 453, 911
Flaherty K., Pipher J., Megeath T., Winston E., Gutermuth R., Muzerolle J., Allen L., Fazio G., 2007, ApJ, 663, 1069
Frerking M., Langer W., Wilson R. W., 1982, ApJ, 262, 590
Froebich D., Scholz A., Raftery C., 2007, MNRAS, 374, 399
Garay G., Mardones D., Brooks K., Videla L., Contreras Y., 2007, ApJ, 666, 309
Georgelin Y. M., Georgelin Y. P., Roux S., 1973, A&A, 25, 337
Ginsburg A., Bally J., Williams J. P., 2011, MNRAS, 418, 2121
Grasladen G., Carrasco L., 1975, A&A, 43, 259
Gutermuth R. et al., 2008, ApJ, 674, 336
Gutermuth R., Megeath T., Myers P., Allen L., Pipher J., Fazio G., 2009, ApJS, 184, 18
Herpin F. et al., 2012, A&A, 542, 76
Heyer M., Snell R., Morgan J., Schloerb F., 1989, ApJ, 346, 220
Heyer M., Brunt C., Snell R., Howe J., Schloerb F., Carpenter J., 1998, ApJS, 115, 241
Hillwig T., Gies D., Bagnuolo W., Huang W., McSwain M., Wingert D., 2006, ApJ, 639, 1069
Hiltner W. A., 1956, ApJ, 120, 454
Hora J. et al., 2004, in Mather J., ed., Proc. SPIE, Vol. 5487, Optical, Infrared, and Millimeter Space Telescopes. SPIE, Bellingham, p. 77
Hosokawa T., Inutsuka S., 2006, ApJ, 648, 131
Hunter D., Massey P., 1990, AJ, 99, 846
Indebetouw R. et al., 2005, ApJ, 619, 931
Israel F., 1977, A&A, 59, 27
Israel F., Felli M., 1978, A&A, 63, 325
Israel F. et al., 2003, A&A, 401, 99
Jackson T., Ivezić Z., Knapp G., 2002, MNRAS, 337, 749
Jose J. et al., 2012, MNRAS, 424, 2486
Kameya O., Morita K-I., Kawabe R., Ishiguro M., 1990, ApJ, 355, 562
Kirsanova M., Sobolev A., Thomasson M., Wiebe D., Johansson L., Seleznev A., 2008, MNRAS, 388, 729
Koenig X., Allen L., Gutermuth R., Hora J., Brunt C., Muzerolle J., 2008, ApJ, 688, 1142
Koposov S., Glushkova E., Zolotukhin I., 2008, A&A, 486, 771
Krassner J., Pipher J., Sharpless S., 1979, A&A, 77, 302
Kraus S. et al., 2006, A&A, 455, 521
Kumar M., Davis C., Grave J., Ferreira B., Froebrich D., 2007, MNRAS, 374, 54
Kurtz S., Churchwell E., Wood D., 1994, ApJS, 91, 659
Kutner M. L., Ulich B. L., 1981, ApJ, 250, 341
Lada C., 1987, in Lada C. J., ed., Proc. IAU Symp. 115, Star-forming Regions. Reidel, Dordrecht, p. 1
Lada C., Lada E., 2003, ARA&A, 41, 57
Lada C., Wooden D., 1979, ApJ, 232, 158

