

# VISION - Vienna survey in Orion

## III. Young stellar objects in Orion A

Josefa Elisabeth Grossschedl<sup>1</sup>, João Alves<sup>1</sup>, Paula S. Teixeira<sup>1,2,3</sup>, Hervé Bouy<sup>4</sup>, Jan Forbrich<sup>1,5,6</sup>, Charles J. Lada<sup>6</sup>, Stefan Meingast<sup>1</sup>, Álvaro Hacar<sup>1,7</sup>, Joana Ascenso<sup>8</sup>, Christine Ackerl<sup>1</sup>, Birgit Hasenberger<sup>1</sup>, Rainer Köhler<sup>1</sup>, Karolina Kubiak<sup>1</sup>, Irati Larreina<sup>1</sup>, Lorenz Linhardt<sup>9</sup>, Marco Lombardi<sup>10</sup>, and Torsten Möller<sup>9</sup>

<sup>1</sup> Universität Wien, Institut für Astrophysik, Türkenschanzstrasse 17, 1180 Wien, Austria,  
e-mail: josefa.elisabeth.grossschedl@univie.ac.at

<sup>2</sup> Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona (IEEC-UB), Martí i Franquès 1, 08028 Barcelona, Spain

<sup>3</sup> Scottish Universities Physics Alliance (SUPA), School of Physics and Astronomy, University of St. Andrews, North Haugh, St. Andrews, Fife KY16 9SS, UK

<sup>4</sup> Laboratoire d'Astrophysique de Bordeaux, Université de Bordeaux, Allée Geoffroy Saint-Hilaire, CS 50023, 33615 PESSAC CEDEX, France

<sup>5</sup> Centre for Astrophysics Research, University of Hertfordshire, Hatfield AL10 9AB, Hertfordshire, UK

<sup>6</sup> Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge, MA 02138, USA

<sup>7</sup> Universidade do Porto, Departamento de Engenharia Física da Faculdade de Engenharia, Rua Dr. Roberto Frias, P-4200-465, Porto, Portugal

<sup>8</sup> Leiden Observatory, Leiden University, P. O. Box 9513, 2300-RA Leiden, The Netherlands

<sup>9</sup> Universität Wien, Fakultät für Informatik, Währinger Straße 29/S6, 1090 Wien, Austria

<sup>10</sup> University of Milan, Department of Physics, via Celoria 16, 20133 Milan, Italy

October 3, 2018

### ABSTRACT

We extend and refine the existing young stellar object (YSO) catalogs for the Orion A molecular cloud, the closest massive star-forming region to Earth. This updated catalog is driven by the large spatial coverage (18.3 deg<sup>2</sup>, ~950 pc<sup>2</sup>), seeing limited resolution (~0.7"), and sensitivity ( $K_s < 19$  mag) of the ESO-VISTA near-infrared survey of the Orion A cloud (VISION). Combined with archival mid- to far-infrared data, the VISTA data allow for a refined and more robust source selection. We estimate that among previously known protostars and pre-main-sequence stars with disks, source contamination levels (false positives) are at least ~7% and ~2.5% respectively, mostly due to background galaxies and nebulosities. We identify 274 new YSO candidates using VISTA/*Spitzer* based selections within previously analyzed regions, and VISTA/*WISE* based selections to add sources in the surroundings, beyond previously analyzed regions. The *WISE* selection method recovers about 59% of the known YSOs in Orion A's low-mass star-forming part L1641, which shows what can be achieved by the all-sky *WISE* survey in combination with deep near-infrared data in regions without the influence of massive stars. The new catalog contains 2978 YSOs, which were classified based on the de-reddened mid-infrared spectral index into 188 protostars, 184 flat-spectrum sources, and 2606 pre-main-sequence stars with circumstellar disks. We find a statistically significant difference in the spatial distribution of the three evolutionary classes with respect to regions of high dust column-density, confirming that flat-spectrum sources are at a younger evolutionary phase compared to Class IIs, and are not a sub-sample seen at particular viewing angles.

**Key words.** stars: formation - nebula: M42 - molecular cloud: L1641 - photometry: infrared - cluster: ONC

## 1. Introduction

It is well established that star formation takes place at the coldest and densest regions of molecular clouds. Although only in the last decades, with the development of infrared (IR) and millimeter facilities, it was possible to image the early stages of the star formation process. Describing the new observables, however, is not a straightforward task, and attempts of classifying young stellar objects (YSOs) (e.g., [Greene et al. 1994](#); [Evans et al. 2009](#)) and deriving an evolutionary path from a dense core to a YSO have been plagued with uncertainties. These are mostly due to the limited sensitivity and resolution of the observations and the intrinsic complexity of the star formation process. For example, objects of similar mass can have very different observables due to the large diversity of an YSO environment and

its geometry alone (e.g., [Whitney et al. 2013](#)). In other words, it is often difficult to establish an evolutionary stage for single sources. However, one can also look at entire populations to statistically infer evolutionary properties. This is now possible with the recent deployment of several space based and ground based IR telescopes that observed most nearby (< 500 pc) star-forming regions (e.g., [Evans et al. 2009](#); [Megeath et al. 2012](#); [Dunham et al. 2015](#)).

To understand and reconstruct the star formation process it is crucial to know the YSOs evolutionary stages. First attempts to classify YSOs into three evolutionary Classes (I, II, III) were presented in the 80s (e.g., [Lada & Wilking 1984](#); [Lada 1987](#)), based on the finding that dusty envelopes and circumstellar disks cause an IR excess. These Classes constitute a smooth evolutionary sequence according to the observed IR spectral energy distri-

bution (SED), where the spectral index  $\alpha$  was defined as a linear fit to the photometric near- (NIR) to mid-infrared (MIR) SED in log-space

$$\alpha = \frac{d \log(\lambda F_\lambda)}{d \log \lambda}, \quad (1)$$

used to estimate the evolutionary Stage<sup>1</sup> (e.g., Robitaille et al. 2006). In the 90s five YSO Classes were established (0, I, flat-spectrum, II, III) (e.g., Greene et al. 1994), which are thought to be connected to the true evolutionary Stage as follows: Class 0 sources (André et al. 1993) are protostars in the very early collapse phase with low black-body temperatures ( $T_{\text{bol}} < 70$  K), and with envelope masses that still dominating the system. They are mostly not detectable in the NIR or MIR and usually require observations at longer wavelengths. Class I YSOs ( $\alpha \gtrsim 0$ ) are protostars (P) which are still embedded and accreting material from a surrounding envelope onto a forming circumstellar disk. Class II YSOs ( $\alpha \lesssim 0$ ) are pre-main-sequence (PMS) stars surrounded by dusty circumstellar disks (D), which have dispersed their envelopes (also called T-Tauri stars). Finally, Class III YSOs are likely evolved PMS stars that emerge when accretion ends and the disks dissipate by stellar radiation or winds (e.g., Pillitteri et al. 2013). They show only very little ( $\alpha \lesssim -1.6$ ) or no IR-excess ( $\alpha \lesssim -2.5$ ). When using selection criteria based on IR photometry only the part of Class IIIs with IR-excess can be identified.

Greene et al. (1994) introduced the flat-spectrum class (hereafter also referred to as flats), lying between Classes I and II with  $\alpha \approx 0$ . This class represents YSOs not easily assignable to either protostars or disks<sup>2</sup>, and it is not clear if they are simply a mixture of or a transitional phase between these two. Therefore, Greene et al. assigned them an uncertain evolutionary status. There are several reasons for this. Firstly, the shape of the SED can be influenced by geometric effects, like disk inclination along the line of sight to the observer (Whitney et al. 2003b,a; Robitaille et al. 2006; Crapsi et al. 2008; Whitney et al. 2013), or by high foreground extinction (Muench et al. 2007; Forbrich et al. 2010). For example, an evolved protostar with an almost depleted envelope or viewed pole-on, and a Class II source where the disk is viewed edge-on or the source is seen through high extinction, may show a similar flat-spectrum SED (Whitney et al. 2003a). On the other hand, there are several studies suggesting a younger physical stage of flats compared to Class IIs. Muench et al. (2007) point out, that flat-spectrum sources are considered to be protostars in a later stage of envelope dispersal or with highly flared disks. Moreover, they find that flat-spectrum sources are intrinsically more luminous than Class IIs, suggesting a different evolutionary stage. Greene & Lada (2002), using NIR spectroscopy, find that accretion rates of flat-spectrum sources lie in between those of Classes I and II (inferred from the veiling excess), suggesting a transitional evolutionary stage. Recently, Furlan et al. (2016) found, based on SED modeling including FIR photometry, that the large majority of their studied sample of flat-spectrum sources require an envelope in their fit, indicating that these objects are still in the protostellar phase, covering different stages in their envelope evolution. At the same time, Carney et al. (2016) concluded from a molecular line study that about 30% of previously identified Class I sources were

more evolved Stage II YSOs. A similar situation was pointed out by Heiderman & Evans (2015), who find that only about 50% of flat-spectrum sources are surrounded by envelopes. Furlan et al. (2016) point out, the differences in their findings could be due to different methods to select flats. Indeed, different conventions do not provide easily comparable samples. The differences are driven by available photometry, the chosen spectral range to construct the spectral index, different class definitions, or even if extinction correction is applied or not (e.g. Lada 1987; Greene et al. 1994; Lada et al. 2006; Muench et al. 2007; Evans et al. 2009; Teixeira et al. 2012). Until all the points above are carefully addressed for a large statistical significant sample, the nature of flat-spectrum sources will remain undetermined.

The goal of this paper is twofold: a) construct the most complete catalog of dusty YSOs in the Orion A giant molecular cloud, and b) use it to infer on the nature of flat-spectrum sources. To achieve this we make use of the deep seeing-limited NIR VISTA photometry from the Vienna Survey In OriON (VISION, Meingast et al. 2016, hereafter, Paper I) to improve on previous YSO catalogs (Megeath et al. 2012, 2016; Furlan et al. 2016). Since the VISTA survey-area is larger than previously analyzed regions (Fig. 1), we will look for new YSO candidates in the surroundings, to improve the spatial completeness of this rich YSO sample. This we do in combination with MIR data from the Wide-field Infrared Survey Explorer (WISE, Wright et al. 2010) and the *Spitzer* Space Telescope (Werner et al. 2004), and with FIR data from the *Herschel* space observatory<sup>3</sup> (Pilbratt et al. 2010). Our analysis, using IR photometry, will not be sensitive to the majority of Class IIIs (see e.g. Pillitteri et al. 2013) and we ignore PMS stars without IR-excess in this paper. Future work should consider the whole young stellar population to determine the complete star-forming history within the cloud.

The structure of the paper is as follows. In Sect. 2 we describe the data and give a brief overview on recent Orion A YSO catalogs. In Sect. 3 we first present our methods to classify the YSOs (Sect. 3.1) and second, we discuss our methods to evaluate the contamination (false positives) of the known YSO population (Sect. 3.2). Example images of these are presented in Appendix A. Finally, we present our methods to select new YSO candidates (Sect. 3.3), with a detailed description of the selection methods given in Appendix B. We will classify the YSO candidates based on extinction corrected spectral indices into Class I, flat-spectrum, and Class II/III sources, with an overview of the resulting updated YSO sample presented in Sect. 4, and the corresponding table in Appendix C. In Sect. 5 we discuss the issues that come with YSO classification and we infer on the meaning of the flat-spectrum sources by looking at their spatial distribution with respect to regions of high dust-column density<sup>4</sup>. Finally, we give a summary in Sect. 6.

## 2. Data

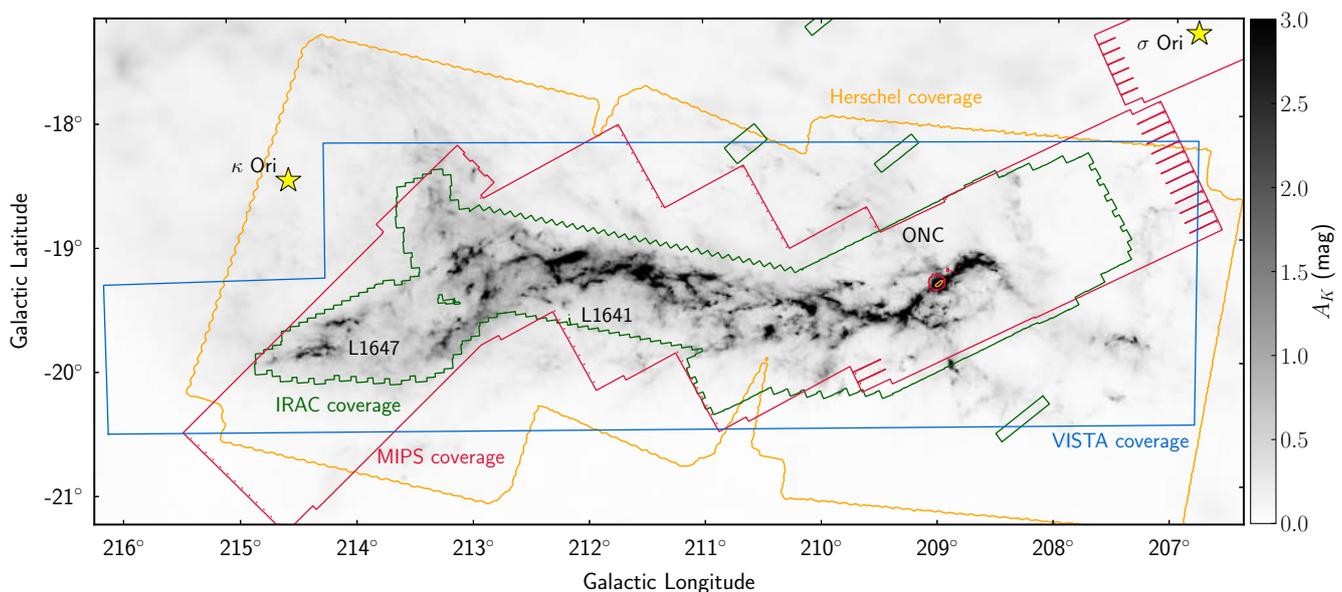
We use archival IR data and the new deep NIR VISTA data to re-examine the already studied YSO population in Orion A (Megeath et al. 2012, 2016; Furlan et al. 2016; Lewis & Lada 2016), and to select new YSO candidates in a larger field covered by VISTA. In Fig. 1 the footprints of the surveys used in this work are shown. The blue VISTA contour is the region investi-

<sup>1</sup> Class is used for the observed SED classification, while Stage refers to the physical configuration.

<sup>2</sup> YSO classes are also called for simplicity: Class 0/I - protostars (P), flat-spectrum sources - flats (F), and Class II/III - disks (D). Disks include Class IIs and the part of Class IIIs with IR-excess. See Tabel 3.

<sup>3</sup> *Herschel* is an ESA space observatory with science instruments provided by European-led Principal Investigator consortia and with important participation from NASA.

<sup>4</sup> Dust column-density, as traced by *Herschel*, is not directly tracing the dense gas. Therefore, the true volume density is unknown.



**Fig. 1.** Survey footprints displayed on top of the *Planck-Herschel-Extinction* dust column-density map (Lombardi et al. 2014). The gray scale shows the line of sight extinction, given in  $A_{K,\text{Herschel}}$  (mag). The contours show the survey coverages for VISTA (blue), *Spitzer*/IRAC1 (green), *Spitzer*/MIPS1 (red), and *Herschel* (orange). The small green boxes are control fields for the *Spitzer*/IRAC bands, which are partially overlapping with VISTA. The red and orange circles, located at the position of the ONC, result from saturated MIPS1 and *Herschel* photometry respectively. Indicated are the two main regions in Orion A, the Lynds dark cloud L1641 (including L1647), and the ONC region. The B-star  $\kappa$ -Ori and the O-star  $\sigma$ -Ori are marked for orientation. See also Figure 1 in Paper I for an overview of some sub-regions.

gated in this work ( $\sim 18.3 \text{ deg}^2$ ), while the green *Spitzer*/IRAC region ( $\sim 7 \text{ deg}^2$ ) was investigated by Meingast et al. (2012, 2016). This improved coverage allows for a spatially more complete sample. The background image is the *Planck-Herschel-Extinction* dust column-density map from Lombardi et al. (2014) (hereafter, *Herschel map*), with a resolution of  $36''$ . The *Herschel map* is used to estimate the total line of sight extinction at the position of the YSO candidates, to distinguish between regions of high and low dust column-density. The dust optical depth was converted by Lombardi et al. (2014) to extinction ( $A_K$ ) using a 2MASS<sup>5</sup> NIR extinction map (Lombardi et al. 2011). They find a linear conversion factor of  $\gamma = 2640$ . Based on a recent extinction map, constructed with VISTA and *Spitzer* data, we use an updated conversion factor of  $\gamma = 3050$  (Meingast et al. 2018, Paper II). Hereafter, we use the abbreviation  $A_{K,\text{Herschel}}$  when referring to extinctions extracted from the *Herschel map*.

### 2.1. VISTA near-infrared data

In the first paper of this series, introducing VISION (Paper I), we obtained deep NIR  $J$ ,  $H$ , and  $K_S$  photometry (see Table 1), covering the entire Orion A cloud, using the Visible and Infrared Survey Telescope for Astronomy (VISTA, Emerson et al. 2006) operated by the European Southern Observatory (ESO). We gain angular resolution and sensitivity compared to previous NIR surveys (e.g. 2MASS), reaching 90% completeness limits of 20.4, 19.9 and 19.0 mag in  $J$ ,  $H$ , and  $K_S$  respectively. The survey reaches a seeing limited resolution of almost  $0.6''$  (median seeing of  $0.72''$ ). Compared to 2MASS, the sensitivity of VISTA goes about 4 – 5 magnitudes deeper, and the resolution improved by about a factor of 3. Therefore, the VISION catalog contains about a factor of 10 more sources in the covered area ( $\sim 800,000$  point-sources). This allows for an improved YSO classification, and a better distinction of background galax-

ies or extended nebulous IR emission from YSO candidates. To estimate the colors and magnitudes of background and extragalactic contamination we use the VISTA control field observed during the survey, which is shifted about  $+22^\circ$  in galactic longitudes ( $l$ ) and lies at about the same galactic latitude ( $b$ ), covering  $\sim 1.8 \text{ deg}^2$  in the sky (10% of the science field coverage, see figure 4 in Paper I).

**Table 1.** Overview of the used photometric NIR and MIR bands.

Survey	Band	$\lambda^a$ ( $\mu\text{m}$ )	$F_{\nu,0}^b$ (Jy)	FWHM <sup>c</sup> ( $''$ )	$A_\lambda/A_{K_S}^d$
VISTA (1)	$J$	1.25	1594.0	0.78	2.50
	$H$	1.65	1024.0	0.75	1.55
	$K_S$	2.15	666.7	0.8	1.00
<i>Spitzer</i> IRAC & MIPS (2)	$I1$	3.6	280.9	1.66	0.64
	$I2$	4.5	179.7	1.72	0.56
	$I3$	5.8	115.0	1.88	0.50
	$I4$	8.0	64.9	1.98	0.51
	$M1$	24.0	7.17	6	0.45
<i>WISE</i> (3)	W1	3.4	309.540	6.1	0.79
	W2	4.6	171.787	6.4	0.55
	W3	12.0	31.674	6.5	0.61
	W4	22.0	8.363	12.0	0.43

**Notes.** <sup>(a)</sup> Central wavelength. <sup>(b)</sup> Zero magnitude flux density. <sup>(c)</sup> Mean image quality. <sup>(d)</sup> The extinction laws  $A_\lambda/A_{K_S}$  for VISTA and *Spitzer* are taken from Meingast et al. (2018), and for *WISE* they are provided by S. Meingast.