Lada C., Lombardi M., Alves J., 2010, ApJ, 724, 687
MacKenzie T. et al., 2011, MNRAS, 415, 1950
Mampaso A., Pismis P., Vilchez J., Phillips J., 1987, Rev. Mex. Astron. Astrofis., 14, 474
Maschberger Th., Clarke C., Bonnell I., Kroupa P., 2010, MNRAS, 404, 1061
McKee Ch., Tan J., 2003, ApJ, 585, 850
Meyer M., Calvet N., Hillendrand L., 1997, AJ, 114, 288
Minkowski R., 1946, PASP, 58, 305
Mirabel I., Ruiz A., Rodríguez L., Cantó J., 1987, ApJ, 318, 729
Morgan L., Thompson M., Urquhart J., White G., 2008, A&A, 477, 577
Moscadelli L., Reid M., Menten K., Brunthaler A., Zheng X., Xu Y., 2009, ApJ, 693, 406
Muench A., Lada E., Lada C., 2000, ApJ, 533, 358
Muench A., Lada E., Lada C., Alves J., 2002, ApJ, 573, 366
Muench A., Lada C., Luhman K., Muzerolle J., Young E., 2007, ApJ, 134, 411
Nakano M., Sugitani K., Niwa T., Itoh Y., Watanabe M., 2008, PASJ, 60, 739
Niwa T., Tachihara K., Itoh Y., Oasa Y., Sunada K., Sugitani K., Mukai T., 2009, A&A, 500, 1119
Noriega-Crespo A. et al., 2004, ApJS, 154, 352
Oey M., Lamb J., Kushner C., Pellegrini E., Graus A., 2013, ApJ, 768, 66
Ogura K., Sugitani K., Pickles A., 2002, AJ, 123, 2597
Ojha D., Ghosh S., Kulkarni V., Testi L., Verma R., Vig S., 2004, A&A, 415, 1039
Ojha D. et al., 2011, ApJ, 738, 156
Olofsson G., 1983, A&A, 120, 1
Patten B. et al., 2006, ApJ, 651, 502
Pestalozzi M., Elitzur M., Conway J., 2009, A&A, 501, 999
Price S. D., Walker R. G., 1976, Interim Report Air Force Geophysics Lab., Optical Physics Div., The AFGL Four Colour Infrared Sky Survey: Catalog of Observations at 4.2, 11.0, 19.8 and 27.4 Micrometers. Air Force Geophysics Lab., Hanscom AFB, MA
Puga E. et al., 2010, A&A, 517, 2
Ray T., Poetzel R., Solf J., Mundt R., 1990, ApJ, 357, 45
Reid M., Wilson C., 2005, ApJ, 625, 891
Reid M., Menten K., Brunthaler A., Zheng X., Moscadelli L., Xu Y., 2009, ApJ, 693, 397
Robitaille T. et al., 2008, AJ, 136, 2413
Rygl K., Bunthaler A., Reid M., Menten K., van Langevelde H., Xu Y., 2010, A&A, 511, 2
Sandell G., Wright M., 2010, ApJ, 715, 919
Sandell G., Goss W., Wright M., Corder S., 2009, ApJ, 699, 31
Schmeja S., 2011, Astron. Nachr., 332, 172
Schmeja S., Klessen R., 2006, A&A, 449, 151
Schmeja S., Kumar M., Ferreira B., 2008, MNRAS, 389, 1209
Schuster M., Marengo M., Patten B., 2006, in Silva D., Doxsey R., eds, Proc. SPIE, Vol. 6270, Observatory Operations: Strategies, Processes, and Systems. SPIE, Bellingham, p. 65
Sharpless S., 1959, ApJS, 4, 257
Snell R., Huang Y., Dickman R., Claussen M., 1988, ApJ, 325, 853
Tej A., Ojha D., Ghosh S., Kulkarni V., Verma R., Vig S., Prabhu T., 2006, A&A, 452, 203
Thompson R., Thronson H., Campbell B., 1983, ApJ, 266, 614
Troland T., Crutcher R., 2008, ApJ, 680, 457
Vlemmings W. H., 2008, A&A, 484, 773
Wang Y. et al., 2011, A&A, 527, 32
Werner M., Becklin E., Gatley I., Matthews K., Neugebauer G., Wynn-Williams C., 1979, MNRAS, 188, 463
Winston E. et al., 2007, ApJ, 669, 493
Wouterloot J., Brand J., 1989, A&A, 80, 149
Wynn-Williams C., Beckling E., Neugebauer G., 1974, ApJ, 187, 473
Xu Y. et al.2006, AJ, 132, 20
Zapata L., Rodríguez L., Kurtz S., 2001, Rev. Mex. Astron. Astrofis., 37, 83
Zinnecker H., Yorke H., 2007, ARA&A, 45, 481

APPENDIX A: BACKGROUND CONTAMINATION ESTIMATES

The detection of distant galaxies in our mid-IR data is expected due to the relative transparency of the molecular clouds at these wavelengths. Since galaxies and YSOs have similar IR colours, it is important to find a way to distinguish between both to avoid including galaxies as part of our YSOs population. A simple method would be to remove all the sources located in the GDA in the colour–magnitude diagrams (e.g. Chavarría et al. 2008; Gutermuth et al. 2009). However, the number of background galaxies will decrease in highly embedded regions due to the dust extinction. Therefore, we complement the GDA with our extinction maps and for each YSO candidate, we calculate the extinction along the line of sight. Then, we apply this extinction to the galaxies in the colour–magnitude diagram to emulate the galaxies colour as if they were affected by the same extinction as the YSO candidate. Finally, we define the GDA in the colour–magnitude diagram and confirm the candidate as a YSO if it is located outside the GDA.