**References.** (1) Meingast et al. (2016); (2) IRAC Instrument Handbook (2015) <http://irsa.ipac.caltech.edu/data/SPITZER/docs/irac/iracinstrumenthandbook>; MIPS Instrument Handbook (2011) <http://irsa.ipac.caltech.edu/data/SPITZER/docs/mips/mipsinstrumenthandbook>; (3) Cutri et al. (2013)

<sup>5</sup> 2MASS - The 2 Micron All-Sky Survey (Skrutskie et al. 2006)

**Table 2.** Comparison of the MGM and FFA16 YSO classification.

	FFA16 (HOPS sources) <sup>d</sup>							Uncertain	not in FFA16 <sup>e</sup>
	ALL	Class 0	Class I	Flats	Class II	Galaxies			
	309 (278) <sup>c</sup>	60 (60)	103 (93)	104 (88)	16 (11)	22 (22)		4 (4)	
MGM Class <sup>b</sup>									
P	330	235 (223)	47	90 (87)	86 (77)	1 (1)	10	1	95
FP	49	13 (8)	2	3 (1)	4 (2)	1 (0)	3	—	36
RP	6	6 (5)	2	1 (0)	—	—	1	2	—
D	2442	39 (30)	1	6 (5)	13 (9)	14 (10)	5	—	2403
not in MGM <sup>d</sup> :	16 (12)	8	3 (0)	1 (0)	—	—	3	1	

**Notes.** In the middle the corresponding overlapping numbers of the two catalogs are listed. <sup>(a)</sup> The top row shows FFA16 classification, including their extra-galactic and uncertain candidates. <sup>(b)</sup> The first column shows MGM classification: protostar (P), faint protostar (FP), red protostar (RP), and disk candidates (D). <sup>(c)</sup> The first number are all listed HOPS sources in FFA16, and the second number in brackets are the sources where SED modeling was applied for sources with sufficient PACS photometry. <sup>(d)</sup> The last row lists the number of sources that are only in FFA16 and not in MGM. <sup>(e)</sup> The last column lists the number of sources that are only in MGM but not in FFA16.

## 2.2. Mid- to far-infrared data, and existing Orion A YSO catalogs

Megeath et al. (2012) have carried out a comprehensive study of the dusty young stellar population in Orion A, presenting a sample of 2818 YSO candidates with IR-excess. The catalog was slightly updated by Megeath et al. (2016) to a new sample of 2827 candidates, which we call hereafter simply MGM sample. They obtained *Spitzer* MIR photometry, using the Infra-Red Array Camera (IRAC, Fazio et al. 2004), and the Multiband Imaging Photometer for *Spitzer* (MIPS, Rieke et al. 2004) (see Table 1). The MGM selection is based on eight band color-color and color-magnitude diagram selections (including 2MASS), also described in Megeath et al. (2009); Gutermuth et al. (2009), and Kryukova et al. (2012). The 2827 YSO candidates are separated into protostars (P, RP, FP, 385) and disk dominated PMS stars (D, 2442), which roughly correspond to Class I and Class II YSO candidates. They give three sub-samples for protostars; the main protostar candidates (P), red protostar candidates (RP) with only a measurement in *M1*, and faint protostar candidates (FP), while the latter is a more unreliable sample (see Table 2, first column).

In addition, we use FIR data from the *Herschel* Photoconductor Array Camera and Spectrometer (PACS, Poglitsch et al. 2010) at 70, 100, and 160  $\mu\text{m}$ . *Herschel* observed Orion A during the *Herschel* Orion Protostar Survey (HOPS, e.g., Stanke et al. 2010; Fischer et al. 2010; Ali et al. 2010; Fischer et al. 2013; Manoj et al. 2013; Stutz et al. 2013; Tobin et al. 2015). Furlan et al. (2016) (hereafter, FFA16) discuss 309 of these HOPS sources in Orion A, with 293 (95%) sources being a sub-sample of the MGM YSOs. Considering the FFA16 sample as an update to MGM, there are 2817 YSO candidates in Orion A, classified into 60 Class 0, 234 Class I, 104 flat-spectrum, and 2419 Class II sources. An overview and comparison of the two catalogs is shown in Table 2. FFA16 classify the sources based on the bolometric temperature and the spectral index from 4 to 24  $\mu\text{m}$  ( $\alpha_{12M}$ ). They perform SED modeling to determine different stellar properties, by combining PACS with *Spitzer* photometry, *Spitzer*/IRS spectra, and APEX 350 and 870  $\mu\text{m}$  data (Stutz et al. 2013). However, modeling was only done for a sub-sample of 278 sources, due to limited PACS photometry for the rest. Out of the total 309 HOPS sources they classify 283 as YSO candidates and the remaining 26 as extra-galactic contamination or uncertain candidates (see Table 2). We further use the *Her-*

*schel*/PACS point-source catalog (HPPSC, Marton et al. 2017) to look for matches which are not in the HOPS catalog.

To select new YSO candidates in regions beyond *Spitzer*/IRAC (Sect. 3.3) we add MIR all-sky photometry from *WISE* (AllWISE data release, Cutri et al. 2013). *WISE* observed in four bands (see Table 1), with the sensitivity limits varying from about 17, 16, 11, to 7 mag for *W1-4*<sup>6</sup>. The wavelength coverage is similar to *Spitzer* (see Table 1 and figure 1 in Jarrett et al. 2011), especially for *W1/I1*, *W2/I2*, and *W4/M1*. The *W3* band covers a broader range around 12  $\mu\text{m}$  and overlaps with *I4* which is centred at 8  $\mu\text{m}$ . Both are influenced by PAH emission (polycyclic-aromatic-hydrocarbons), which is excited by UV radiation and emitted in the IR. Hence, typical sources of PAH emission are massive star-forming regions. This leads to higher contamination in these bands, especially near the Orion Nebula Cluster (ONC, e.g. Hillenbrand & Hartmann 1998; Lada et al. 2000). Also star-forming galaxies show PAH emission, which can be erroneously identified as YSOs, which will be addressed in Sect. 3.2. The lower resolution and sensitivity of *WISE* compared to *Spitzer* results in higher confusion caused by extended MIR emission. Especially the *W4* band is significantly contaminated by extended thermal emission, amplified by its low resolution.

## 2.3. Ancillary Data

Orion A is one of the most favorable sites to study star formation, being the closest massive star-forming region to earth ( $\sim 414$  pc Menten et al. 2007). Hence, there is a large number of studies and data available, especially for the prominent ONC region (see figure 1 in Paper I). The mentioned catalogs (MGM, FFA16) include members already reported in earlier smaller scale studies. To perform a more complete study, we add the following published datasets.

Several spectroscopic and optical surveys are available for the ONC region (Hillenbrand 1997; Hillenbrand & Carpenter 2000; Da Rio et al. 2009; Szegedi-Elek et al. 2013) and the dark cloud L1641 (Fang et al. 2009, 2013; Hsu et al. 2012, 2013; Da Rio et al. 2016). Spectroscopic surveys provide information on spectral types and on extinction, and allow classification into classical and weak-line T-Tauri stars (CTTS, WTTS). These are PMS stars showing typical emission (e.g.  $H\alpha$ ) or absorption (e.g.

<sup>6</sup> Given as *w?mpro* in the AllWISE catalog, abbreviated as *W?* in this work. The “?” is used as placeholder for 1, 2, 3, or 4.

Lithium, Li I) lines, which are indicators for youth. For example, strong  $H\alpha$  emission is caused by gas accretion onto the surface of the stellar photosphere. Accordingly, it probes the gaseous component of the circumstellar disk, while IR-excess probes the dusty component. Kim et al. (2013, 2016) present a study of transition disks (TD, Cieza et al. 2008, 2010; Muzerolle et al. 2010) for Orion A. These are circumstellar disks with inner dust holes filled with gas, and optically thick outer gas+dust disks (Teixeira et al. 2012). They show no or weak excess from 3 to 8  $\mu\text{m}$  (probing the inner disk) but a significant excess at longer wavelengths ( $\lambda \gtrsim 10 \mu\text{m}$ , probing the outer disk).

Moreover, we add optical data from the Sloan Digital Sky Survey DR12 (SDSS, Alam et al. 2015), which does not cover all of Orion A, but large parts near the ONC, and the Pan-STARRS survey (Flewelling et al. 2016), which covers the whole region. The optical data allows to construct more complete SEDs. These are helpful when investigating especially critical sources, with unclear classification.

To further confirm the young nature of stars we add X-ray observations from XMM-Newton and Chandra. XMM-Newton data is available for the L1641 region (Pillitteri et al. 2013) and the L1647 region ( $\kappa$  Ori, Pillitteri et al. 2016)<sup>7</sup>, and Chandra data is available for the ONC (COUP, Getman et al. 2005b,a) and for regions north and south to the ONC (SFINCS, Getman et al. 2017). This information is listed in the final catalog (Table C.1) with a simple flag “X”, which indicates if the source was detected in X-rays.

## 2.4. Combined data catalog

We combine the different data sets to one data catalog, adopting the cross-match radius to the resolution. First, VISTA is cross-matched with the whole *Spitzer* data catalog<sup>8</sup>, containing the MGM YSO sample. Second, data of the 309 HOPS (FFA16) sources are added, of which most are a sub-sample of the MGM catalog. Next, AllWISE MIR data is cross-matched. Due to the lower angular resolution of *WISE* ( $\sim 6''$ , *Spitzer*  $\sim 1.7''$ , VISTA  $\sim 0.7''$ ) multiple VISTA sources can lie inside one unresolved *WISE* source, which can lead to misidentifications. This can contribute to contamination and incompleteness of the final sample in ways that are difficult to characterize. This is addressed in Sect. 5.1 where we discuss the completeness of our final sample. Finally, all auxiliary data are added to complement the data catalog with the available information from the literature.

## 3. Methods

In this Section we present our methods to classify the YSO candidates, to revisit existing catalogs, and to add new candidates.

### 3.1. YSO classification

The YSO classification in this Paper is not solely based on a classical spectral index classification, but rather a combination of investigating various spectral index ranges, of including FIR information, visual inspection, and individual SED inspection. As an initial estimate, we adopt the YSO classification based on the MIR spectral index similar to Greene et al. (1994), as given

**Table 3.** Adopted YSO classification based on the MIR SED.

Class designation	A.d. <sup>a</sup>	Spectral Index
Class 0/I (protostars)	P	$+0.3 < \alpha$
flat-spectrum sources (flats)	F	$-0.3 \leq \alpha \leq +0.3$
Class II (thick disks)	D	$-1.6 < \alpha < -0.3$
Class III (anemic or thin disks)	D, AD	$-2.5 < \alpha \leq -1.6$
Class III (disk-less PMS) / MS stars <sup>b</sup>	III, MS	$\alpha \leq -2.5$

**Notes.** <sup>(a)</sup> Alternative designation. <sup>(b)</sup> Disk-less pre-main-sequence stars or main-sequence stars.

in Table 3 and we refine the classification by using the methods listed above.

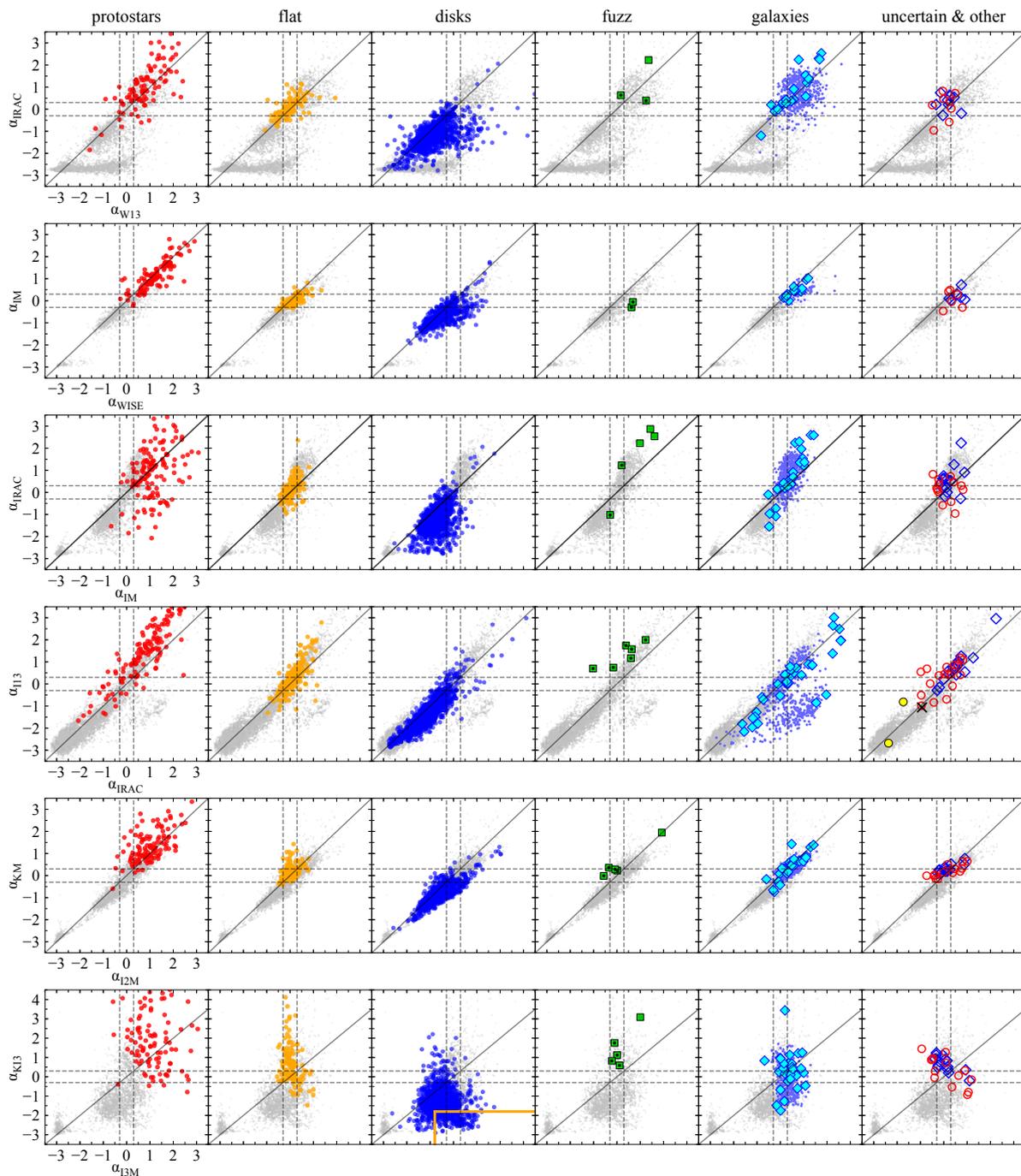
The lower spectral index limit for sources with IR-excess is given by Lada et al. (2006) with  $\alpha_{\text{IRAC}} > -2.56$ . Below, the SED reflects the photosphere of the star. They state that effects of different spectral types have no significant influence on this value, therefore it is an upper limit for sources with no IR-excess. We adopt a value of  $-2.5$ , due to the uncertainties in the *Spitzer* photometry, influencing especially sources near the ONC. Although the value is defined for  $\alpha_{\text{IRAC}}$ , it can also be applied to other spectral index ranges as an upper limit, which is highlighted in Fig. 2, where we compare various spectral indices. The scatter at the main-sequence (MS) star locus at about  $-3$  is caused by extincted sources.

Class IIIs, by the definition of Greene et al. (1994), include sources with weak IR-excess, due to optically thin disk remnants, also called anemic disks (AD, Lada et al. 2006). The distinction between Classes II and III was set due to findings of Andre & Montmerle (1994), where they find a sharp threshold in millimeter flux density at  $\alpha \approx -1.5$ . However, we do not separate the disk bearing YSOs, meaning Class IIIs and the part of Class IIIs with IR-excess, but call them collectively disks (D), when analyzing the sample in Sect. 5.3. Nevertheless, we label these sources separately in the final catalog with “D” and “AD”, respectively. In a similar manner, we do not distinguish between Classes I and 0, and call them collectively protostars (P). Again, we label these candidates separately in the final catalog with “I” and “0”, respectively, mainly based on the information from Stutz et al. (2013) and FFA16. Sources with flat-spectra likely correspond to YSO candidates with envelope remnants on the verge to the disk dominated PMS phase (FFA16), but it can also be an effect of disk inclination or foreground extinction, as highlighted in Sect. 1. Therefore, this class remains suspicious, and will be addressed in more detail in Sect. 5.2.

To calculate the spectral index we use all available photometry from 2 to 24  $\mu\text{m}$ . This range from the NIR to the MIR is often used to define the classes (e.g., Dunham et al. 2015; Heiderman & Evans 2015; Kim et al. 2016). However, as mentioned above, we also compare with other ranges like given in Table 4, while differences are highlighted in Fig. 2. For example, a comparison of  $\alpha_{\text{IRAC}}$  and  $\alpha_{\text{IM}}$  shows (Fig. 2, third row), when using only  $\alpha_{\text{IRAC}}$ , some protostars would be shifted to later classes. Furthermore, shorter wavelength ranges probe the inner disk, while longer wavelengths from about 8  $\mu\text{m}$  on-wards, probe the outer disk or envelope. This gives information on transition disk YSOs (TD). These sources can be misclassified as flats or even protostars, depending on the wavelength range used. They can be selected by using, for example,  $\alpha_{\text{K13}}$  and  $\alpha_{\text{I3M}}$ , highlighted at the bottom row of Fig. 2 by the orange lines over-plotted on the disk candidate plot (third column). Again, we do not separate TDs

<sup>7</sup> Download from <https://nxs.esac.esa.int>; the coordinates provided by Pillitteri et al. (2016) resulted in an erroneous cross-match.

<sup>8</sup> Available at: [http://astrol.physics.utoledo.edu/~megeath/Orion/The\\_Spitzer\\_Orion\\_Survey.html](http://astrol.physics.utoledo.edu/~megeath/Orion/The_Spitzer_Orion_Survey.html)



**Fig. 2.** Comparing various observed spectral indices ( $\alpha$ ). Shown are YSO candidates and false positives, with symbols like in Fig. 3. A slope one line is given (black solid line), highlighting where different spectral indices give the same value. The range for flat-spectrum sources is indicated by dashed lines. The two top rows show a comparison of *Spitzer* and *WISE* spectral indices, covering approximately the same spectral range. The protostars contain sources with declining  $\alpha$ ; we found that most of these are influenced by visible outflows. The used spectral indices are listed in Table 4. The solid orange outline (last row, third column) indicates a transition disk selection (see text for more explanations).

in our statistical analysis, but we label them in the final catalog separately as “TD”.

Not only the used spectral index but also foreground extinction can affect the classification by shifting, for example, Class II sources to the flat-spectrum or Class I regime (e.g. Muench et al. 2007; Forbrich et al. 2010), whereas longer MIR wavelength bands are less effected by extinction. Muench et al. (2007) point out that background stars can not mimic a Class I source when using  $\alpha_{I3M}$ , even at extinctions as high as  $A_K \approx 20$  mag (see their figure 20).

To correct for extinction we de-redden the photometry, by estimating the line of sight extinction towards each source individually, relative to the  $K_S$  band ( $A_K$ ). We denote the de-reddened spectral index as  $\alpha'$ . When available, we use literature values for line of sight extinctions obtained via spectral surveys (Hillenbrand 1997; Fang et al. 2009; Kim et al. 2013, 2016; Furlan et al. 2016). To convert from  $A_V$  to  $A_K$  we use  $A_K/A_V = 0.112$  from Rieke & Lebofsky (1985). Else we calculate the line of sight foreground extinction with the NICER technique (Near-Infrared Color Excess Revisited, Lombardi & Alves 2001), using the ex-

**Table 4.** List of used spectral indices.

Spectral indices	band range	$\lambda$ range ( $\mu\text{m}$ )	e.g. used by
$\alpha_{\text{KM}}$	$K_S$ to $M1$	2 - 24	Dunham et al. (2015) Heiderman & Evans (2015) Kim et al. (2016)
$\alpha_{\text{IM}}$	$I1$ to $M1$	3 - 24	-
$\alpha_{\text{IRAC}}$	$I1$ to $I4$	3 - 8	Lada et al. (2006)
$\alpha_{\text{I2M}}$	$I2$ to $M1$	4 - 24	FFA16
$\alpha_{\text{I3M}}$	$I3$ to $M1$	5 - 24	Muench et al. (2007)
$\alpha_{\text{KI3}}$	$K_S$ to $I3$	2 - 5	-
$\alpha_{\text{KW3}}$	$K_S$ to $W3$	2 - 12	-
$\alpha_{\text{KW}}$	$K_S$ to $W4$	2 - 22	-
$\alpha_{\text{W13}}$	$W1$ to $W3$	3 - 12	-
$\alpha_{\text{WISE}}$	$W1$ to $W4$	3 - 22	-
$\alpha_{\text{KW12M}}$	$K_S$ to $M1$	2 - 24	-

**Notes.** The top block shows spectral indices including VISTA and *Spitzer* bands, and the bottom block VISTA and *WISE* bands. All available bands are used between the individually given ranges, however, *Spitzer* and *WISE* are not mixed, except for the last spectral index, where we combine  $K_S$  with  $W12$  and  $M1$ , used for sources beyond the IRAC coverage but still inside the MIPS coverage.

tionction laws listed in Table 1. The method uses NIR JHK photometry, with the intrinsic color derived from the VISTA control field. In our case the intrinsic color corresponds to the location of M-stars. If not all of the three NIR bands have valid measurements we use only two NIR bands for an estimate (i.e.,  $E(J-H)$ ,  $E(J-K_S)$ , or  $E(H-K_S)$ ). For sources with only one or no NIR observation, we use the new PNICER technique (Meingast et al. 2017), which is a probabilistic machine learning approach, enabling the inclusion of MIR bands, to estimate extinction. The finally used method (to determine extinction) is given in Table C.1 in column “AK\_Method”. For YSOs or galaxies the calculated line-of-sight extinction might overestimate the actual foreground extinction, due to the intrinsic reddening by circumstellar material, and galaxies show redder colors also due to dust and star formation. We compare the individual line-of-sight extinction with the total line of sight extinction at the position of each source as extracted from the *Herschel* map ( $A_{\text{K,Herschel}}$ ), and find that especially for galaxies, extinctions estimated from their IR colors ( $A_{\text{K,IR}}$ ) are mostly larger than  $A_{\text{K,Herschel}}$ . Therefore, we use  $A_{\text{K,Herschel}}$  if  $(A_{\text{K,Herschel}} + \text{err}_{A_{\text{K,Herschel}}}) < A_{\text{K,IR}}$  to de-redden such sources<sup>9</sup>.

Unfortunately, we can not use the same set of spectral indices for all sources due to different survey coverages and different sensitivities of the various bands. For sources lying outside the IRAC covered region, thus new YSO candidates (Sect. 3.3), we use *WISE* spectral indices, or a combination with VISTA or *Spitzer*/ $M1$  (see Table 4). This leads to an inconsistent classification, as highlighted in Fig. 2. The spectral index  $\alpha_{\text{W13}}$  is significantly different to  $\alpha_{\text{IRAC}}$ , where contaminated  $W3$  photometry (PAH emission) produces a shift of MS stars (or Class IIIs) to redder colors at the bottom of the diagram. The spectral index  $\alpha_{\text{WISE}}$  shows a better correlation with  $\alpha_{\text{IM}}$ , but not all  $M1$  observed sources also show a significant  $W4$  measurement due to the inferior sensitivity of *WISE*. Nevertheless, newly selected candidates beyond the *Spitzer* coverage are less affected by contamination and therefore have overall more reliable *WISE* photometry.

<sup>9</sup> The error of  $A_{\text{K,Herschel}}$  is extracted from the *Herschel* error map (Lombardi et al. 2014).

To summarize the classification process, for the final YSO classification we do not use the MIR spectral index blindly, by strictly following a simple cut using a single spectral index range. Instead, we look at different spectral indices, as listed in Table 4, and check if they consistently show a rising, flat, or declining slope, and compare with the de-reddened spectral indices  $\alpha'$ . For border cases, especially close to the flat-spectrum range, with no clear trend for the different spectral indices, we individually check the SEDs to make a final decision, and investigate the FIR range, basically adopting the classification by FFA16. Due to considering different spectral index ranges, the flat-spectrum sources in Fig. 2 (showing the observed spectral index  $\alpha$ ), do not fall exactly in the range of  $-0.3 < \alpha < 0.3$  for all indices.

Additionally, we use visual information, as also discussed below (Sect. 3.2). The VISTA images reveal outflows, cavities, jets, and reflection nebulae, which were taken into account as confirmation for the protostellar nature. For example, jets and outflow shocks close to the source affect the 4.5  $\mu\text{m}$  range ( $I2$  or  $W2$ ) (Evans et al. 2009), and reflected light in the outflow cavity can affect NIR bands (Crapsi et al. 2008). Moreover, the silicate absorption feature, located at about 10  $\mu\text{m}$  ( $I4$  or  $W3$ ), is caused by protostellar envelopes or edge-on disks (Crapsi et al. 2008), or by layers of high column-density in front of the source. Therefore, many protostars do not show a rising  $\alpha_{\text{IRAC}}$ , while clearly rising when including  $M1$  (Fig. 2). The protostellar nature is also clarified with FIR data, because most protostars correspond to a PACS point sources (FFA16), and show a clear peak in the FIR, which is not reflected in the MIR spectral index.

Finally, contaminated photometry can produce a fake IR-excess, which is not always easy to account for. Especially high-mass star-forming regions can cause a lot of such contamination (image artifacts, saturation, nebulosities, extended emission, cloud-edges). We exclude bands if their photometry where found to result from contamination during visual inspection.

### 3.2. Revisiting the known YSO population - Methods to evaluate false positives

The *Spitzer* based MGM catalog is currently the reference for the YSO population in Orion A, with updates from FFA16 and Lewis & Lada (2016) (hereafter, LL16). In this section we present a combination of methods to evaluate the contamination (false positives) of existing YSO catalogs. These are visual inspection, position of sources in color-color and color-magnitude diagrams, effects of extinction, source morphology (extension flags) from VISTA, and information from the literature.

- 1. Visual Inspection.** All previously identified YSO candidates were visually inspected, using the VISTA images (Paper I), and images of *Spitzer*/IRAC/MIPS, *WISE*, 2MASS, *Herschel*/PACS, DSS, and SDSS<sup>10</sup>. This enables us to identify resolved galaxies (G), IR nebulosities (fuzz), or image artifacts like diffraction-spikes or airy-rings (“C” for contamination). The abbreviations given in parenthesis are also used in the final catalog (Table C.1) in column “Class”. Visual inspection can further be used as a confirmation for protostars, which show visible outflows, jets, reflection nebulae, or cavities. Examples of these can be found in Appendix A.
- 2. Color-color and color-magnitude diagrams.** In parallel to visual inspection, various color-color (CCD) and color-magnitude diagrams (CMD) are checked, similar to those

<sup>10</sup> VISTA, *WISE*, 2MASS, PACS, DSS, SDSS images are available via Aladin (Bonnarel et al. 2000; Boch & Fernique 2014).

used by [Gutermuth et al. \(2009\)](#) and [Megeath et al. \(2012\)](#). We specifically check if the color of each source is consistent with the typical color of the proposed class. The diagrams include photometry in the NIR and MIR from VISTA and *Spitzer*. We show examples in Figs. 3 and 4 separately for the three different YSO classes (protostars, flats, disks, as classified in Sect. 3.1), and for false positives (fuzz, galaxies, uncertain and other objects). In the fifth column, showing galaxies including false positives, one can see that YSOs and some types of galaxies occupy similar color spaces in most diagrams. Here, the false-positive YSO identifications are typically found toward the bright end of the galaxy locus.

3. **Extinction.** As an additional indicator, the total column-density toward single sources can be used. For example, it is unlikely that protostars are associated with low extinction regions since they are still surrounded by their dust envelope. As investigated by [Lada et al. \(2010\)](#), a typical star formation extinction threshold is  $A_K > 0.8$  mag. We use the *Herschel* map to infer the total dust column-density towards each source ( $A_{K,Herschel}$ ). However, this needs to be handled carefully, since we can not rule out the presence of unresolved structure beyond the resolution of *Herschel* ( $36''$ ,  $\sim 0.07$  pc @  $414$  pc). Therefore, if no or only little extinction is located at the position of a candidate protostar, the source is further investigated (see also, [LL16](#)), to look for other signs of youth (e.g. outflows, PACS detection). Else it is flagged as uncertain (“U”). On the other hand, if a source is above the adopted threshold, does not immediately confirm its YSO nature. For example, bright galaxies can be detectable through extinction as high as  $A_K \sim 2$  mag. Finally, disk sources do not necessarily have to be connected to regions of higher dust column-density. During their typical age of a few million years (e.g., [Evans et al. 2009](#); [Dunham et al. 2015](#)) they could have already moved away from their birthplace, or the clouds out of which they have formed might have dissipated ([LL16](#)).
4. **Source morphology.** The VISTA source catalog provides two extension flags. ClassSex refers to a source’s morphology as determined by the source extraction algorithm SExtractor, while ClassCog derives the morphology from variable aperture photometry in combination with machine learning techniques (for details see [Paper I](#)). Values close to 0 indicate an extended object, values close to 1 point-like morphology. These flags, however, are not a universal discriminator between galaxies and stars, because protostars are often associated with extended emission or outflows. For this reason, we always use these flags in combination with visual inspection. While many galaxies are associated with extended morphology, faint extra-galactic objects can also appear point-like. These, however, are mostly identified in the various CCDs and CMDs. Special cases are active galactic nuclei (AGNs), which might be more difficult to distinguish, since they can appear more bright and point-like, while also showing similar IR colors as protostars or flat-spectrum sources. These can contribute to residual contamination in the final catalog.
5. **Additional surveys.** We searched the SIMBAD astronomical databases ([Wenger et al. 2000](#)) and the literature (see Sect. 2.3) for additional classification information. Oftentimes, young stars are already marked as emission line stars (Em\*), flare stars (Flare\*), variable stars (V\*, Orion\_V\*, Irregular\_V\*), or T-Tauri stars (TTau\*, WTTS, CTTS). Since this information is very heterogeneous, we generally do not use it for our classification. Only suspicious sources (faint, unresolved, untypical colors), which do not have an

entry in these additional surveys, are marked as uncertain candidates (U). We include this information in the final catalog (Table C.1).

To summarize, galaxies (“G”) are identified morphologically using visual inspection and extension flags in combination with colors, magnitudes, and information about extinction. If no clear morphological identification is possible, we flag some sources as uncertain galaxy candidates (“UG”), if their colors, magnitudes, and location at low extinction suggest the extra-galactic nature. Hence, they belong to the uncertain candidates. Fuzzy contamination, like nebulosities, cloud-edges, or Herbig-Haro objects, are generally identified visually (“fuzz”), as well as photometric contamination like image artifacts (“C”).

### 3.3. New YSO candidates

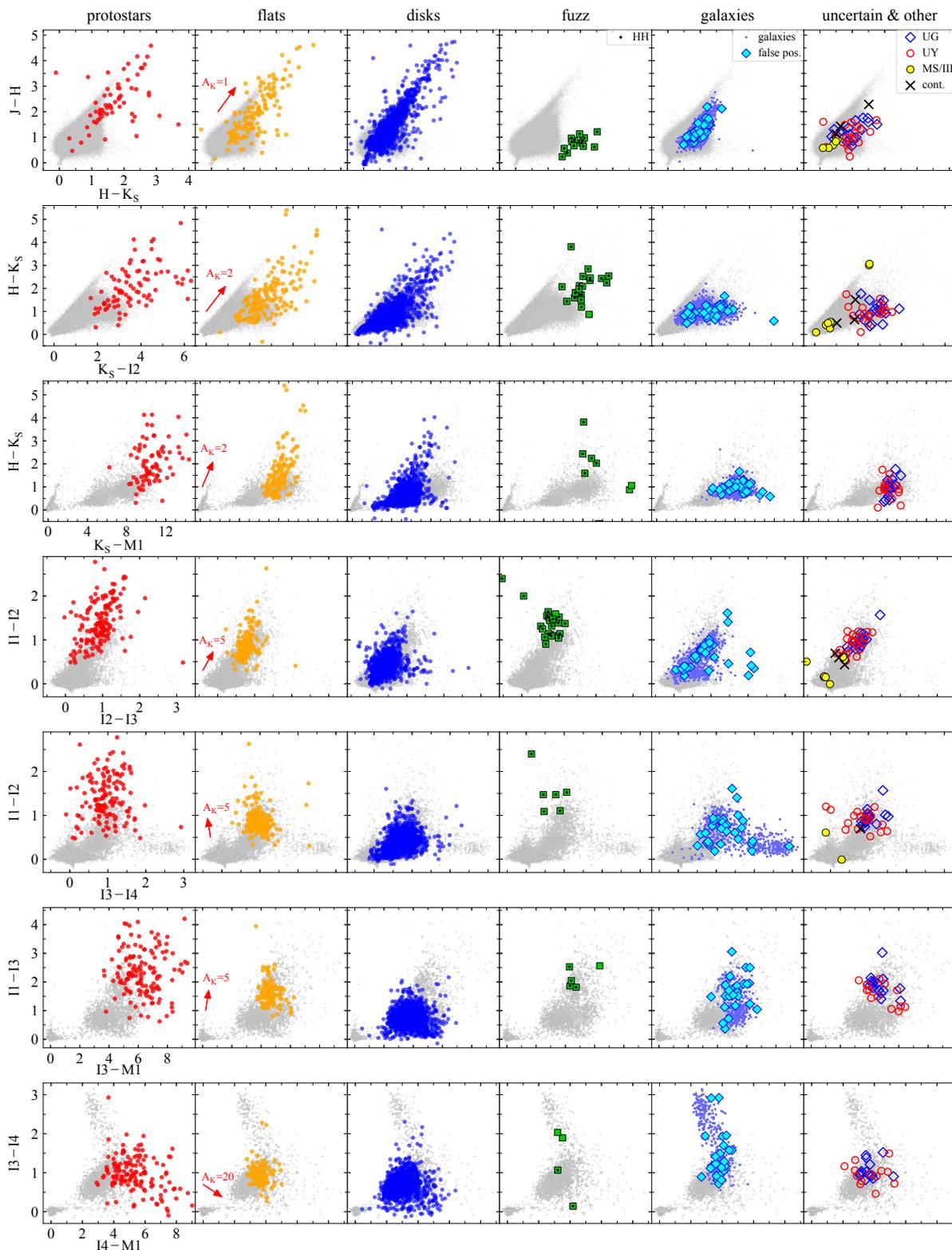
Here we shortly describe our methods to add new YSO candidates, while the detailed procedure is described in Appendix B. The methods are mainly based on NIR and MIR color-color and color-magnitude diagram selection criteria, and we also add some few sources using PACS photometry. To add new YSO candidates in the surroundings of the *Spitzer*/IRAC surveyed region (outside IRAC regions, green contour, Fig. 1), we make use of the larger coverage of VISTA (blue contour). To this end, we construct color and magnitude diagrams using VISTA combined with *WISE* and partially *Spitzer*/M1 (red contour). *WISE* requires special treatment, especially concerning the two longer wavelength bands *W3* and *W4*, due to the low resolution and high contamination caused by extended MIR emission, already highlighted in Sect. 2.2. The selection conditions for *WISE* data are informed by previous works ([Jarrett et al. 2011](#); [Rebull et al. 2011](#); [Koenig et al. 2012](#); [Koenig & Leisawitz 2014](#); [Koenig et al. 2015](#)), however we adjust them for our purpose (see Appendix B.2). Moreover, we select new YSO candidates also inside the IRAC region in combination with *Spitzer* photometry, by applying different selection criteria compared to previous studies, and by including VISTA instead of 2MASS. Additionally, we use the PACS point source catalog and visual inspection of the PACS images, to add further new YSO candidates.

## 4. Results

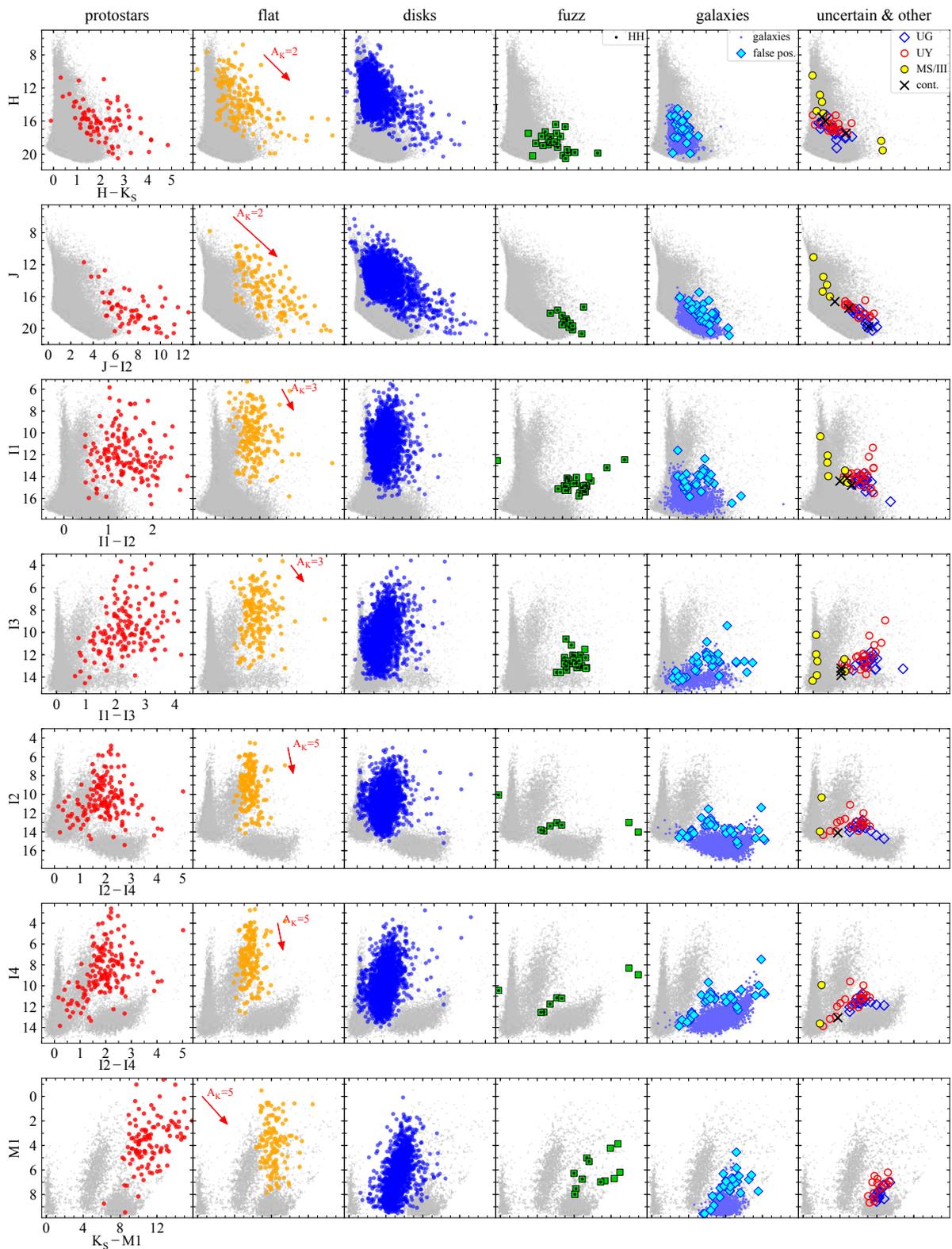
In this section we first summarize our results for the revisited catalogs, second, the new YSO candidates, and finally we give an overview of the updated YSO catalog.