We use different colour–magnitude diagrams for the different YSOs: [4.5] versus [4.5] − [8.0] for Class I and Class II candidates, K versus K−[4.5] for IR-excess candidates and J versus J−K for sources detected only in the NIR bands. We use NIR and *Spitzer*-IRAC galaxy observations from the IRAC Shallow Survey (Eisenhardt et al. 2004), which covers 8.5 square degrees of the sky in the NOAO Deep Wide-Field Survey in Boötes and used 3 × 30 second frames (deeper than our IRAC observations). Fig. A1 shows as an example the GDA in the [4.5] versus [4.5] − [8.0] colour–magnitude diagram for W5-east at $A_K = 0$. In total we identified 11 Class I and 29 Class II sources as background contamination in W5-east, 11 Class I and 2 Class II in S235, 15 Class I and 5 Class II

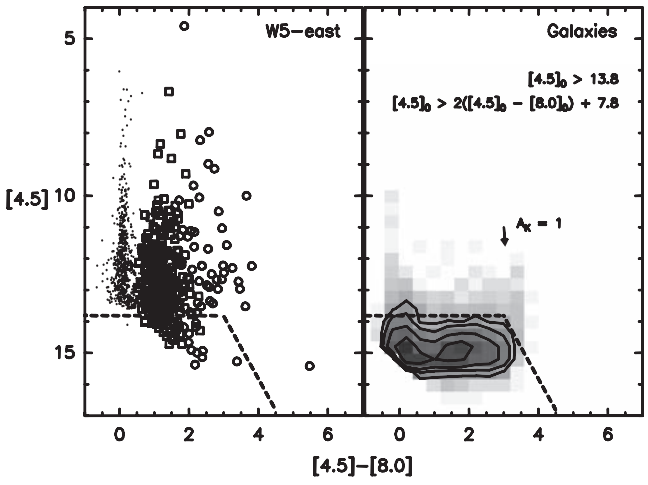

Figure A1. Colour–magnitude diagram used to remove background contamination in region W5-east. This diagram is used to eliminate contaminants detected in the four IRAC bands (K versus K−[4.5]] and J versus J−K diagrams are used for the other YSOs classification schemes). Dashed lines enclose the galaxy-dominated areas (GDA) in each chart. The chart on the left shows all the detected sources. Candidate YSOs are labelled as circles (Class I) and squares (Class II). Other detections are shown as dots. The chart on the right shows the galaxies from the IRAC Shallow survey for $A_K = 0$. Contours begin at 0.5 galaxy per mag^{-2}. The arrow shows the extinction vector for $A_K = 1$. For each YSO candidate i with associated extinction A_{K_i}, the GDA moves along the extinction vector to reach the YSO extinction. Then, the candidate is removed if it is located inside the GDA.

in S252, 1 Class I and 1 Class II in S254-S258 and 7 Class I and 9 Class II in NGC 7538.

Unresolved asymptotic giant branch (AGB) stars will have colours similar to YSOs in the near and mid-IR wavelengths. We estimate the amount of AGB contaminants over the observed fields using the AGB star distribution derived by Jackson, Ivezić & Knapp (2002) and Robitaille et al. (2008). We find that the number of AGB stars for each region FOV is of the order of 1. This is due to the location of the studied regions towards the outer parts of Galaxy. Since the AGB star contamination is negligible with respect to the number of YSOs, we did not perform a further correction in this matter.

APPENDIX B: COMPARING YSO DETECTIONS WITH PREVIOUS PUBLICATIONS IN W5-EAST AND S235

Ogura et al. (2002) and Nakano et al. (2008) identified 114 stars with Hα emission in region W5-east over the same FOV as in our observations. Of those, we identify 84 (74 per cent) as YSOs: 3 as Class I, 65 as Class II and 16 as having IR-excess. In the same region, and using IRAC+NIR data, Koenig et al. (2008) found 292 YSOs (including Class I, Class II and transition discs following their classification scheme). Of those, we identified 251 (86 per cent): 36 Class I, 191 Class II and 24 IR-excess sources. In region S235, Dewangan & Anandarao (2011) detected 230 YSOs using also IRAC+NIR data. Of those, we detected 220 (96 per cent); 46 Class I, 165 Class II and 9 sources with IR-excess.

The longer integration times and better spatial resolution of our NIR observations have a direct impact in the number of identified YSOs. In region W5-east, we identify 478 YSOs, while Koenig et al. (2008) identified 292 over the same FOV. In region S235, we identify 630 YSOs while Dewangan & Anandarao (2011) identified 230 in the same FOV. Approximately 60 per cent of the new YSOs