### 4.1. Results for revisited YSO candidates

With the methods described above we revisit the 2839 previously identified YSO candidates from MGM (2827) and [FFA16](#) (283), resulting in 2704 ( $\sim 95\%$ ) YSO candidates in the updated catalog (Table C.1). We re-classify them as described in Sect. 3.1, with the resulting number-counts for each Class listed in Table 5. Out of the 135 ( $\sim 5\%$ ) excluded candidates, there are 92 ( $\sim 3\%$ ) false positives and 43 ( $\sim 2\%$ ) uncertain candidates. Most of the uncertain sources are faint objects with untypical properties (colors, magnitudes, location), for which we can not tell with our criteria and the available data, if they are faint YSOs, extra-galactic (e.g. AGNs), background giants (e.g. AGBs), or even brown dwarfs. Follow up observations are needed to clarify the nature of these sources, like spectra (e.g. NIR spectra, [Greene & Lada 2002](#)), or looking for envelope tracers (e.g. HCO<sup>+</sup>, [Heiderman & Evans 2015](#)). We still expect a residual degree of contamination in our

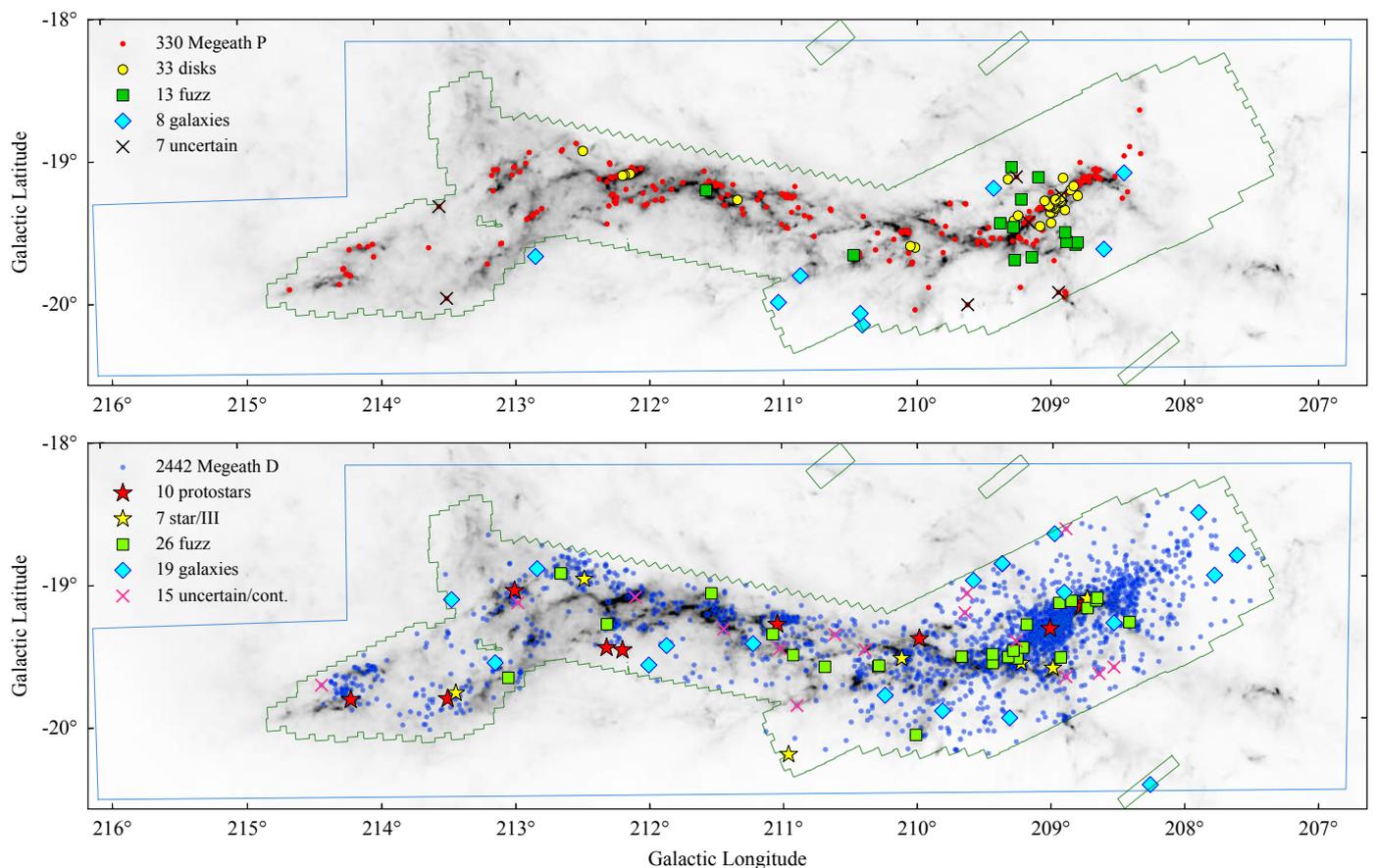


**Fig. 3.** Seven selected color-color diagrams, showing YSO candidates and false positives. Gray dots represent the background, which are all sources with errors  $< 0.2$  mag for the given bands. From left to right we show the three YSO classes, protostars (red), flat-spectrum sources (orange), disk sources (blue); and contaminating objects: fuzzy nebulous contamination (green), galaxy contamination (cyan), and finally other contamination or uncertain objects, which are MS stars or Class III candidates (filled yellow circles), contamination due to image artifact (black crosses), uncertain candidates (red open circles), and uncertain galaxy candidates (blue open diamonds). In column four showing fuzzy contamination, we highlight Herbig-Haro objects with a black dot, while sources without dot are mostly cloud-edges or other nebulous structures. In column five we show galaxies (dot symbols) and previous YSO candidates identified as galaxies (false positives, filled cyan diamonds). This highlights that galaxies often occupy similar color spaces as YSOs, especially the ones previously classified as YSO candidates. However, most tend to be fainter for the distance of Orion A (see CMDs, Fig. 4).



**Fig. 4.** Seven selected color-magnitude diagrams, showing YSO candidates and false positives, as in Fig. 3. The CMDs highlight the brightness differences of the classes and contaminating objects. False positives (fuzz, galaxies) and uncertain sources tend to be fainter. Galaxies, previously identified as YSO candidates (false pos., cyan diamonds) are generally brighter than average galaxies (bluish dots, other identified galaxies). Looking at YSO candidates, flats tend to be overall brighter in the MIR compared to disk candidates.

final selection mainly due to AGNs or AGBs. AGNs especially influence the flat and protostar range (Stern et al. 2005), and AGBs the anemic disk (Class III) range (Dunham et al. 2015). The false positives include six sources that do not show any IR-



**Fig. 5.** Distribution of the MGM YSO candidates showing false positives and re-classification, without considering flat-spectrum sources, displayed on the *Herschel* map. Top: The 330 MGM protostar candidates (P, red dots), of which 33 are re-classified as disk candidates (yellow filled circles). Bottom: The 2442 MGM disk candidates (D, blue dots), of which 10 are re-classified as protostar candidates (red filled stars), and 7 as MS stars or Class III sources (yellow filled stars). Other symbols are false positives, like extra-galactic contamination (cyan filled diamonds), and fuzzy nebulous contamination (green filled squares) as given in the legend. The “x” symbol marks uncertain sources and sources contaminated by image artifacts (“uncertain/cont.”).

excess beside some reddening effects due to extinction. These are flagged as “star” or Class III candidate (“III”, if X-ray source or emission line star, see Sect. 2.3). Statistical overviews of the different types of contamination are listed for the MGM and FFA16 samples in Tables 6 and 7, respectively. For the whole MGM sample we get a lower limit of contamination of about 3% to 5%, while for the FFA16 sample (with applied SED modeling, first row Tab. 7) we get only a contamination fraction of < 1%.

In Figure 5 we show the location of the MGM YSO candidates and the proposed false positives, since this sample is the most used reference up to date. The top map shows the 330 more reliable MGM protostar candidates (P), and the bottom shows the 2442 MGM disk candidates (D). Many of the more scattered<sup>11</sup> previously found protostar candidates at low dust column-density turned out to be extra-galactic contamination, or are uncertain sources (see also LL16). Contamination due to MIR nebulosities tends to be located close to the ONC region, as expected, due to higher contamination caused by the bright nebula. We compare our findings with the contamination estimates discussed in Megeath et al. (2012). For the region inside the IRAC coverage ( $\sim 7$  deg<sup>2</sup>) they expect about 43 false positives due to extra-galactic contamination. We find 37 galaxies (G) and

18 galaxy candidates (UG) in this region. There are further 25 uncertain YSO candidates, which might also be of extra-galactic nature. For the sub-samples D, P, and FP, Megeath et al. (2012) estimate about  $\sim 11$ ,  $\sim 20$ , and  $\sim 13$  extra-galactic contaminants, respectively. Inside their given errors this corresponds roughly to

**Table 5.** Re-classification summarized.

Sample	YSOs <sup>b</sup>	YSO Classes <sup>a</sup>		
		P	F	D
Revisited <sup>c</sup>	2704	182 (6.7%)	176 (6.5%)	2346 (86.8%)
New inside <sup>d</sup>	154	1 (0.7%)	3 (1.9%)	150 (97.4%)
New outside <sup>d</sup>	120	5 (4.2%)	5 (4.2%)	110 (91.7%)
New all <sup>d</sup>	274	6 (2.2%)	8 (2.9%)	260 (94.9%)
Total <sup>c</sup>	2978	188 (6.3%)	184 (6.2%)	2606 (87.5%)

**Notes.** The percentages in parenthesis are relative to the YSO counts of each sample given in Col. 2. <sup>(a)</sup> Classification from this work. Class 0/I protostars (P), flat-spectrum sources (F), Class II/III pre-main-sequences stars with disks (D). <sup>(b)</sup> Total number of YSO candidates of the given samples. <sup>(c)</sup> Reclassification of revisited sources for the combined MGM and FFA16 sample (Sects. 3.2, 4.1). <sup>(d)</sup> Classification for new sources (Sects. 3.3, 4.2), separated between candidates selected in and outside the IRAC region. <sup>(e)</sup> Total = Revisited + New all

<sup>11</sup> i.e. not connected to regions of high dust column-density, or less clustered environments.

**Table 6.** The MGM YSO sample revisited in numbers.

MGM		This work												
Class	Nr	YSOs				False positives <sup>b</sup>				Uncertain <sup>c</sup>		Total contamination <sup>d</sup>		
		All <sup>a</sup>	P	F	D	Galaxies	Fuzz	MS/III	Artifacts	UG	UY	f.p.	f.p.+U	
All	2827	2695 (95.3%)	174	175	2346	37	44	7	4	18	22	92 (3.3%)	132 (4.7%)	
D	2442	2375 (97.3%)	10	58	2307	19	26	7	3	—	12	55 (2.3%)	67 (2.7%)	
P	330	302 (91.5%)	159	110	33	8	13	—	—	2	5	21 (6.4%)	28 (8.5%)	
FP	49	15 (30.6%)	2	7	6	10	4	—	—	15	5	14 (28.6%)	34 (69.4%)	
RP	6	3 (50.0%)	3	—	—	—	1	—	1	1	—	2 (33.3%)	3 (50.0%)	
P,FP,RP	385	320 (83.1%)	164	117	39	18	18	—	1	18	10	37 (9.6%)	65 (16.9%)	

**Notes.** Shown is the re-classification and proposed contamination (false positives) from this work. Uncertain sources are given separately. <sup>(a)</sup> The total number of remaining YSO candidates is the sum of the three classes (P, F, D). <sup>(b)</sup> Different types of false positives: Galaxies, Nebulosities (Fuzz), main-sequence stars or Class III candidates (MS/III), and contamination from image artifacts (C). <sup>(c)</sup> The number of uncertain objects are given separately for uncertain galaxy candidates (UG), and uncertain YSO candidates (UY). <sup>(d)</sup> Summarized contamination, giving a lower and upper limit based on the sum of false positives (f.p.) and the sum when including uncertain candidates (f.p.+U)

**Table 7.** The FFA16 HOPS sample revisited in numbers. Similar to Table 6.

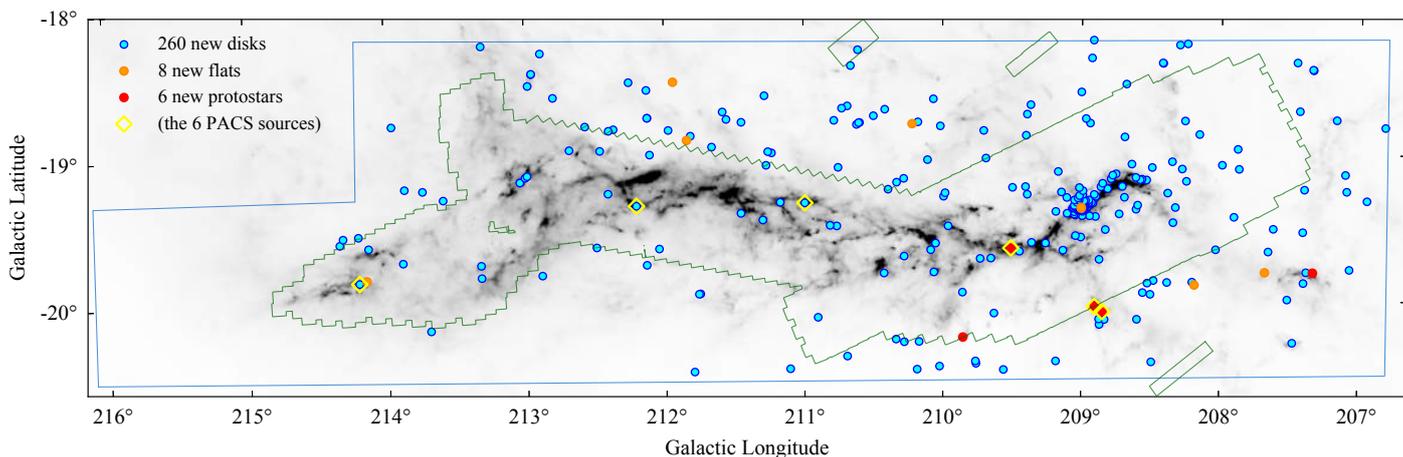
FFA16			This work							
Type	Class	Nr.	YSOs				False positives			Uncertain
			All	P	F	D	Galaxies	Fuzz	Artifacts	UY
YSOs	Modeled	252	250	149	83	18	1	1	—	—
YSOs	All	283	272	151	93	28	5	2	—	4
	Class 0	60	60	57	3	—	—	—	—	—
	Class I	103	93	83	6	4	4	2	—	4
	Flat	104	103	11	80	12	1	—	—	—
	Class II	16	16	—	4	12	—	—	—	—
Other	Galaxies	22	10	2	5	3	6	2	1	3
	Uncertain	4	1	1	—	—	—	1	1	1

**Table 8.** The 44 low  $A_K$  MGM protostar candidates revisited by LL16, compared to our results. Similar to Tables 6 and 7.

LL16		This work						
Type	Nr.	YSOs			False positives		Uncertain	
		P	F	D	Galaxies	Fuzz	UY	
all MGM low- $A_K$	P	44	1	13	8	8	10	4
YSOs	Stage I	10	1	4	2	1	—	2
	Stage II	18	—	9	5	3	—	1
Other	Galaxies	4	—	—	—	4	—	—
	Fuzz	9	—	—	—	—	9	—
	Uncertain	3	—	—	1	—	1	1

the 19, 8 (+2 UG), and 10 (+16 UG), that we found for each sub-sample. These are still lower limits, since, as already mentioned, remaining contamination by point-like AGNs can not be ruled out entirely. However, considering that our findings correspond well with the MGM contamination estimates, or tend to be even slightly above, these are likely a negligible fraction. Considering contamination due to nebulosities in the MIR, we get about 3.9% and 1.1% in the MGM P and D samples, respectively. Although Megeath et al. (2012) estimated it to be a more negligible fraction. This fact highlights the unfortunate sensitivity to point-like outflow knots and cloud edges of MIR observations, influencing especially a protostar sample. However, Figure 3 shows that for example, Herbig-Haro objects often show distinct colors in the NIR (first row, forth column) and in some MIR regimes. Hence, a careful color selection can mitigate at least some of these contaminants.

Furthermore, we compare our findings to the results in LL16 (Table 8), who revisited a sub-sample of 44 MGM protostars (P) that are located at low dust column-density ( $A_K < 0.8$  mag). These 44 sources are of interest to test the assumption that protostars (or star-formation) are connected to a certain extinction threshold (Lada et al. 2010). LL16 concluded that 10 out of the 44 low- $A_K$  MGM Ps are likely Stage I candidates based on SED modeling (Robitaille et al. 2006), and discuss scenarios to explain the absence of significant dust at the location of these sources, including source migration/ejection, and dust dissipation due to protostellar outflows. They use the same *Herschel map* (Lombardi et al. 2014) to estimate the extinction at the position of each YSO. However, based on the updated conversion factor from optical depth to  $A_K$  (see Sect. 2), the number of MGM protostars below the extinction threshold changes to 42. One of the two sources, which are now above the threshold, was classified as fuzz (MGM 1286), and the other as Stage I protostar



**Fig. 6.** The distribution of the 274 new YSO candidates selected in this work. Most sources were classified as new disk candidates (blue), and only few as flat-spectrum (orange) and protostar (red) candidates (see legend). The six extra PACS sources (3 P, 3 TD, Sect. 4.2), are additionally highlighted with yellow open diamonds. Sources beyond the green IRAC coverage are new VISTA/WISE or VISTA/M1 selected YSO candidates, located in areas not included in previous studies. **Notes.** There are only 5 red dots visible, marking the 6 new protostar candidates, because two of them are lying very close to each other, so the dots are blended.

(MGM 333) by LL16. The latter was classified as flat by FFA16. However, we classify it as Class II candidate, due to the declining spectral index when including the K-band. Also it lacks a PACS counterpart and is visible in the optical (Pan-STARRS  $g = 20.6$  mag). Moreover, it was classified as transition-disk candidate in Kim et al. (2013, 2016), which explains the flattish  $\alpha$  in the MIR. In total, we find only one reliable protostar candidate among their rest 9 Stage I sources (Table 8). Five more are likely more evolved YSO candidates (1 disk, 4 flats), one is a galaxy, and two are uncertain sources, which need more investigation, to test the theory of an extinction threshold correctly. The rest 34 MGM protostar candidates are re-classified by LL16 into 18 Stage II candidates, and 16 false positives or uncertain candidates. We confirm 14 of these Stage II sources as YSO candidates (5 disks, 9 flats) and we re-include one of their uncertain sources as Class II candidate.

Most of the FFA16 HOPS YSOs we confirm as reliable candidates ( $\sim 96.1\%$ ), especially those where SED modelling was possible ( $\sim 99.2\%$ , see also Table 7). FFA16 list 22 galaxy candidates, of which 19 are MGM YSO candidates, where we find that 10 out of these are likely good YSO candidates. One reason for this misidentification could be insufficient data quality. For example, if some bands are contaminated by artifacts or extended emission the SED might not fit to any YSO model. Also spectra in star-forming regions, which are affected by PAH emission, can be a combination of the YSO plus the nebulous surroundings, which can produce similar spectra as star-forming galaxies. Indeed, most of these sources are near regions of extended emission, and are at the same time associated with high extinction regions. This makes it unlikely that these are background galaxies, also given the fact that they are well visible in the NIR. Moreover, there is one modelled FFA16 Class I candidate, and four modelled LL16 YSO candidates (one Stage I, three Stage II), which we identify as resolved galaxies from visual inspection of the VISTA images. These findings are of interest, as they show that modelling alone is not always reliably separating YSO candidates from extra-galactic contamination, which was also pointed out by Evans et al. (2009) and Furlan et al. (2016).

#### 4.2. Results for new YSO candidates

With the color based NIR and MIR selection criteria (see Appendix B) we are able to add 268 new YSO candidates inside the whole VISTA coverage. Separating selections from inside (VISTA/Spitzer) and outside (VISTA/WISE/M1)<sup>12</sup> the IRAC region, we select 152 and 117 new YSO candidates, respectively.

We add further six YSO candidates by using the PACS point source catalog and PACS images. Two of these are new protostar candidates, located at a prominent young clustering, south-west of the ONC (Har04-145 cluster, see Appendix A and Fig. A.1), of which one is likely a new undiscovered Class 0 protostar (ID 116363), not yet discussed in previous works. Furthermore, we add another new protostar candidate (ID 213612) - detected during visual inspection - inside the IRAC region right next to a known Class 0 source (MGM 1121, separation  $\sim 5''$ ). This new candidate shows a prominent outflow cavity in the NIR. Both, the Class 0 and the new candidate, lie on top of an elongated PACS source, and are also highlighted by Tobin (2017) as protostar binary candidate. Finally, we add three transition disk candidates (ID 377204, 459841, 522530). These sources show no NIR or MIR excess, but a clear PACS excess, indicating an outer disk. Visually they seem to be surrounded by reflection nebulae in the NIR. See also Appendix A, Fig. A.2.

In total we add 274 new YSO candidates to the Orion A catalog inside the VISTA coverage, with 154 selected inside and 120 selected outside the IRAC region. The sources are classified with the methods discussed in Sect 3.1 into 6 new protostar candidates, 8 new flat-spectrum candidates, and 260 new disk candidates (Table 5). Sources inside the IRAC coverage might have been missed previously due to different selection criteria, and by adding VISTA we gain sensitivity in the NIR. YSOs selected near the ONC often lack longer wavelength measurements ( $\lambda > 5 \mu\text{m}$ ), which can lead to erroneous classification of these sources. There are 67 such new disk candidates (333 total) with the longest measured wavelength at  $4.5 \mu\text{m}$  (I2), mostly near the ONC. This lack of longer MIR band leads to less reliable classifications. Therefore, some of these disks can still be flat-

<sup>12</sup> The VISTA/WISE based selection adds 104, leading to 117 in combination with the VISTA/M1 selection (including the red MIPS coverage) beyond the IRAC coverage. Therefore, 43 sources are selected by both methods, meaning 13 are only selected by VISTA/M1.

spectrum or protostar candidates. They can also be influenced by contamination near the ONC, that is not reflected in the photometry error, and for the same reason the extinction correction can be erroneous. However, visual inspection of these sources does not show signs of deep embeddedness or outflows, therefore, the Class II status is more likely.