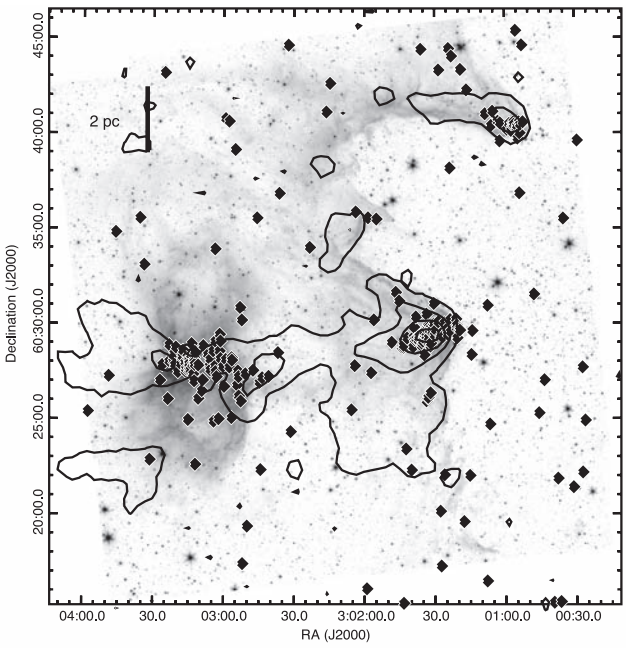

Figure B1. Distribution of YSOs identified in this work that were not identified as either Class I or Class II by Koenig et al. (2008) in region W5-east. The contours show the ^{13}CO surface density distribution. Contours are between 4×10^{15} and 4×10^{16} cm^{-2}.

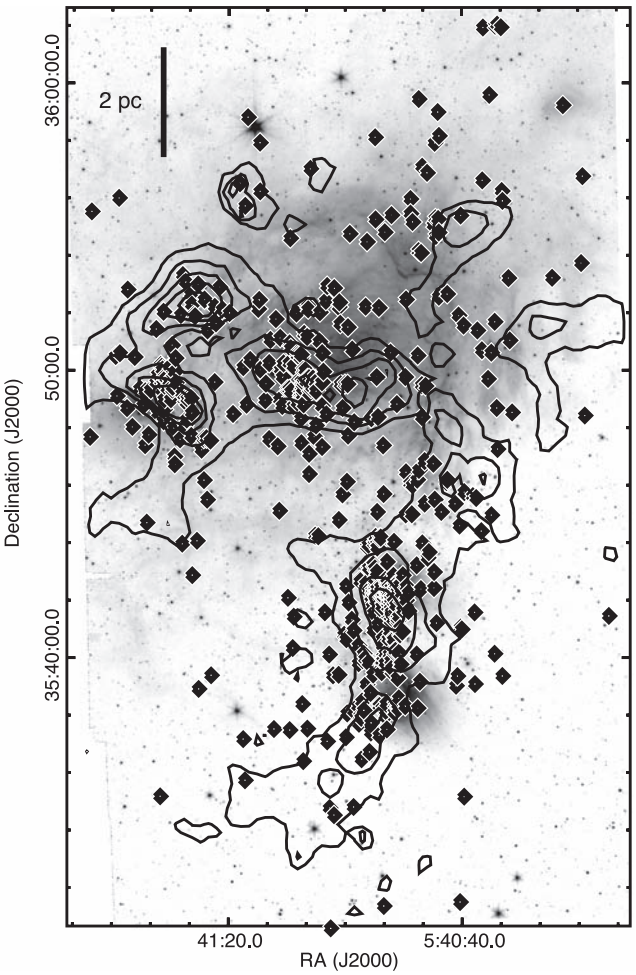

Figure B2. Distribution of YSOs identified in this work that were not identified by Dewangan & Anandarao (2011) in region S235. The contours show the ^{13}CO surface density distribution. Contours are between 4×10^{15} and 1×10^{16} cm^{-2}.

are detected using NIR observations (66 and 59 per cent in W5-east and S235, respectively). Figs B1 and B2 show the distribution of the YSOs that were not identified by Koenig et al. (2008) and Dewangan & Anandarao (2011) in W5-east and S235, respectively. In both regions, those sources are distributed in groups whose location correlates well with the molecular material and embedded clusters. If they were background galaxies, they would be distributed homogeneously over the field. We conclude that this population is mainly composed of YSOs.

APPENDIX C: ESTIMATING THE NON-DETECTED CLUSTER POPULATION

We identify the young stellar population in regions located at a few kpc from the Sun. At those distances, our completeness limit allows us to detect stars down to ~0.5–1 $M_{\odot}$ and we would like to estimate the percentage of sources below this limit. For example, if we assume that the embedded clusters follow an Orion IMF (Muench et al. 2002), the percentage of non-detected members would be up to 50 per cent. Now, if we account for the cluster extinction, this percentage may be even higher. In the following, we explain the algorithm used to estimate the percentage of non-detected cluster members.

The completeness limit will vary from one cluster to another due to their different properties like extinction, extended emission brightness and separation between cluster members. To account for these differences, we calculated the 90 per cent completeness limit for each cluster individually. This is done by adding synthetic stars over the cluster *K*-band image and then counting the percentage of synthetic stars recovered. The completeness limit (CL_K) corresponds to the magnitude where this percentage goes below 90 per cent. After this, we calculate the percentage of synthetic stars (κ) with *K* magnitudes fainter than CL_K from a synthetic cluster having similar properties as the observed clusters (see Appendix D) and estimate the corrected number of cluster members (N^*_{YSO}) from

$$N^*_{YSO} = \frac{N_{YSO}/(1 - IR)}{1 - \kappa}. \quad \text{(C1)}$$

Where *IR* is the percentage of cluster members without IR-excess (see Section D1). One important assumption when creating the synthetic cluster is the assumed IMF. In our case, we used the Orion-Trapezium IMF from Muench et al. (2002). However, in Section 5.5, we show that the high-mass IMF slope (Γ_{HM}, see Table 7) for some clusters differs from the Orion-Trapezium value ($\Gamma_{HM} = -1.21$). This difference becomes the main source of error when estimating percentage of detected sources κ and the corrected number of cluster members N^*_{YSO}. We account for this by calculating κ for $\Gamma_{HM} = -1.00$ and $\Gamma_{HM} = -1.60$ (which encloses the observed Γ_{HM}values from Table 7). We find that κ varies by on average 10 per cent, which translates into a factor of less than 3 for N^*_{YSO}. The error bars in Fig. 26 are derived from those calculations.

APPENDIX D: ADDING CLUSTER PROPERTIES TO SYNTHETIC CLUSTER

We use synthetic clusters to estimate the KLF and the percentage of non-detected cluster members. Those synthetic clusters are created using the Monte Carlo code described in Muench et al. (2000). This code allows us to convolve synthetic clusters with ad hoc extinction and/or IR-excess distribution functions in order to reproduce individual cluster properties. For the extinction distribution we use values from our extinction maps.

D1 IR-excess

The IR-excess of the observed sources, given by their $H - K$ colour, was calculated assuming that affects only the *K*-band and that the intrinsic $H - K$ colour of cluster members is 0.3. We estimate the IR-excess by first derredening the YSOs to the CTTS loci from Meyer, Calvet & Hillendrand (1997) transformed to the K-4.5 versus $H - K$ colour diagram (using the extinction law from Flaherty et al. 2007), and then moving them to their intrinsic $H - K$ colour. Then, the IR-excess distribution function is convolved with the synthetic cluster *K*-band magnitudes.

In addition, we use the ratio between the number of YSOs and the number of field subtracted *K*-band detections inside the convex hull area (see Section D2) to estimate the percentage of cluster members without IR-excess.

D2 Field contamination in KLF

The KLF is derived from all the detections in the *K*-band in an area enclosing the cluster. Evidently, not all those detections are cluster members and hence the observed luminosity function is contaminated by the presence of field stars. We remove this contamination by subtracting a control field KLF from the cluster KLF. In order to do this properly, we added the extinction due to the molecular cloud to the field stars. This is done by convolving the control field with the extinction distribution of the cluster members. Once the control field is convolved, we construct the KLF, normalize it by the area and subtract it from the cluster KLF.

APPENDIX E: GAUSSIAN DECOMPOSITION OF ^{13}CO SPECTRUM IN EMBEDDED CLUSTERS

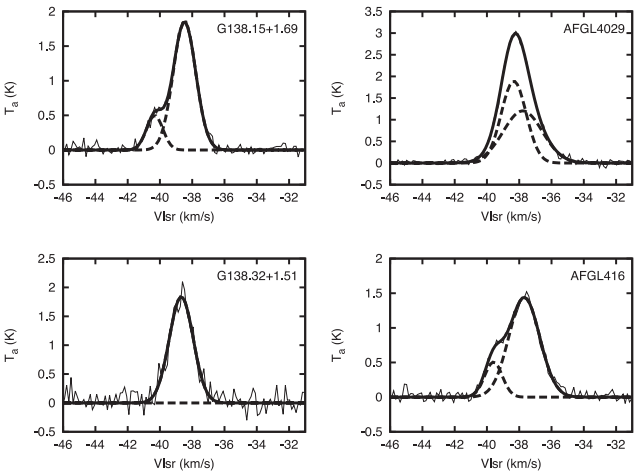

Figure E1. Integrated spectrum over the corresponding convex hull area for embedded clusters in region W5-east. The spectrum is decomposed in several Gaussian components. The continuous line shows the components added. Different components are shown by segmented lines. Parameters are summarized in Table E1.

Table E1. Clusters ^{13}CO spectrum decomposition.