In Figure 6 the new YSO candidates are shown on top of the *Herschel* map including VISTA and *Spitzer* survey contours. The new candidates in the surroundings are often located near the IRAC coverage, especially near the ONC region. Beyond the L1641 region to the galactic south-east we find almost no new YSOs, whereas to the galactic north of L1641 and to the galactic south-west we find some scattering of new candidates. Overall, the distribution of the new candidates highlights the influence of the massive ONC, by showing a larger scatter near this region. A more detailed analysis of the (2D) distribution of our final sample will be discussed in Sect. 5.3.

#### 4.3. Final YSO sample and YSO re-classification

The updated YSO catalog for Orion A contains 2978 YSO candidates with IR-excess, located inside the VISTA coverage. Included are the revisited 2704 YSO candidates (2839 minus 135), plus the newly selected 274 sources. The final catalog is presented in Appendix C, which contains a `Class_flag` for revisited (1), new (2), rejected (3), and uncertain candidates (4). The 2978 YSO candidates are classified into 2606 disks (Class II/III), 184 flat-spectrum, and 188 protostar (Class 0/I) candidates (Table 5). The flat-spectrum sources are composed of 58 previously identified MGM disks (32 %), and 117 MGM protostars (64%).

We reclassify 10% of the MGM protostar candidates as disk candidates (33 disks out of 330 Ps, see also Fig. 5). These reclassified sources are mostly near the ONC. Reasons for the different classification are mainly due to extinction or contamination effects. When correcting for extinction, some sources do not show significant IR-excess to be classified as protostars by our methods. In addition, the different classification method compared to Megeath et al. (2012) can lead to different results, because some candidates do not show significant excess even without dereddening. Unfortunately, sources in the ONC region often lack longer wavelength detections due to saturation (e.g., missing IRAC3-4 or MIPS1). Using solely NIR colors can be ambiguous, therefore, we used visual inspection and a more detailed SED examination for a final decision (see Sect. 3.2). For example, if the VISTA image shows a bluish NIR source and if the source has an optical counterpart it is very unlikely to be an embedded protostar. There are also rather exotic protostar candidates, showing typical red NIR to MIR colors but without a MIPS1 or PACS counterpart (in surroundings where these bands are not yet saturated). Other sources have a catalog entry for MIPS1 while the images show only extended fuzzy counterparts from the surrounding cloud structure, therefore, these are contaminated by extended emission. Overall, the ONC is a very critical area, and the YSO classification in this region is likely more prone to errors than in other regions.

We check especially for protostars and flats if they show a corresponding *Herschel*/PACS counterpart by 1) using HOPS information from FFA16, 2) using the PACS point source catalog, and 3) visually inspecting the PACS images. Out of the 184 flat-spectrum sources, 102 (55%) coincide with a PACS point-source, suggesting that these flats might still be associated with envelopes. For the rest, there are 34 (18%) without PACS, and for 48 (26%) we can not be sure, due to extended emission, crowded regions, or saturated regions near the ONC. The flats

with PACS are overall brighter than those without. Concerning the 188 Class 0/I candidates, there are 168 sources (89%) with a clear PACS counterpart. Out of the remaining 20 sources there are 6 with no counterpart and for 14 we can not tell due to mentioned contamination issues. This makes these 20 sources suspicious or more uncertain protostar candidates. For the disk candidates we also check if there are cross-matches with the PACS point source catalog, however, we did not check all disk candidates visually for PACS counterparts. Out of the 2606 disk candidates, 249 (~10%) are clearly associated with a PACS point source.

The resulting spatial distribution of the three YSO classes is presented in Fig. 7. By eliminating false positives, the distribution of protostars now appears to be less scattered and more confined to regions of high dust column-density. Moreover, protostars and flats show a similar distribution and are almost equal in sample sizes. Both seem to be connected or located near regions of high dust column-density, whereas the disk sources are already more dispersed, while also larger in number. Hence, we quantify this behavior in Sect. 5.3.

## 5. Discussion

In this section we firstly discuss the completeness of the YSO sample, secondly the issues that come with YSO classification, especially concerning the flat-spectrum class, and finally the distribution of the three YSO classes with respect to regions of high dust column-density.

### 5.1. Completeness

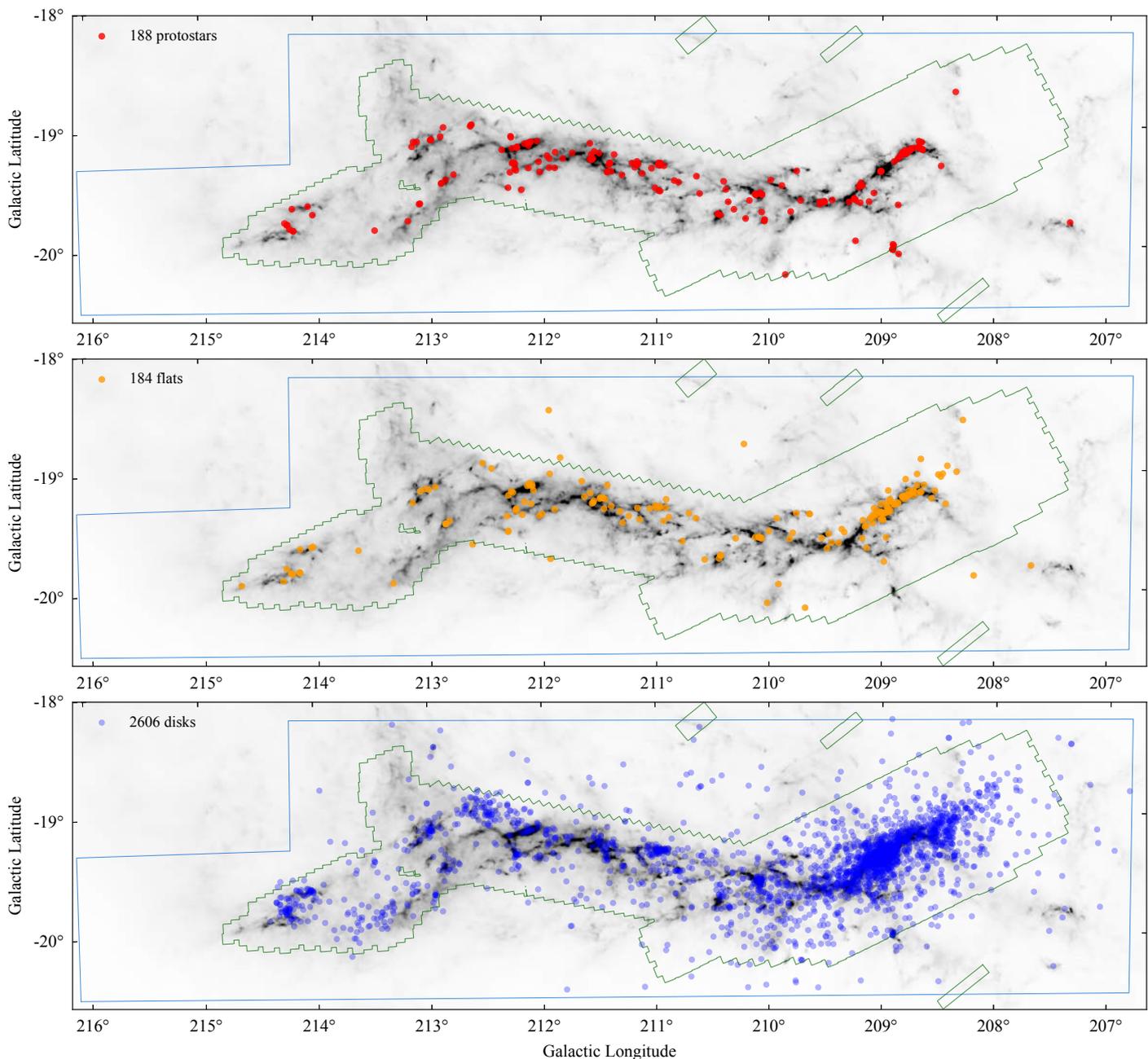
Estimating the completeness of our selection, or any similar selection, is complicated. We will partly refer to Megeath et al. (2016), who estimated the completeness of the MGM sample in two ways. First, they estimated the nebular background and source confusion, using the route median square deviations (RMEDSQ) of the IRAC pixels surrounding each YSO candidate. This gives an estimate of the incompleteness due to local MIR background emission, which is spatially varying, and increasing with stellar density. Second, they used *Chandra* X-ray data at the ONC (COUP, Getman et al. 2005a,b), to estimate the incompleteness in the crowded ONC region, which is affected by very bright IR nebulosity, and high extinction. They do this by carefully comparing the number of COUP sources with and without IR counterparts to their known *Spitzer* YSOs (MGM sample). With this approach they correct the number of YSO candidates with IR-excess in Orion A from 2821<sup>13</sup> to 3191, using the COUP correction, and finally to 4199, using the correction due to local MIR background emission. This means an incompleteness of about 49% for the Orion A sample inside the IRAC coverage.

The YSO sample in this paper, inside the IRAC coverage, included the revisited 2692 MGM sources<sup>14</sup> plus the new 154 candidates added inside the IRAC region, leading to 2846 YSO candidates. The final number is similar to the original MGM sample size, therefore, we adopt their completeness estimate of about 50% as an upper limit.

We now focus on the COUP covered region containing 630 MGM sources. Megeath et al. (2016) estimate 370 extra sources after applying the COUP correction, meaning there should be

<sup>13</sup> This number does not include the six red protostar candidates (RP), since the completeness was estimated for IRAC.

<sup>14</sup> 2827 MGM sample minus 135 (92 false positives and 43 uncertain)

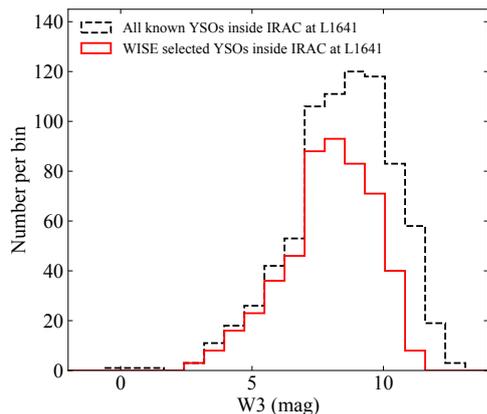


**Fig. 7.** Distribution of all 2978 YSO candidates as selected and classified in this work. The Classes are shown separately from top to bottom: protostar (Class 0/I, top, red), flat-spectrum (middle, orange), and disk candidates (Class II/III, bottom, blue), displayed on the *Herschel* map. The *Spitzer/IRAC* and VISTA coverage contours are shown in green and blue respectively. This highlights the location of new sources in the surroundings, and the limitations of the VISTA survey coverage.

about 1000 YSOs with IR-excess in the small coverage. We add 73 new candidates in this region (Sect. 3.3), of which 56 are X-ray detected COUP sources, meaning that we were only able to add about 15% of the estimated missing sources towards the ONC, or about 20% including the 17 sources without an X-ray counterpart. Assuming MGM completeness, we are still missing about 30% of YSOs with IR-excess towards the ONC. Also of note in this context, about 75% of the newly identified YSOs with IR-excess have an X-ray counterpart within the COUP coverage. This provides an independent support for these new candidates. Considering the whole YSO sample (revisited + new), there are about 81% IR YSOs with an X-ray counterpart within the COUP coverage.

The *WISE* completeness is not directly comparable to the *Spitzer* completeness. The inferior resolution and sensitivity of *WISE* misses faint sources and sources in crowded regions. To test the VISTA/*WISE* selection presented in this work (Appendix B.2), we check how many sources can be recovered inside the IRAC region, restricting this analysis to L1641 ( $l > 210^\circ$ ). This is a fair comparison for regions not as complicated as the ONC in the MIR (*WISE* saturates towards the ONC). We are able to recover about 59% of previously known YSO candidates in L1641. This shows what can be achieved with *WISE* in combination with deep NIR data in low-mass star-forming regions. Including the ONC region we recover only about 38%, highlighting the influence of massive-star-forming regions on low resolution MIR data.

To test the effect of crowding on *WISE* based selections, we redo the recovery test by comparing to only those MGM YSOs in L1641 with no other *Spitzer* source closer than  $6''$  as nearest neighbor. Surprisingly, we do not find a significant difference, and get again a recovery rate of about 59% when comparing only to non-crowded *Spitzer* sources<sup>15</sup>. This suggests that *WISE* is mainly limited by sensitivity issues, since we are losing mostly faint YSO candidates, due to our error and magnitude cuts and various steps to clean the *WISE* data of extended emission. This is highlighted in Fig. 8, comparing the *W3* magnitude of all known YSO candidates in the IRAC L1641 region to those selected by VISTA/*WISE*. The *WISE* selection is especially incomplete for sources fainter  $W3 \gtrsim 7$  mag.



**Fig. 8.** Histogram of the *WISE* *W3* magnitude for all known YSO candidates inside the IRAC region at L1641 ( $l > 210^\circ$ ) with  $w3sigpro < 0.5$  (black dashed line), compared to VISTA/*WISE* selected YSOs in the same region (red solid line).

We can use this recovery rate to estimate the completeness of our VISTA/*WISE* selection outside IRAC, where we added 104 new YSO candidates with this method. If the YSO density beyond the IRAC coverage is similar to a low-mass star-forming region like L1641, we would expect about 72 additional YSO candidates in the surroundings. Adding this to the 104, a *Spitzer* based selection would have found 176. Now we can add the Megeath et al. (2016) completeness estimate for the *Spitzer* YSOs of 49%. With this we get an upper limit of new YSO candidates beyond the IRAC region of 359 sources. The combined VISTA/*WISE* and VISTA/*M1* selections give 117 new YSO candidates outside IRAC, meaning we selected only about 1/3 of possible new candidates. However, the completeness was estimated by Megeath et al. (2016) based on bright MIR nebosity near the *Spitzer* YSOs in the whole Orion A region, including the ONC. Since regions outside IRAC are less influenced by background MIR emission, the 359 are indeed an upper limit, as it is likely that we are missing less sources towards these regions. Moreover, the YSO density decreases beyond the IRAC coverage, meaning less crowded sources, which also suggests that, locally, we likely miss less than 2/3 of the YSO candidates.

Comparing the  $\sim 59\%$  recovery rate with other *WISE* based selections from the literature, we find that we recover slightly more than Koenig et al. (2015), with a 50% recovery rate. They compare the Koenig & Leisawitz (2014) *WISE* YSO selection scheme with *Spitzer* selected YSO candidates in various regions. Another previous *WISE* based study covering the whole Orion A region (Marton et al. 2016), using machine learning based selec-

tion criteria, recovers about 20% of YSO candidates at Orion A, with a small fraction of contamination ( $\sim 3\%$ ). Compared to our selection, we recover almost twice as much, considering the 38% recovery rate when including the ONC.

## 5.2. Inferring on the meaning of flat-spectrum sources

In this work we perform YSO classification based on the MIR spectral index, defined by the observed IR excess. This is sometimes just a rough estimate of the true evolutionary Stage, however, for low-mass stars the method is a well established tool (e.g. Lada et al. 2006). Using a grid of SED models would give more detailed results, by taking into account inclination and/or extinction effects (e.g. Whitney et al. 2003b; Robitaille et al. 2006; Crapsi et al. 2008; Forbrich et al. 2010; Furlan et al. 2016), although to accurately model an entire YSO population would require complete reliable photometric observations, covering wide wavelength ranges, ideally reaching from the optical into the FIR and mm-range, as shown by Furlan et al. (2016). Additionally, obtaining spectra (e.g. *Spitzer*/IRS) gives further useful insights into emission and absorption lines ( $H_\alpha$ ,  $H_2O$ , Si, PAH). However, this would exceed the scope of this paper, where we focus on a statistically significant sample, while a detailed analysis using SED modelling is generally only possible for smaller sub-samples of a YSO population.

Therefore, we like to review the various uncertainties influencing the reliability of YSO classification. Particularly uncertain are the flat-spectrum sources, since they are at the border between Classes I and II, spanning a narrow spectral index range. The physical significance of defining flat-spectrum sources between  $-0.3 < \alpha < 0.3$ , as suggested by Greene et al. (1994), is highly debatable. For example, Teixeira et al. (2012) use  $-0.5 < \alpha < 0.5$ . Also, if there would be a physically meaningful separate Class between Classes I and II, one would expect three distinct over-densities in the various CCDs and CMDs, which is not observed. To make things even more complicated, more massive stars disperse their disks faster (Lada et al. 2006), so different SED shapes are expected just as a consequence of the mass distribution of the YSOs, even if all stars had the same age. Moreover, (Whitney et al. 2004) show that also the luminosity (or mass) of the central YSO, and not only of the disk, influences the SED shape. For example, the emission of the stellar photosphere of a low mass star peaks in the NIR, while for more massive stars it peaks in the optical. This means the latter contribute less to the NIR part of the SED, which leads to a slightly more rising observed IR SED, even if the disk mass and extend is the same as that of the lower mass star. This suggests that classification is a function of luminosity, which can actually be seen in some color-magnitude diagrams (see Fig. 4), or when plotting  $\alpha$  versus a magnitude. Therefore, the traditional definition of flat-spectrum sources seems to introduce a bias towards brighter sources. Moreover, Heiderman & Evans (2015) looked for envelope tracers in a significant sample of Class 0/I and flat-spectrum sources, and find that about 50% of flats are true Stage 0/I sources. Therefore, they conclude, that nothing distinctive occurs within the flat-spectrum category, suggesting that this category has no physical significance.

On top of that, classification can be influenced by different geometric effects, like disk inclination or extinction effects, as highlighted in the introduction. Crapsi et al. (2008) find that seeing the disk near edge-on, can be responsible for most observed flats, also suggested by Chiang & Goldreich (1999) and Whitney et al. (2003b,a, 2004). To test the effect of disk inclination, we use the SED models of Robitaille et al. (2006) for Class II

<sup>15</sup> The non-crowded VISTA/*WISE* recovery rate compared to all *Spitzer* sources in L1641 is about 53%.

YSOs. For sources with more than 2 solar masses and inclinations greater than  $75^\circ$  (disk is seen near edge-on) we find that less than 1% are misidentified as flat-spectrum sources. However, the majority of YSOs are low-mass stars ( $M < 2 M_\odot$ ), for which we estimate that about 3.6% would show flat spectra due to disk inclination effects, which corresponds to about half of the observed flat-spectrum sources in our sample. This is similar to the  $\sim 50\%$  Stage II sources, that Heiderman & Evans (2015) find in their flat sample. However, Muench et al. (2007) find that flats tend to be overall more luminous than disks. At the same time they point out that edge-on disks tend to be sub-luminous, due to obscuration by the disk. If flats were caused largely by inclination effects, this would contradict the first statement. We checked the luminosity of the YSOs by calculating the bolometric luminosity ( $L_{\text{bol}}$ ) with the method from Myers & Ladd (1993). We get a median of  $(0.2 \pm 0.1)L_\odot$ ,  $(0.7 \pm 0.6)L_\odot$ , and  $(1.4 \pm 1.1)L_\odot$ , for disks, flats, and protostars, respectively. For the dereddened photometry we get  $(0.3 \pm 0.3)L_\odot$ ,  $(1.6 \pm 1.4)L_\odot$ , and  $(2.0 \pm 1.7)L_\odot$ . Indeed, the flats are overall more luminous than the disks, also, they lie in-between the disks and protostars. This gives the impression that flat-spectrum sources can be interpreted as a transitional evolutionary class.

Moreover, Furlan et al. (2016) showed that most of their investigated sample of flat-spectrum sources in Orion A show signs of envelopes when applying SED modelling. They point out, that this sample likely represents protostars at different stages in their envelope evolution. Megeath et al. (2012), who investigated the whole dusty YSO population, only presented a simple color based separation into disk dominated PMS stars (D) and protostars (P) which are similar to Classes II and I, respectively. The flat-spectrum sources in this paper are composed of 32% previously classified MGM disks and 68% MGM protostars, which would also suggest at first guess that these sources are likely younger compared to the average Class II, and not simply disk inclination effects. Moreover, LL16 point out, that 14<sup>16</sup> out of their investigated 44 low- $A_K$  MGM protostars are modeled as Stage II YSOs, which show overall a rather flat SED. They suggest that these are still very young, likely at the beginning of the disk dominated PMS phase, and therefore were misclassified as protostars previously. Also pointed out by LL16, the median spectral index for disks between 2 and 8  $\mu\text{m}$  ( $\alpha_{\text{KI}}$ ) is about  $-1.33$ , which is the expected value for a spatially flat accretion (or re-processing) disk. We get a similar median for this spectral index of  $-1.42$  to  $-1.28$  (de-reddened and observed). This suggests that the majority of the disk sources are not highly flared. This might be explained by sufficient dust settling onto the circumstellar disk during the Class II evolution (D'Alessio et al. 1999; Lewis & Lada 2016).