Name	T_1 (K)	V_{LSR1} (km s^{-1})	FWHM$_1$ (km s^{-1})	T_2 (K)	V_{LSR2} (km s^{-1})	FWHM$_2$ (km s^{-1})	T_3 (K)	V_{LSR3} (km s^{-1})	FWHM$_3$ (km s^{-1})	T_4 (K)	V_{LSR4} (km s^{-1})	FWHM$_4$ (km s^{-1})
G138.15+1.69	1.85	−38.5	1.65	0.50	−40.3	1.20	–	–	–	–	–	–
AFGL4029	1.89	−38.3	1.81	1.20	−37.8	2.84	–	–	–	–	–	–
G138.32+1.51	1.84	−38.7	1.74	–	–	–	–	–	–	–	–	–
AFGL416	1.43	−37.7	2.26	0.51	−39.6	1.29	–	–	–	–	–	–
G173.51+2.79	3.60	−21.4	1.39	–	–	–	–	–	–	–	–	–
G173.63+2.69	2.36	−17.4	1.06	1.37	−16.6	2.84	–	–	–	–	–	–
S235C	4.13	−16.2	1.72	–	–	–	–	–	–	–	–	–
S235AB	2.63	−16.7	1.76	0.68	−16.4	4.11	0.48	−15.3	1.13	–	–	–
G173.66+2.78	1.34	−19.7	4.14	–	–	–	–	–	–	–	–	–
S235	2.69	−19.8	1.74	1.69	−19.8	3.08	–	–	–	–	–	–
G173.62+2.88	4.89	−20.7	1.36	1.15	−21.8	1.57	0.61	−19.8	3.67	–	–	–
G173.67+2.87	2.87	−18.8	1.69	1.89	−19.7	3.29	–	–	–	–	–	–
G189.79+0.29	4.07	10.0	1.93	3.53	7.3	1.97	1.65	8.7	1.29	–	–	–
G189.84+0.29	4.80	8.8	2.82	2.55	6.6	1.76	–	–	–	–	–	–
G189.95+0.22	3.27	5.8	1.88	1.29	7.8	1.79	–	–	–	–	–	–
S252A	6.10	8.3	3.57	0.22	4.3	2.75	–	–	–	–	–	–
G189.94+0.33	2.96	8.7	2.16	1.84	7.4	1.65	0.42	7.6	4.42	–	–	–
S252C	1.34	8.7	2.94	0.52	5.8	3.53	–	–	–	–	–	–
G189.95+0.54	0.50	9.5	1.27	1.14	14.6	1.48	0.44	5.2	2.44	–	–	–
S252E	1.26	8.5	1.72	0.75	7.5	3.20	–	–	–	–	–	–
G192.69−0.25	2.05	6.7	2.44	–	–	–	–	–	–	–	–	–
G192.54−0.15	0.76	24.1	1.67	–	–	–	–	–	–	–	–	–
S256	3.42	7.3	2.16	–	–	–	–	–	–	–	–	–
G192.65−0.08	4.21	6.5	1.90	–	–	–	–	–	–	–	–	–
S255-2 and S255N	2.82	6.6	2.14	1.81	8.0	3.01	–	–	–	–	–	–
G192.55−0.01	0.67	5.9	2.19	0.45	8.3	1.18	–	–	–	–	–	–
G192.75−0.08	2.79	7.1	1.97	–	–	–	–	–	–	–	–	–
G192.63+0.00	2.70	6.5	1.57	0.68	6.2	4.16	–	–	–	–	–	–
G192.70+0.03	3.55	8.0	1.39	0.47	7.7	4.00	–	–	–	–	–	–
G192.75+0.00	1.93	8.2	1.81	–	–	–	–	–	–	–	–	–
S258	3.55	8.2	1.72	0.70	6.8	1.20	–	–	–	–	–	–
G111.44+0.79	3.38	−57.8	5.05	1.81	−51.2	3.06	1.67	−53.8	2.87	–	–	–
G111.44+0.75	2.80	−54.5	6.53	1.23	−50.3	1.90	0.44	−60.7	3.78	–	–	–
G111.48+0.80	3.35	−57.6	5.03	2.16	−52.5	2.26	0.63	−49.9	2.09	–	–	–
G111.49+0.81	4.37	−58.9	3.92	1.90	−53.1	3.97	–	–	–	–	–	–
G111.47+0.75	3.14	−56.4	5.88	1.90	−53.2	2.14	1.73	−51.1	2.35	–	–	–
NGC 7538	2.82	−56.4	4.00	1.46	−53.1	3.22	0.87	−49.1	2.51	0.54	−59.0	5.38
G111.57+0.81	1.74	−55.5	4.16	0.45	−52.7	1.15	0.19	−50.7	4.49	–	–	–
G111.61+0.77	1.50	−52.8	2.23	1.42	−56.0	2.77	0.38	−49.3	3.90	–	–	–
G111.59+0.64	0.87	−48.8	2.02	0.82	−50.9	5.31	0.20	−56.6	6.51	–	–	–
G111.65+0.63	2.54	−50.5	3.60	1.27	−47.8	2.16	–	–	–	–	–	–