We can contribute to this discussion by looking at the spatial distribution of the various YSO classes with respect to regions of high dust column-density. Figure 7 suggests a stronger connection of protostars and also flats to these regions, while disk sources are more dispersed. This stronger connection to denser regions of protostars and flat-spectrum sources was also pointed out by Heiderman et al. (2010) and Heiderman & Evans (2015). However, if flats were a result of disk inclination effects, they should be more evenly distributed, similar to confirmed disks. Unfortunately, the sample sizes are not directly comparable. There are about a factor of 10 more disks than flats or protostars. To make sure that we are not dealing with small number statistics, we will quantify the distribution in the next section.

### 5.3. Distribution of YSOs with respect to regions of high dust column-density

The spatial distribution of YSOs in Orion A was investigated by Gutermuth et al. (2011), Pillitteri et al. (2013) and Megeath et al. (2016), and a connection of protostars with high column-density was pointed out by Megeath et al. (2012). Such a behaviour was also highlighted, for example, by Muench et al. (2007); Jørgensen et al. (2007); Lada et al. (2010), or Hacar et al. (2016). Moreover, Heiderman & Evans (2015) indicated the same for flats. Furthermore, Teixeira et al. (2012) show that sources with thick disks are stronger connected to high-extinction regions as compared to more evolved anemic disks in NGC 2264. Recently, Hacar et al. (2017) found a strong correlation of protostars (Class 0/I) with the dense gas structure in NGC 1333 (not only high column-density), by observing  $\text{N}_2\text{H}^+$  line emission, as dense gas tracer. However, they do not find a significant connection of flat spectrum sources with dense gas. Unfortunately, we do not know (yet) the distribution of dense gas (volume-density) in the whole Orion A region. However, we can use the dust column-density from *Herschel* to infer on the connection of YSOs to a certain column-density threshold. For this we use a star formation threshold of  $A_K > 0.8$  mag as suggested by Lada et al. (2010). This is now possible for a larger field as compared to previous studies, since most of the above investigated regions are limited by the available survey coverages, while the majority of YSOs connected to Orion A should be present inside the VISTA coverage.

To quantify the distribution of the three YSO classes with respect to regions of high dust column-density, we evaluate the closest distance<sup>17</sup> to the next *Herschel map* pixel above the threshold (Fig. 9). The resulting normalized cumulative distribution function of the distances (given in pc) is presented in Fig. 10, with the bin-size corresponding to *Herschel* resolution. The displayed confidence intervals are obtained with bootstrapping. To this end, we draw random values out of each sample with replacement with 2000 iterations, while the sub-samples have the same size as the original sample size of each class. The resulting distributions are significantly different from each other within  $3\sigma$ . Essentially all protostars are seen in projection of regions of high dust column-density ( $99.5\% \pm 0.5\%$ ), while flats also show a stronger connection ( $86.4\% \pm 2.6\%$ ) compared to disks ( $63.1\% \pm 0.9\%$ )<sup>18</sup>. To get a measure for the background we check the distribution of all 800,000 VISTA sources, and find only about 7.7% of these sources are projected on regions above the extinction threshold.

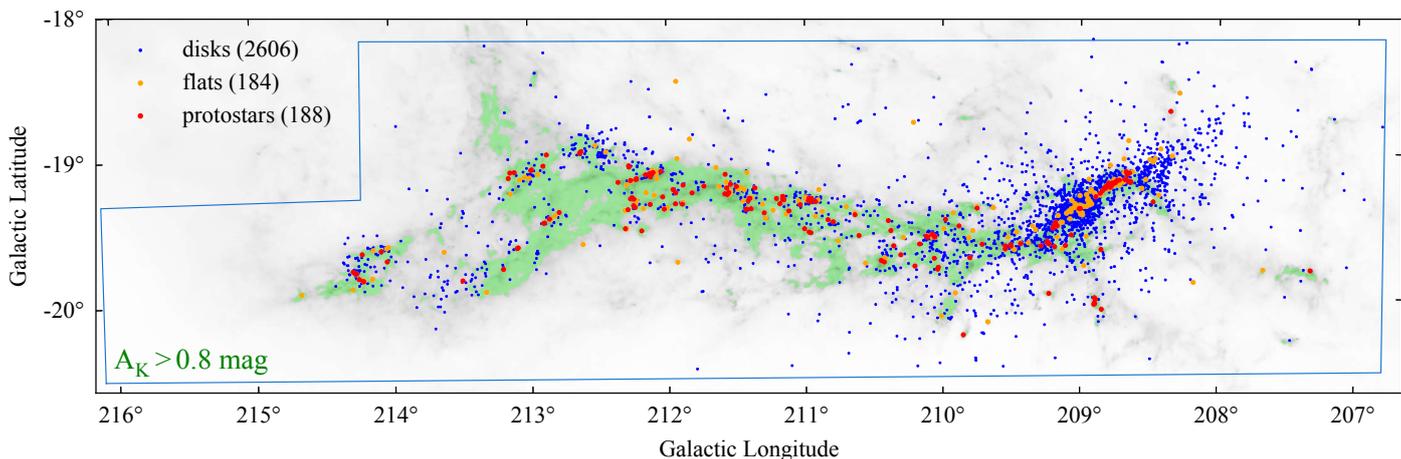
Looking at the original MGM YSO catalog, there are  $90.0\% \pm 1.7\%$  Ps and  $64.8\% \pm 1.0\%$  Ds projected above the threshold. Compared to the values when using the updated catalog, we see that the MGM protostars show a less clear connection to regions of high dust column-density, while the disk samples are similar within the errors. Differences are due to the exclusion of false positives and YSO reclassification by including flat-spectrum sources.

To check the influence of the chosen flat-spectrum range of  $-0.3 < \alpha < 0.3$  on the spatial distribution result in Fig. 10, we re-did the above test with a larger range of  $-0.5 < \alpha < 0.5$  (e.g., Teixeira et al. 2012). We find that the resulting distributions still show a significant difference between the classes, and the

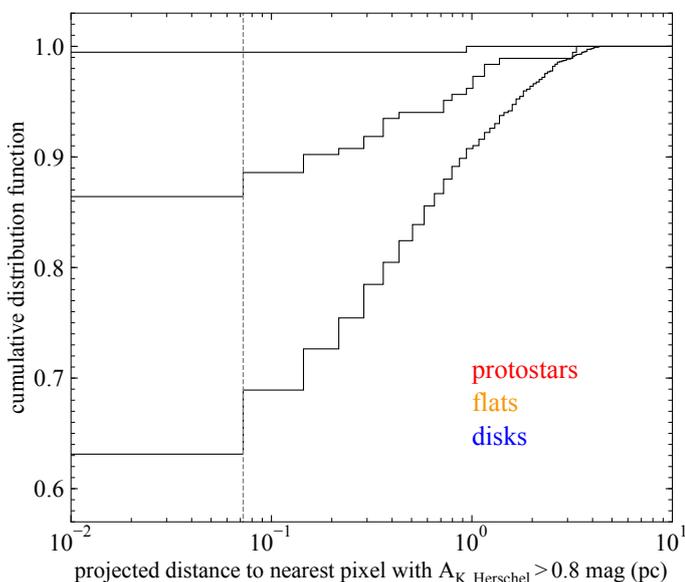
<sup>17</sup> `astropy.coordinates.match_coordinates_sky`  
(<http://www.astropy.org>)

<sup>18</sup> The percentages in parenthesis give the value of the first bin in Fig. 10, with the standard deviation error as uncertainty.

<sup>16</sup> Actually they find 18 Stage II sources, however, we excluded 4 false positives.



**Fig. 9.** The green shaded area indicates the extinction threshold of  $A_{K,\text{Herschel}} > 0.8$  mag. Superimposed are the YSO candidates (see legend). For each source we calculate the projected distance to the closest pixel in the *Herschel map* (green) above the extinction threshold. The resulting normalized cumulative distribution is shown in Fig. 10.



**Fig. 10.** Normalized cumulative distribution of the projected distances of the YSOs to the nearest *Herschel map* pixel with  $A_{K,\text{Herschel}} > 0.8$  mag (green pixels, Fig. 9). Separated in colors are the three YSO classes; Class I (protostars, red), flat-spectrum sources (flats, orange), and Class II/III (disks, blue). The observed distributions are shown as black lines, and the color shaded areas show the confidence interval contours at 68.3%, 95%, and 99.7% ( $1\sigma$ ,  $2\sigma$ , and  $3\sigma$ ). The first bin, indicated by the vertical gray dashed line, gives the resolution of *Herschel* ( $36''$ ,  $0.07$  pc @  $414$  pc). Hence, sources in the first bin are projected directly on top of regions of high dust column-density.

fraction of sources projected on top of high column-density stays essentially the same for each class.

Next, we investigate the possibility that the flat-spectrum sources presented in this paper are a simple mix of disks and protostars. For this we created a random mix of these two classes following the ratio protostars/disks ( $0.072$ ) to create a population of “synthetic flat sources”. In this case, it is clear that the distribution of the synthetic flat sources ( $65.8\% \pm 3.5\%$ ) is substantially different from the observed distribution of flats, being actually very similar to the distribution of disks, and can be ruled out. As a second more stringent test, we did the same experiment but only for disks that would be observed as proto-

stars due to inclination effects (estimated to be about 3.6% of the total sample of disks, see Sect. 5.2). This leads to an almost even number of Ps (90) and Ds (94) to be drawn randomly from these samples. In this case we find that the spatial distribution of “synthetic flat sources” is similar within  $3\sigma$  to the observed one, while being marginally different from each other within the  $1\sigma$  range, with  $80.8\% \pm 2.6\%$  projected on regions of high column-density. Nevertheless, we would expect that in the latter scenario the flat sources would be on average fainter because of the obscuration of the edge-on disk, which is the opposite of what is observed (see Sect. 5.2). This argument was also made in Muench et al. (2007). These simple experiments suggest that a flat-sources population created solely as a mix of protostars and disks is unlikely, or in other words, that most flats tend to be younger and therefore closer to the protostellar phase.

Overall, the spatial distribution of protostars in Orion A supports previous findings, for example by Hacar et al. (2017) in NGC 1333, but, we also find a stronger correlation of flats to high column-density regions as compared to Class IIs. This is not supported by Hacar et al. (2017), who does not find a significant correlation of flat-spectrum sources with dense gas (as traced by  $\text{N}_2\text{H}^+$ ). One reason could be that dust column-density does not only probe the real dense gas, and by using an extinction threshold of  $A_K > 0.8$  mag, we also include lower density and more diffuse regions, which are not included when specifically using dense gas tracers. However, by investigating a much larger region in Orion A compared to NGC 1333<sup>19</sup> and by having a larger statistically significant sample of YSOs, we can use the column-density as indicator for regions of dense cloud material. This suggests that flat-spectrum sources are indeed a younger evolutionary stage, and not simply disk inclination effects. Even if the flat sources are not anymore directly connected to dense gas, like in NGC 1333, they are still located near regions of higher column-density. This indicates that they did not have enough time to disperse sufficiently to show the same distribution as the more evolved Class IIs.

To make a stronger statement about the distribution of classes, follow up observations are needed to confirm the YSO nature of uncertain sources, and also of the scattered flat-spectrum sources. As shown for example by Heiderman & Evans (2015), many of previously classified protostars and flat-

<sup>19</sup> The investigated Orion A region is about a factor of 300 larger than that of NGC 1333.

spectrum sources that were found in regions of low dust column-density turned out to be background contamination. This can also be the case for some of the sources in the updated Orion A YSO catalog.

## 6. Summary

We have revisited and validated previous YSO catalogs (Megeath et al. 2012, 2016; Furlan et al. 2016; Lewis & Lada 2016) of the Orion A star-forming region using deep NIR ESO-VISTA data (VISION, Paper I), and added new YSO candidates in the larger field covered by VISTA, in combination with *Spitzer*, *WISE*, and *Herschel*/PACS. We summarize our results as follows:

1. We identified 274 new YSO candidates (6 protostars, 8 flat-spectrum sources, and 260 PMS stars with disks) of which 268 were selected by combining VISTA, *Spitzer*, and *WISE* selection criteria, and 6 candidates were selected by including *Herschel*/PACS photometry. A total of 120 candidates were found in regions beyond the previously analyzed *Spitzer*/IRAC survey. The rest are selected in regions covered by *Spitzer*/IRAC and were likely missed in previous works due to a) different selection criteria, b) sensitivity and saturation issues at longer wavelengths, and c) crowding and nebula contamination, especially in regions near the ONC.
2. Among the previously known 330 protostars and 2442 disk sources from Megeath et al. (2012, 2016), contamination levels are at least 6.4% and 2.2% respectively, mostly due to background galaxies or unresolved nebulosities, which were identified visually. These numbers are lower limits, because we can not rule out a residual degree of contamination, mainly due to unresolved galaxies (AGNs), or background giants. With this we conclude that previous surveys of Orion A are largely reliable, especially concerning the more evolved Class IIs, but they are incomplete due to limited survey areas and sensitivity issues near regions of bright nebula. The latter is also still an issue for this updated catalog, although we were able to slightly reduce the incompleteness near the ONC region for sources with IR-excess, and we extend the spatial completeness by using the larger field observed by VISTA.
3. The new catalog contains 2978 YSO candidates, including the 274 new and the 2704 revisited YSO candidates. They are classified as 188 (6.3%) protostar, 184 (6.2%) flat, and 2606 (87.5%) disk candidates, using extinction corrected spectral indices.
4. Within the *Chandra* observed COUP field, 81% of the IR YSOs, and 75% of the newly identified IR YSOs (Sect.5.1) appear to be associated with X-ray YSOs from Getman et al. (2005b). This provides independent support for the validity of these new IR selected YSOs. Considering the whole YSO sample from this work, about 38% of all IR YSOs have an X-ray counterpart, while large areas of the VISTA observed region are not covered by X-ray surveys.
5. We estimate that a search for YSOs in Orion A using *WISE* and VISTA recovers about 59% of the known YSOs in the region of L1641 ( $l < 210^\circ$ , excluding the ONC) inside the *Spitzer*/IRAC coverage. This shows what can be achieved by the all-sky *WISE* survey in combination with deep NIR data in regions not contaminated by massive star formation.
6. The spatial distribution of flat-spectrum sources relative to regions of high dust column-density ( $A_K > 0.8$  mag) shows

a significantly stronger correlation with these regions compared to Class IIs. This is a strong indication that they correspond to a younger evolutionary phase, and are not simply affected by disk/envelope projection effects or foreground extinction. With this we confirm earlier works (e.g. Furlan et al. 2016) that show that flats are likely protostars at different stages in their envelope evolution. Other studies find (e.g. Heiderman & Evans 2015) that only about 50% of flat-spectrum sources show signs of envelopes when using dense gas tracers. However, with the result from this work we can add to this discussion; maybe not all flats are still associated with envelopes and dense gas, but most are likely still very young, either at the end of the embedded protostellar phase or at the beginning of the disk-dominated PMS phase, since they had not enough time to disperse similarly as the majority of the Class II YSOs. To establish a final conclusion in the future, detailed SED modelling and follow-up observations are needed for all uncertain sources.

*Acknowledgements.* Author Großschedl gratefully acknowledges funding by the Austrian Science Fund (FWF) under project number P 26718-N27. This work is based on observations made with ESO Telescopes at the La Silla Paranal Observatory under program ID 090.C-0797(A). This work is part of the research programme VENI with project number 639.041.644, which is (partly) financed by the Netherlands Organisation for Scientific Research (NWO). AH thanks the Spanish MINECO for support under grant AYA2016-79006-P. J. Alves and T. Möller are part of the Research Platform Data Science @ Uni Vienna (<https://datascience.univie.ac.at>). We thank John A. Lewis for the very useful discussions on the validation of the YSO classification. This work is based on data obtained from (1) the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration; (2) the *Spitzer* Space Telescope, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA; (3) *Herschel*, which is an ESA space observatory with science instruments provided by European-led Principal Investigator consortia and with important participation from NASA. This work has made use of (1) TOPCAT <http://www.starlink.ac.uk/topcat>; (2) “Aladin sky atlas” developed at CDS, Strasbourg Observatory, France (Bonnarel et al. 2000; Boch & Fernique 2014); (3) the SIMBAD database and Vizier catalog access tool, operated at CDS, Strasbourg, France; (4) Python, <https://www.python.org>; and (5) Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013); (6) the Sloan Digital Sky Survey (SDSS), which is a joint project of The University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, the Max-Planck-Institute for Astronomy, Princeton University, the United States Naval Observatory, and the University of Washington. Apache Point Observatory, site of the SDSS, is operated by the Astrophysical Research Consortium. Funding for the project has been provided by the Alfred P. Sloan Foundation, the SDSS member institutions, the National Aeronautics and Space Administration, the National Science Foundation, the U.S. Department of Energy, and Monbusho, <https://www.sdss.org>; (7) the Digitized Sky Surveys DSS, which is an optical all-sky survey produced at the Space Telescope Science Institute under U.S. Government grant NAG W-2166.

## References

- Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, *ApJS*, 219, 12  
 Ali, B., Tobin, J. J., Fischer, W. J., et al. 2010, *A&A*, 518, L119  
 Andre, P. & Montmerle, T. 1994, *ApJ*, 420, 837  
 André, P., Ward-Thompson, D., & Barsony, M. 1993, *ApJ*, 406, 122  
 Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, 558, A33  
 Boch, T. & Fernique, P. 2014, in *Astronomical Society of the Pacific Conference Series*, Vol. 485, *Astronomical Data Analysis Software and Systems XXIII*, ed. N. Manset & P. Forshay, 277  
 Bonnarel, F., Fernique, P., Bienaymé, O., et al. 2000, *A&AS*, 143, 33  
 Bujarrabal, V., Young, K., & Fong, D. 2008, *A&A*, 483, 839  
 Carney, M. T., Yıldız, U. A., Mottram, J. C., et al. 2016, *A&A*, 586, A44  
 Chiang, E. I. & Goldreich, P. 1999, *ApJ*, 519, 279  
 Cieza, L. A., Schreiber, M. R., Romero, G. A., et al. 2010, *ApJ*, 712, 925  
 Cieza, L. A., Swift, J. J., Mathews, G. S., & Williams, J. P. 2008, *ApJ*, 686, L115