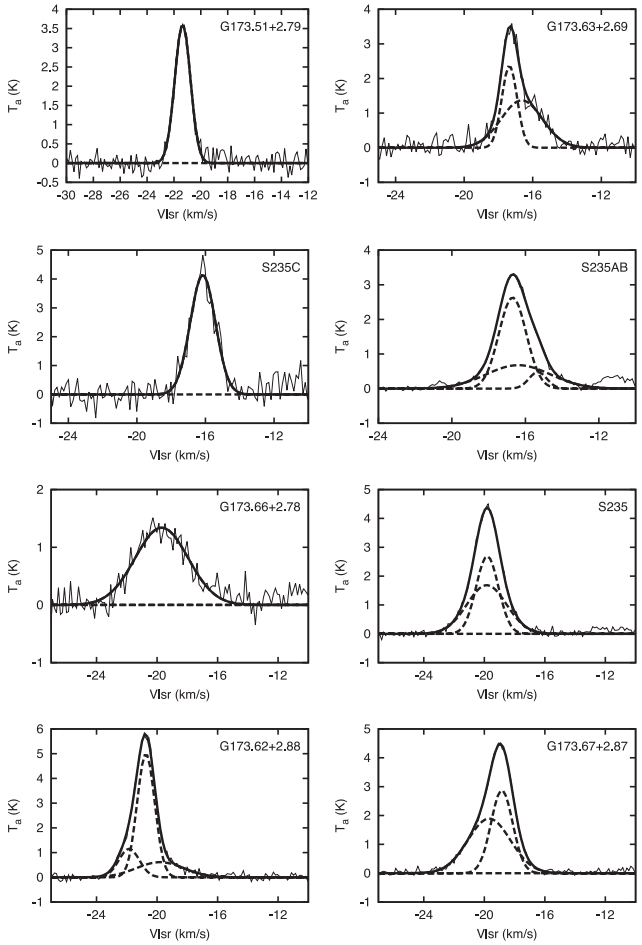

Figure E2. Same as Fig. E1 but for region S235.

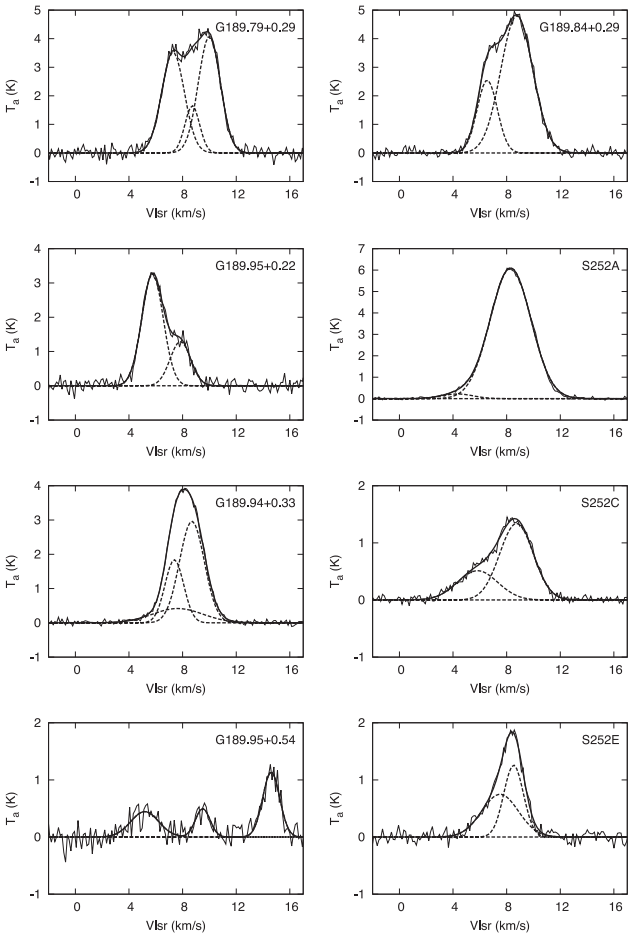

Figure E3. Same as Fig. E1 but for region S252.

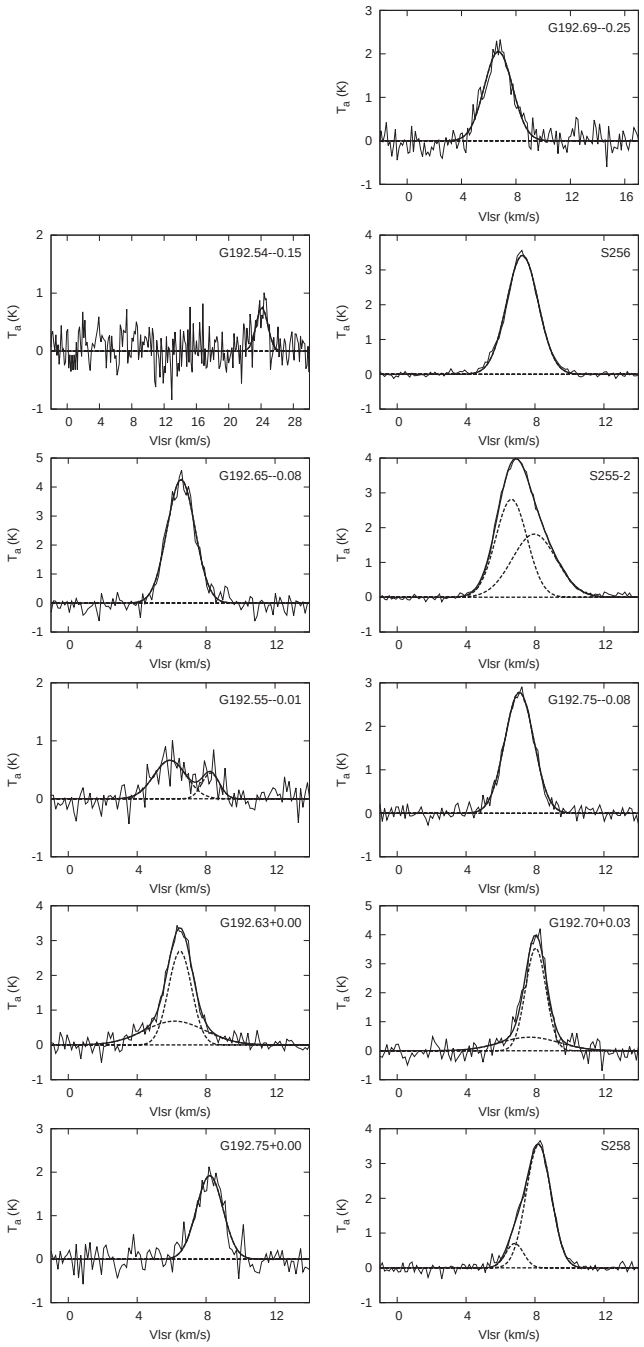

Figure E4. Same as Fig. E1 but for region S254-S258.

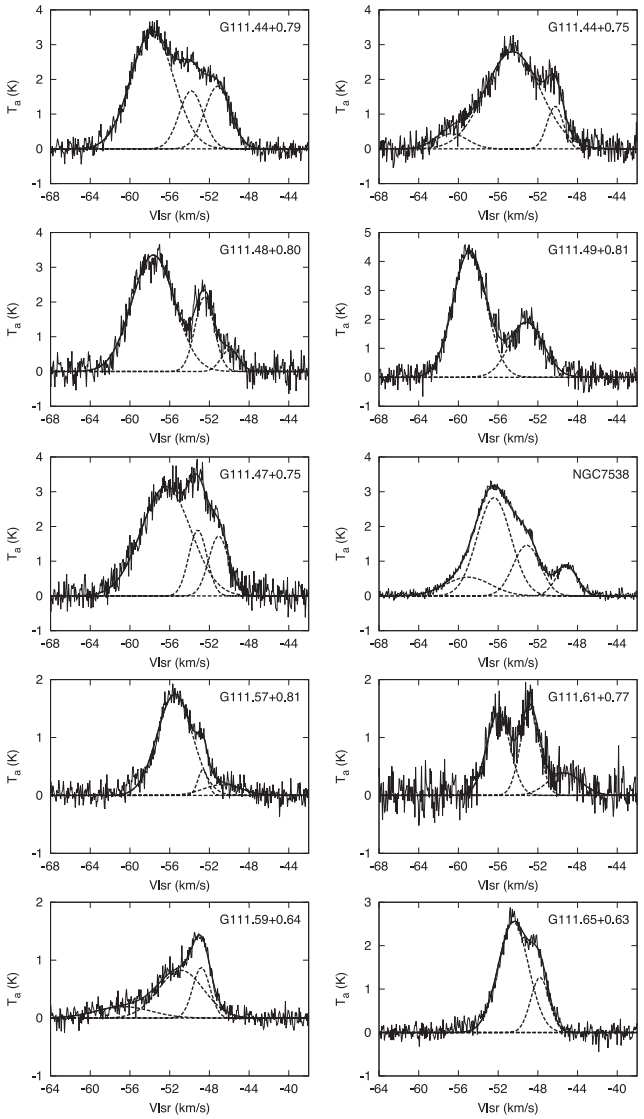

Figure E5. Same as Fig. E1 but for region NGC 7538.

SUPPORTING INFORMATION

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

Table 2. Identified YSOs per region (http://mnras.oxfordjournals.org/lookup/suppl/doi:10.1093/mnras/stu224/-/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 paper.

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