- Crapsi, A., van Dishoeck, E. F., Hogerheijde, M. R., Pontoppidan, K. M., & Dullemond, C. P. 2008, *A&A*, 486, 245
- Cutri, R. M., Wright, E. L., Conrow, T., et al. 2013, Explanatory Supplement to the ALLWISE Data Release Products, Tech. rep.
- Da Rio, N., Robberto, M., Soderblom, D. R., et al. 2009, *ApJS*, 183, 261
- Da Rio, N., Tan, J. C., Covey, K. R., et al. 2016, *ApJ*, 818, 59
- D'Alessio, P., Calvet, N., Hartmann, L., Lizano, S., & Cantó, J. 1999, *ApJ*, 527, 893
- Davis, C. J., Cervantes, B., Nisini, B., et al. 2011, *A&A*, 528, A3
- Dunham, M. M., Allen, L. E., Evans, II, N. J., et al. 2015, *ApJS*, 220, 11
- Emerson, J., McPherson, A., & Sutherland, W. 2006, *The Messenger*, 126, 41
- Evans, II, N. J., Dunham, M. M., Jørgensen, J. K., et al. 2009, *ApJS*, 181, 321
- Fang, M., Kim, J. S., Pascucci, I., et al. 2017, *AJ*, 153, 188
- Fang, M., Kim, J. S., van Boekel, R., et al. 2013, *ApJS*, 207, 5
- Fang, M., van Boekel, R., Wang, W., et al. 2009, *A&A*, 504, 461
- Fazio, G. G., Hora, J. L., Allen, L. E., et al. 2004, *ApJS*, 154, 10
- Fischer, W. J., Megeath, S. T., Ali, B., et al. 2010, *A&A*, 518, L122
- Fischer, W. J., Megeath, S. T., Stutz, A. M., et al. 2013, *Astronomische Nachrichten*, 334, 53
- Flewelling, H. A., Magnier, E. A., Chambers, K. C., et al. 2016, *ArXiv e-prints* [arXiv:1612.05243]
- Forbrich, J., Tappe, A., Robitaille, T., et al. 2010, *ApJ*, 716, 1453
- Furlan, E., Fischer, W. J., Ali, B., et al. 2016, *ApJS*, 224, 5
- Getman, K. V., Broos, P. S., Kuhn, M. A., et al. 2017, *ApJS*, 229, 28
- Getman, K. V., Feigelson, E. D., Grosso, N., et al. 2005a, *ApJS*, 160, 353
- Getman, K. V., Flaccomio, E., Broos, P. S., et al. 2005b, *ApJS*, 160, 319
- Greene, T. P. & Lada, C. J. 2002, *AJ*, 124, 2185
- Greene, T. P., Wilking, B. A., Andre, P., Young, E. T., & Lada, C. J. 1994, *ApJ*, 434, 614
- Gutermuth, R. A., Megeath, S. T., Myers, P. C., et al. 2009, *ApJS*, 184, 18
- Gutermuth, R. A., Pipher, J. L., Megeath, S. T., et al. 2011, *ApJ*, 739, 84
- Hacar, A., Alves, J., Forbrich, J., et al. 2016, *A&A*, 589, A80
- Hacar, A., Tafalla, M., & Alves, J. 2017, *ArXiv e-prints* [arXiv:1703.07029]
- Heiderman, A. & Evans, II, N. J. 2015, *ApJ*, 806, 231
- Heiderman, A., Evans, II, N. J., Allen, L. E., Huard, T., & Heyer, M. 2010, *ApJ*, 723, 1019
- Hillenbrand, L. A. 1997, *AJ*, 113, 1733
- Hillenbrand, L. A. & Carpenter, J. M. 2000, *ApJ*, 540, 236
- Hillenbrand, L. A. & Hartmann, L. W. 1998, *ApJ*, 492, 540
- Hillenbrand, L. A., Hoffer, A. S., & Herczeg, G. J. 2013, *AJ*, 146, 85
- Hsu, W.-H., Hartmann, L., Allen, L., et al. 2012, *ApJ*, 752, 59
- Hsu, W.-H., Hartmann, L., Allen, L., et al. 2013, *ApJ*, 764, 114
- Jarrett, T. H., Cohen, M., Masci, F., et al. 2011, *ApJ*, 735, 112
- Jørgensen, J. K., Johnstone, D., Kirk, H., & Myers, P. C. 2007, *ApJ*, 656, 293
- Kim, K. H., Watson, D. M., Manoj, P., et al. 2016, *ApJS*, 226, 8
- Kim, K. H., Watson, D. M., Manoj, P., et al. 2013, *ApJ*, 769, 149
- Koenig, X., Hillenbrand, L. A., Padgett, D. L., & DeFelippis, D. 2015, *AJ*, 150, 100
- Koenig, X. P. & Leisawitz, D. T. 2014, *ApJ*, 791, 131
- Koenig, X. P., Leisawitz, D. T., Benford, D. J., et al. 2012, *ApJ*, 744, 130
- Kryukova, E., Megeath, S. T., Gutermuth, R. A., et al. 2012, *AJ*, 144, 31
- Lada, C. J. 1987, in *IAU Symposium*, Vol. 115, *Star Forming Regions*, ed. M. Peimbert & J. Jugaku, 1–17
- Lada, C. J., Lombardi, M., & Alves, J. F. 2010, *ApJ*, 724, 687
- Lada, C. J., Muench, A. A., Haisch, Jr., K. E., et al. 2000, *AJ*, 120, 3162
- Lada, C. J., Muench, A. A., Luhman, K. L., et al. 2006, *AJ*, 131, 1574
- Lada, C. J. & Wilking, B. A. 1984, *ApJ*, 287, 610
- Lewis, J. A. & Lada, C. J. 2016, *ApJ*, 825, 91
- Lombardi, M. & Alves, J. 2001, *A&A*, 377, 1023
- Lombardi, M., Alves, J., & Lada, C. J. 2011, *A&A*, 535, A16
- Lombardi, M., Bouy, H., Alves, J., & Lada, C. J. 2014, *A&A*, 568, C1
- Manoj, P., Watson, D. M., Neufeld, D. A., et al. 2013, *ApJ*, 763, 83
- Marton, G., Calzoletti, L., Perez Garcia, A. M., et al. 2017, *ArXiv e-prints* [arXiv:1705.05693]
- Marton, G., Tóth, L. V., Paladini, R., et al. 2016, *MNRAS*, 458, 3479
- Meehan, L. S. G., Wilking, B. A., Claussen, M. J., Mundy, L. G., & Wootten, A. 1998, *AJ*, 115, 1599
- Megeath, S. T., Allgaier, E., Young, E., et al. 2009, *AJ*, 137, 4072
- Megeath, S. T., Gutermuth, R., Muzerolle, J., et al. 2012, *AJ*, 144, 192
- Megeath, S. T., Gutermuth, R., Muzerolle, J., et al. 2016, *AJ*, 151, 5
- Meingast, S., Alves, J., & Lombardi, M. 2018, *A&A*, 614, A65
- Meingast, S., Alves, J., Mardones, D., et al. 2016, *A&A*, 587, A153
- Meingast, S., Lombardi, M., & Alves, J. 2017, *A&A*, 601, A137
- Meisner, A. M. & Finkbeiner, D. P. 2014, *ApJ*, 781, 5
- Menten, K. M., Reid, M. J., Forbrich, J., & Brunthaler, A. 2007, *A&A*, 474, 515
- Moneti, A. & Reipurth, B. 1995, *A&A*, 301, 721
- Muench, A. A., Lada, C. J., Luhman, K. L., Muzerolle, J., & Young, E. 2007, *AJ*, 134, 411
- Muzerolle, J., Allen, L. E., Megeath, S. T., Hernández, J., & Gutermuth, R. A. 2010, *ApJ*, 708, 1107
- Myers, P. C. & Ladd, E. F. 1993, *ApJ*, 413, L47
- Ogura, K. & Walsh, J. R. 1991, *ApJ*, 101, 185
- Pilbratt, G. L., Riedinger, J. R., Passvogel, T., et al. 2010, *A&A*, 518, L1
- Pillitteri, I., Wolk, S. J., & Megeath, S. T. 2016, *ApJ*, 820, L28
- Pillitteri, I., Wolk, S. J., Megeath, S. T., et al. 2013, *ApJ*, 773, 80
- Poglitsch, A., Waelkens, C., Geis, N., et al. 2010, *A&A*, 518, L2
- Rebull, L. M., Koenig, X. P., Padgett, D. L., et al. 2011, *ApJS*, 196, 4
- Reipurth, B. 1989, *A&A*, 220, 249
- Rieke, G. H. & Lebofsky, M. J. 1985, *ApJ*, 288, 618
- Rieke, G. H., Young, E. T., Engelbracht, C. W., et al. 2004, *ApJS*, 154, 25
- Robitaille, T. P., Whitney, B. A., Indebetouw, R., Wood, K., & Denzmore, P. 2006, *ApJS*, 167, 256
- Rolph, C. D., Scarrott, S. M., & Wolstencroft, R. D. 1990, *MNRAS*, 242, 109
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, 131, 1163
- Stanke, T., Stutz, A. M., Tobin, J. J., et al. 2010, *A&A*, 518, L94
- Stern, D., Eisenhardt, P., Gorjian, V., et al. 2005, *ApJ*, 631, 163
- Stutz, A. M., Tobin, J. J., Stanke, T., et al. 2013, *ApJ*, 767, 36
- Szegedi-Elek, E., Kun, M., Reipurth, B., et al. 2013, *ApJS*, 208, 28
- Teixeira, P. S., Lada, C. J., Marengo, M., & Lada, E. A. 2012, *A&A*, 540, A83
- Tobin, J. 2017, in *Star Formation from Cores to Clusters*, 59
- Tobin, J. J., Stutz, A. M., Megeath, S. T., et al. 2015, *ApJ*, 798, 128
- Vieira, S. L. A., Corradi, W. J. B., Alencar, S. H. P., et al. 2003, *AJ*, 126, 2971
- Wang, J.-J. 2002, *Chinese J. Astron. Astrophys.*, 2, 260
- Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, *A&AS*, 143, 9
- Werner, M. W., Roellig, T. L., Low, F. J., et al. 2004, *ApJS*, 154, 1
- Whitney, B. A., Indebetouw, R., Bjorkman, J. E., & Wood, K. 2004, *ApJ*, 617, 1177
- Whitney, B. A., Robitaille, T. P., Bjorkman, J. E., et al. 2013, *ApJS*, 207, 30
- Whitney, B. A., Wood, K., Bjorkman, J. E., & Cohen, M. 2003a, *ApJ*, 598, 1079
- Whitney, B. A., Wood, K., Bjorkman, J. E., & Wolff, M. J. 2003b, *ApJ*, 591, 1049
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, *AJ*, 140, 1868

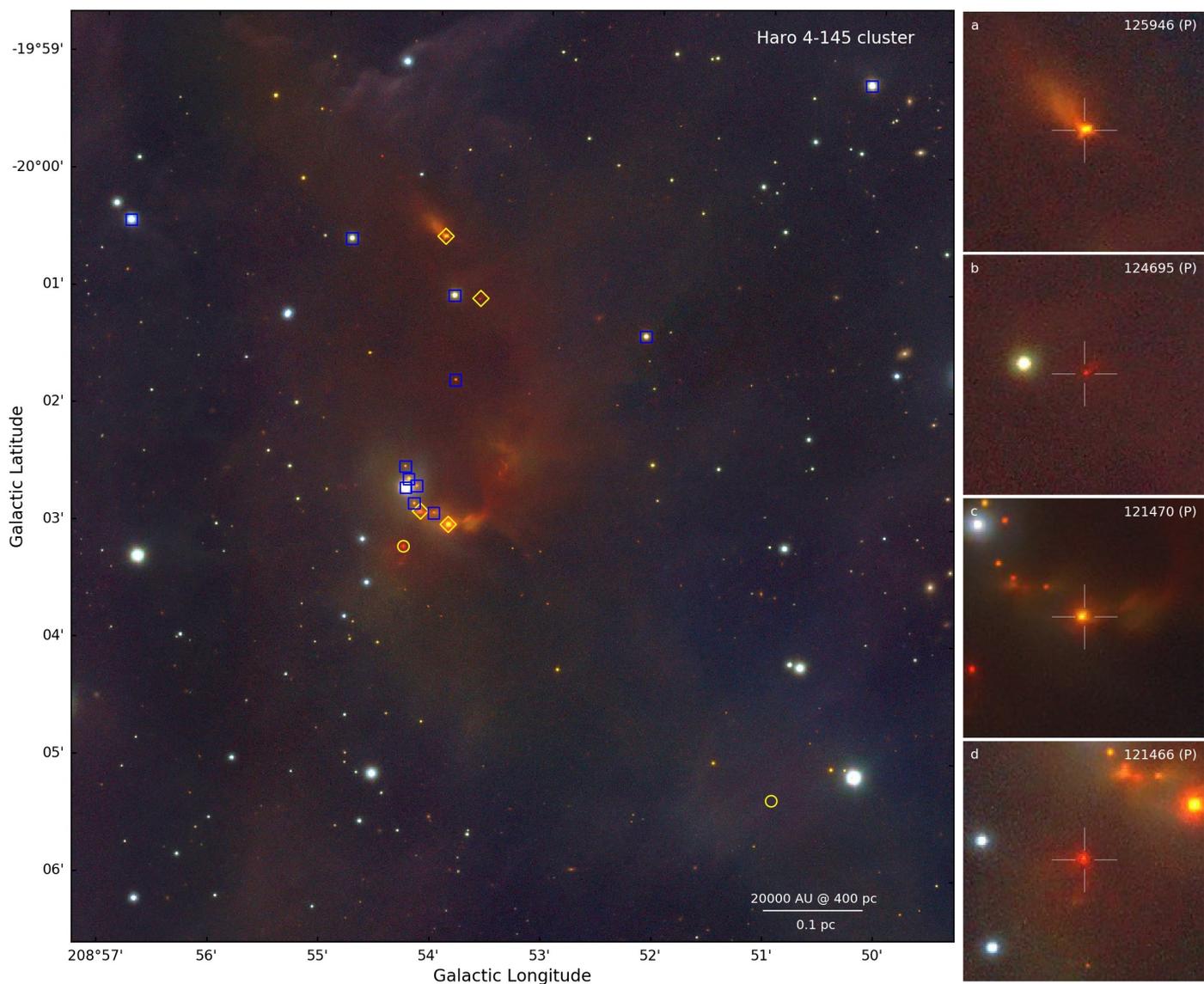
## Appendix A: Example images

In this section we present example image cut-outs of selected new YSO candidates, revisited interesting sources, and contaminating objects. The latter are objects that were erroneously identified as YSO candidates previously (false positives).

### Appendix A.1: New candidates and interesting objects

First we want to highlight some interesting objects, including revisited and new YSO candidates, shown in Figs. A.1 and A.2. The images show  $50'' \times 50''$  VISTA JHK three color composites centered on the objects, oriented in galactic coordinates. The sources are addressed in more detail below, with the listed small letters pointing to the corresponding VISTA cut-outs.

**a,b,c,d) Haro 4-145 clustering.** This clustering (see also Sect. 3.3) was already discussed by Wang (2002), who listed six YSO candidates (IRS 1 to IRS 6), calling it Orion A-W star-forming region. We call it “Haro 4-145 cluster”, named after the bright Class II YSO at its center. A visualization of this region using a VISTA three color composite is shown in Fig. A.1. The clustering is located at the *Spitzer*/IRAC survey border, therefore some sources were missed by MGM. It contains at least 15 YSO candidates (6 P, 9 D), of which 9 are in the MGM sample (3 P, 6 D), and 6 are selected in this work. One is a new VISTA/*WISE* selected protostar candidate, which is a known H<sub>2</sub>O maser (TFT95b H<sub>2</sub>O 053014.409-053750.890, Meehan et al. 1998). During visual inspection we found two new protostar candidates, which also have an entry in the PACS point source catalog. One candidate (ID 121466) is very close to the water maser, and was missed by our VISTA/*WISE* selection, likely because of the very crowded region and overlapping *WISE* PSFs. This source is already listed in Wang (2002) as IRS 2. The second new candidate is more isolated and lies farther to the galactic south ( $\sim 3.6'$ ). It only gets visible as a point-source in *Herschel*/PACS and has no NIR or MIR counterpart, located in a region of high dust column-density ( $A_{K,Herschel} \approx 4.6$  mag). It is



**Fig. A.1.** Clustering of YSOs south-west of the ONC (Haro 4-145 cluster) shown by a  $7.5' \times 7.9'$  VISTA 3-color composite, in galactic coordinates. The cluster harbours at least 6 protostars (yellow symbols) and 9 disk candidates (blue squares). There are 3 more disk candidates visible in the image a bit more scattered. The two new extra protostar candidates from Sect. 3.3 are the two yellow circles, while previously known protostar candidates are shown by diamond symbols. The source at the bottom right is the new Class 0 candidate, only detected as a point-source in *Herschel*/PACS. The cloud structure (region of higher dust column-density) gets visible in this NIR VISTA image. The slightly reddish illuminated cloud, containing the YSOs, seems to be separate from the background cloud structure. Moreover, background sources like galaxies are visible in this image. The four cutouts ( $50'' \times 50''$ ) on the right show zoom-ins on four protostar candidates in this region (see text).

likely a new Class 0 candidate (ID 116363). Four of the six protostar candidates located in the Haro 4-145 clustering are shown in zoom-ins on the right side in Fig. A.1. The two at the top (a, b) were selected already by MGM (125946, 124695), while the two at the bottom (c, d) are selected in this work (121417, 121466) and were also highlighted in Paper I. Source c is the known water maser, and source d is one of the new protostar candidates in this region. The source at the bottom of the overview image, which is the new Class 0 candidate, is not highlighted separately, because it is not visible in the NIR.

**e) New protostar candidate.** One of the new protostar candidates (ID 83961), selected by VISTA/WISE, is located in the galactic west of the ONC at a small visible clump of higher dust column-density, visible in the *Herschel* map at  $(l, b) \sim (207.3, -19.8)$ . VISTA resolves two sources, while WISE shows only one point-source covering both. The second VISTA source

(the right one) could also be a possible new YSO candidate. The projected spatial distance between the two sources is about  $4.5''$  ( $\sim 1800$  AU @ 400 pc). The source was already highlighted in Paper I.

**f) Protostar HH 83 IRS.** The source HH 83 IRS (or IRAS 05311-0631, ID 150542) lies at the border of the IRAC coverage, but is observed by *I2* and *I4*. It was already highlighted in Paper I and discussed in Reipurth (1989); Ogura & Walsh (1991); Moneti & Reipurth (1995), and Davis et al. (2011). The source coincides with an optical jet (Reipurth 1989) and a reflection nebula (Re 17, Rolph et al. 1990), well visible in VISTA. The protostar seems to be an isolated star-forming event in the galactic-south of L1641-North, at the location of a clump of higher dust column-density ( $A_{K, \text{Herschel}} \approx 2.5$  mag), well visible in the *Herschel* map at  $(l, b) \sim (209.85, -20.27)$ . The source is listed in the



**Fig. A.2.** VISTA  $50'' \times 50''$  cutouts of selected 18 YSO candidates. Sources e-p are new candidates (two top rows), and sources q-v (bottom row) are previously known objects. See text for more explanations.

HPPSC, but is not included in the HOPS catalog. The distance to the object is given with 450 pc (Reipurth 1989).

**g) New protostellar binary in L1641.** Next to a known Class 0 (MGM 1121) we select a second protostar candidate visually (ID 213612, see also Sect. 3.3), using PACS and VISTA images ( $\sim 5''$  separation, 2000 AU @ 400 pc). The new candidate shows a prominent outflow cavity. The known Class 0 is the faint reddish dot, to the bottom right from the new source, overshadowed by the outflow of the new candidate. Both sources correspond to an elongated PACS point-source. The system was also highlighted by Tobin (2017) as protostellar binary candidate.

**h) New edge-on disk.** With VISTA/WISE we select a new interesting object (ID 404173) off of the cloud to the galactic north of L1641 at  $(l, b) \sim (210.22, -18.81)$ . We classify the source as flat-spectrum candidate based on its spectral index, but the VISTA image shows an extended object, likely an edge-on disk candidate. The PACS images show a clear point source, as well as the *Herschel* map, with a peak extinction value of  $A_{K, \text{Herschel}} \approx 0.33$  mag, while the immediate surrounding pixels are of lower extinction ( $A_{K, \text{Herschel}} \lesssim 0.1$  mag). This indicates a massive dusty disk or envelope. It even seems that the source lies in the middle of an almost circular excavated area (see *Herschel* map) with an extension of about  $30'$ . Its flattish edge-on appearance is similar to Gomez's Hamburger (Bujarrabal et al. 2008), which is suggested to be an A-type PMS star with a massive circumstellar disk, that is not associated with any inter-stellar molecular cloud. The new edge-on candidate needs more investigation to determine its true nature, and to rule out that it is not an exotic extra-galactic source, or a post AGB star.

**i) New flat-spectrum candidate UY Ori.** One new flat-spectrum candidate selected by VISTA/WISE (ID 107068) is the known Herbig Ae-Be (HAeBe) star UY Ori (B9 III, Vieira et al. 2003). The WISE point source actually covers two resolved VISTA sources. The second source is maybe a companion, with a projected distance of  $\sim 2.4''$  ( $\sim 960$  AU @ 400 pc). We classify the source as a flat-spectrum source, even though the WISE

spectral index is slightly above 0.3 ( $\alpha_{\text{WISE}} = 0.43$ ). Though the local low extinction ( $A_{K, \text{Herschel}} \approx 0.1$  mag), its location far from high extinction regions ( $\sim 30'$ ), and the brightness and colours in the NIR and optical suggest that it is not a protostar. Moreover, as discussed for example by Whitney et al. (2004), the higher luminosity and temperature of high mass stars require a different interpretation of the observed spectral index, and the classification does not follow the same standards as for low mass T-Tauri stars.

**j) New flat-spectrum candidate ID 95901.** This new flat-spectrum candidate (also highlighted in Paper I), selected by VISTA/WISE, is a prominent bright YSO. It is located near the region of the already mentioned new protostar candidate ID 83961 (source e, distance  $\sim 20'$ ), at the galactic west of the ONC. It is located at an isolated clump of higher dust column-density ( $A_{K, \text{Herschel}} \approx 1$  mag) and is associated with a prominent PACS point-source.

**k) New flat-spectrum candidate ID 502202.** This new flat-spectrum candidate, selected by VISTA/WISE, appears to be a double star in VISTA, similar to UY Ori. It is associated with a PACS point-source, and interestingly, it also coincides with a point-like peak in the *Herschel* map ( $A_{K, \text{Herschel}} \approx 0.36$  mag).

**l) New flat-spectrum candidate ID 521492.** This is the only new flat-spectrum candidate selected inside IRAC (by VISTA/I24, Appendix B.1.2). It appears to be a double star, and was missed previously maybe due to source confusion, or different selection criteria.

**m) New flat-spectrum candidate or AGN?** The last new flat-spectrum candidate (ID 565713) discussed here is a rather suspicious YSO candidate. It could also be a bright AGN (Blazar), since it is located a bit off from the cloud ( $\sim 30'$ ) to the north of L1641 at  $(l, b) \sim (211.94, -18.51)$ , in a region of low extinction ( $A_{K, \text{Herschel}} \approx 0.15$  mag). It is the only of the six new flat-spectrum sources that dose not have a PACS counterpart. Furthermore, the surroundings show a loose clustering of background galaxies.

**add h-m)** This six new flat candidates are rather untypical flats (prominent edge-on disk, early spectral-type, luminous, double-star, or suspicious AGN?). Moreover, four of them (except ID 95901 and 521492) are located at regions of low dust column-density ( $A_{K,Herschel} < 0.4$  mag).

**n,o,p) Three new PACS transition disks.** Three of the new YSO candidates show only an excess in PACS, but not in the NIR or MIR, and we classify them as transition disk candidates (ID 377204, 459841, 522530). All three seem to be surrounded by some nebulous haze, especially source 459841 (**o**) shows a prominent reflection nebula.

**q) New disk.** Source 104064 is a prominent new disk candidate (also highlighted in Paper I), selected by VISTA/WISE, to the galactic south-west of the cloud at  $(l, b) \sim (208.6, -20.1)$ . It is located at an isolated small clump of higher dust column-density ( $A_{K,Herschel} \approx 1$  mag), and is also associated with a PACS point-source. The VISTA image reveals a prominent reflection nebula and a spiral-like structure.

**r) Suspicious protostar 2MASS J05344694-0544512.** This source is a suspicious protostar candidate (ID 197736, MGM 1238, HOPS 24), by showing a whitish point-source in VISTA, with no signs of reddening or outflows in the NIR. Moreover, it has no clear PACS counterpart. Still, we kept it listed as protostar candidate due to its rising MIR SED.

**s) V1314 Ori** was previously identified as protostar candidate (MGM 1503, HOPS 49) and listed as FU Ori-type star in Simbad. The source is quite bright in the optical and NIR (Pan-STARRS  $g = 16.0$  mag, VISTA  $K_S = 12.5$  mag), with no significant signs of reddening in this wavelength ranges. We reclassify the source as disk candidate, even though the colors and spectral indices point to a protostar candidate, maybe due to high variability. It is also listed as CTTS in Szegedi-Elek et al. (2013). Still, there is the possibility that it is a pole-on protostar.

**t) V2168 Ori** is an emission line M-star (Hillenbrand et al. 2013), listed as galaxy in FFA16. It coincides with some outflow or line-of-sight nebulosity. The VISTA image even gives the impression, that the source was just floating out of the cloud.

**u) V2275 Ori** is listed as extra-galactic (EG) in the membership list of the COUP catalog by Getman et al. (2005a). However, its color and brightness suggest a Class II YSO, also listed as such previously (MGM 1559, HOPS 51).

**v) TU Ori** is a known prominent YSO and was selected as a new disk candidate by VISTA/I24. It is given as an F or G spectral type in the literature (Hillenbrand et al. 2013), located near the ONC ( $\sim 2.4'$  distance to Trapezium). It is not listed in MGM, despite its brightness. It is at the border to be a flat-spectrum source, and shows quite inconsistent spectral indices, maybe due to contamination effects in this region.

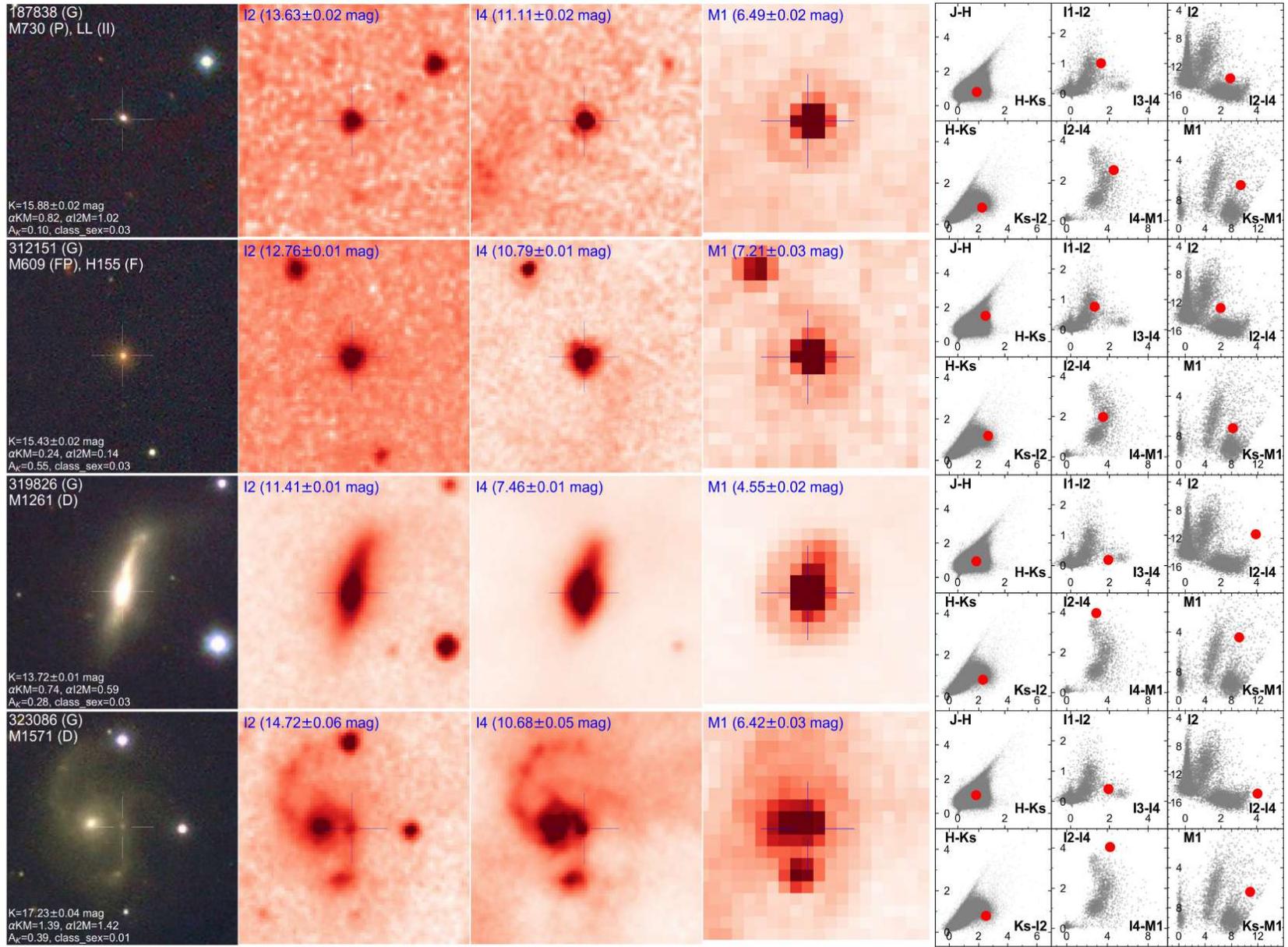
## Appendix A.2: Contaminating objects

In this section we show examples of contaminating objects in Figs. A.3 to A.5. The images are constructed as follows. The first column shows a VISTA JHK three color composite ( $50'' \times 50''$ ) centered on the object and oriented in galactic coordinates. The second to fourth columns are *Spitzer* I2, I4, and M1 cutouts respectively, showing the same region as VISTA. In the upper left corner of the VISTA cutouts we give the running ID from this work with the classification in brackets. If present, the MGM index and classification are given as M#(class), the HOPS index and classification as H#(class), and/or the classification from LL16 as LL(class). At the bottom of the VISTA cutouts we give some selected information on each source: the  $K_S$  magnitude (given as K), the spectral indices  $\alpha_{KM}$  and  $\alpha_{12M}$ , the line-of-sight

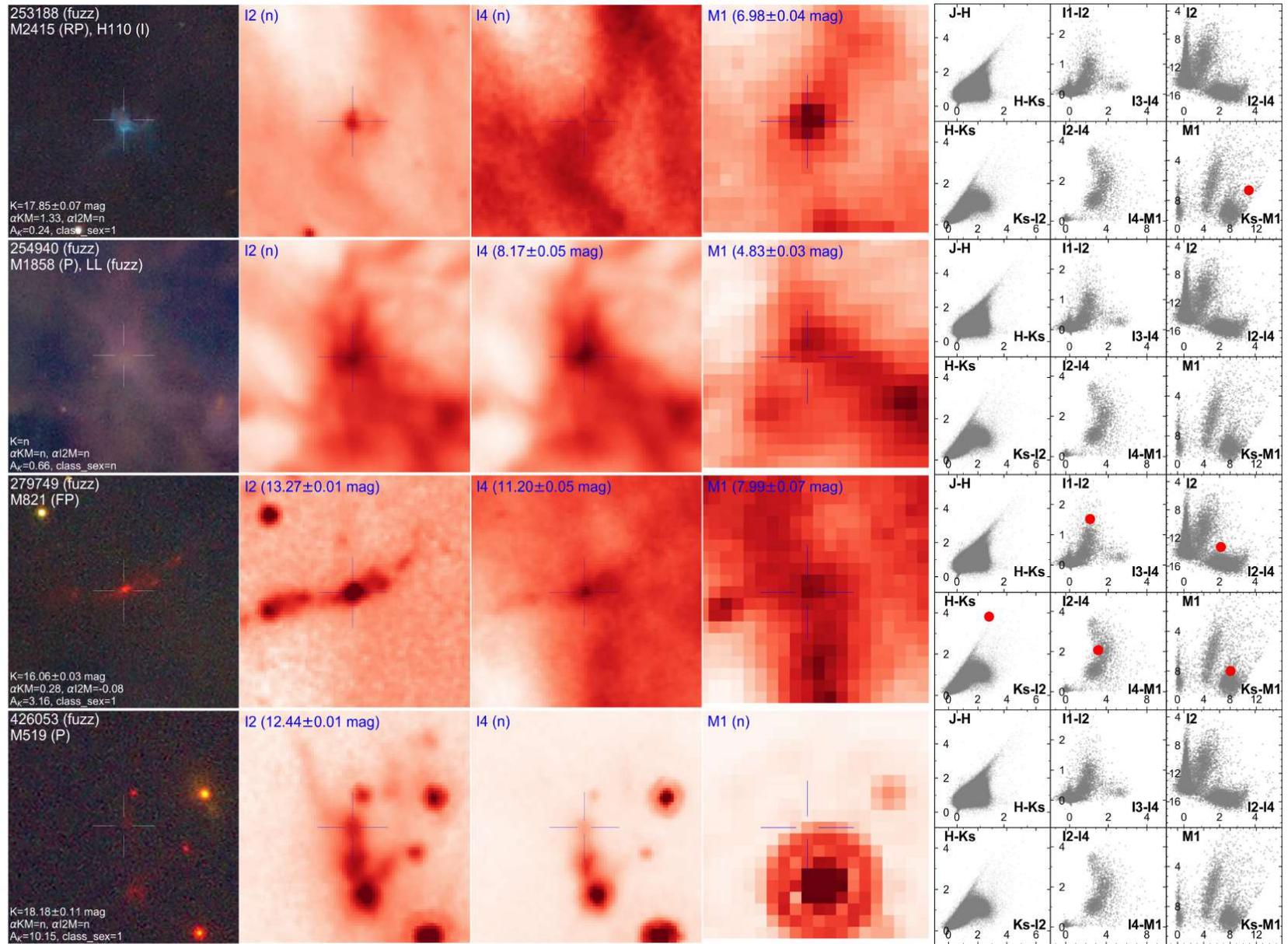
extinction from the *Herschel* map at the position of the source (given here as  $A_K$ ), and the extension flag ClassSex. The last columns show six selected color-color and color-magnitude diagrams with the source highlighted by a red dot. These diagrams demonstrate the color spaces the source occupies. Some contaminating objects, especially nebulosities, do not have measurements in all bands, and therefore the diagrams do not include the source. We still keep the diagrams for consistency. Besides, the fact that fuzzy nebulous sources are not detected in all bands is another clue to identify them.

Figure A.3 shows four examples of galaxies that were erroneously classified as YSO candidates in MGM (false positives). Some even show bright spiral galaxies, for example MGM 1261 and MGM 1517. In the latter case, one can see two bright star-forming regions in one spiral arm, that were both classified as disk candidates previously. Concerning faint and less prominent extra-galactic objects, the VISTA extension flags are used additionally to identify these as extended sources. In combination with the position at low extinction ( $A_{K,Herschel} \lesssim 1$  mag), we can rule out that these sources are embedded protostar. Moreover, visual inspection at various wavelengths helps to rule out that the extension is not caused by outflows, surrounding a protostellar candidate. The color and magnitude diagrams (on the right) further help to confirm the nature of some sources. Galaxies tend to have red colors, similar to YSOs, but are also mostly fainter and are located close to the galaxy clump in the CMDs (last column, bottom right over-densities).

In Figs. A.4 and A.5 we show examples of VISTA identified fuzzy objects, like nebulosities, cloud edges, and Herbig-Haro objects, that were erroneously classified as YSO candidates previously. Here, the *Spitzer* cutouts are important, to rule out that the source is not an embedded source that only gets point-like at longer wavelengths. Some sources indeed almost show a point-like structure in the MIR, for example, source M 2415 in the M1 cutout. Though, it is likely caused by heated material, maybe due to a shocked outflow, appearing in blue in the VISTA image. The extinction at the cite of this source is on the order of  $A_{K,Herschel} \sim 0.24$  mag, which makes it unlikely that it is a deeply embedded source, because then it should also show up in I4 where no point-source can be identified. The CCDs and CMDs are shown for completeness, however, fuzzy nebulous detections which seem point-like at some wavelengths and resolutions are mostly not detected in all bands which construct the here presented diagrams.



**Fig. A.3.** Examples of galaxies which were classified as YSO candidates in MGM. From left to right: VISTA three color composite, *Spitzer* I2, I4, and M1 cutouts (50'' × 50''), oriented in galactic coordinates, and examples of six color-color and color-magnitude diagrams, showing the displayed source as red dot. See text for more explanations.



**Fig. A.4.** Examples of fuzzy objects (nebularities, cloud edges, or parts of Herbig Haro objects) which were classified as YSO candidates in MGM. See Fig. A.3 and text for more explanations.

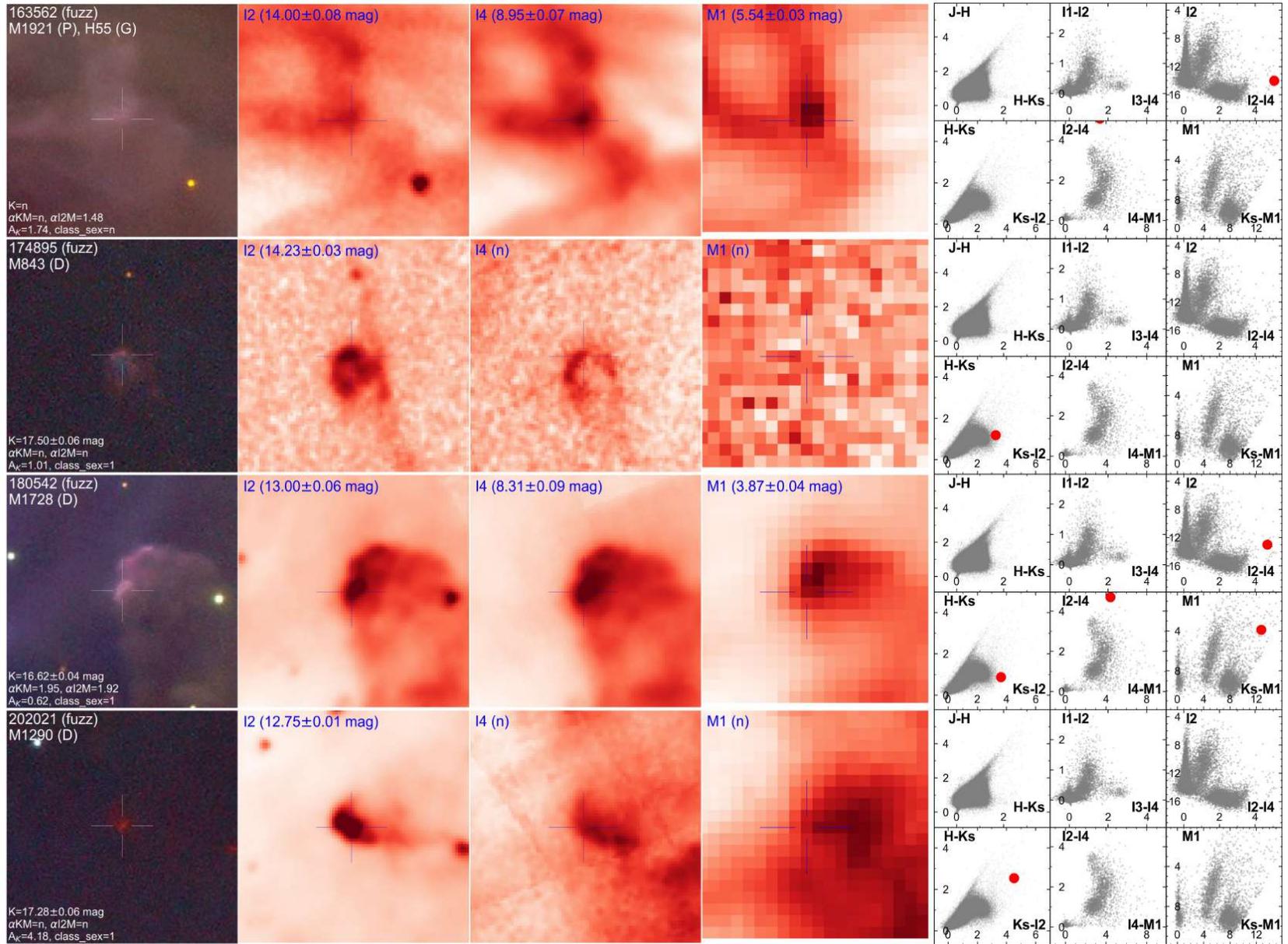


Fig. A.5. Same as Fig. A.4

## Appendix B: Selection conditions for new YSO candidates

Here we present the detailed description of the YSO selection methods for the new YSO candidates (see also Sect. 3.3). We use six individual selection criteria with different band combinations to exploit various color-color and color-magnitude spaces, and the different sensitivities of the available bands. Three criteria are used to select sources inside the IRAC observed region, and the other three for outside IRAC. To summarize the basic procedure for each selection, we 1) apply individual error-cuts to get rid of inferior photometry, 2) use a color-magnitude diagram to exclude faint unresolved star-forming galaxies and the majority of active galactic nuclei (AGN-cut), 3) use a color-color diagram to exclude MS-stars and reddened source due to extinction, and finally 4) cut further color regions that are confused with contaminants. Such contaminants can be shocked blobs of gas outflows around young stars (influencing  $I2$  or  $W2$ ), or bright extended nebula emission, especially influencing e.g.  $W2 - W3$  or  $W2 - W4$  colors.

We decide the selection conditions by comparing with control field plots, when available, and by checking the color spaces of known YSOs and contaminating objects, found in Sect. 3.2 (see Figs. 3 and 4). With this we try to get a high recovery rate and at the same time avoid contamination as good as possible. The following figures display in gray all sources left after applying basic error-cuts. Diagrams for selections inside the IRAC region show the recovered protostars and disks in red and blue respectively, and new candidates are highlighted with yellow diamonds. New candidates selected outside IRAC in combination with *WISE* are shown separately as blue open squares. Selected sources that turned out to be false positives, or that are uncertain candidates, are marked with a black cross.

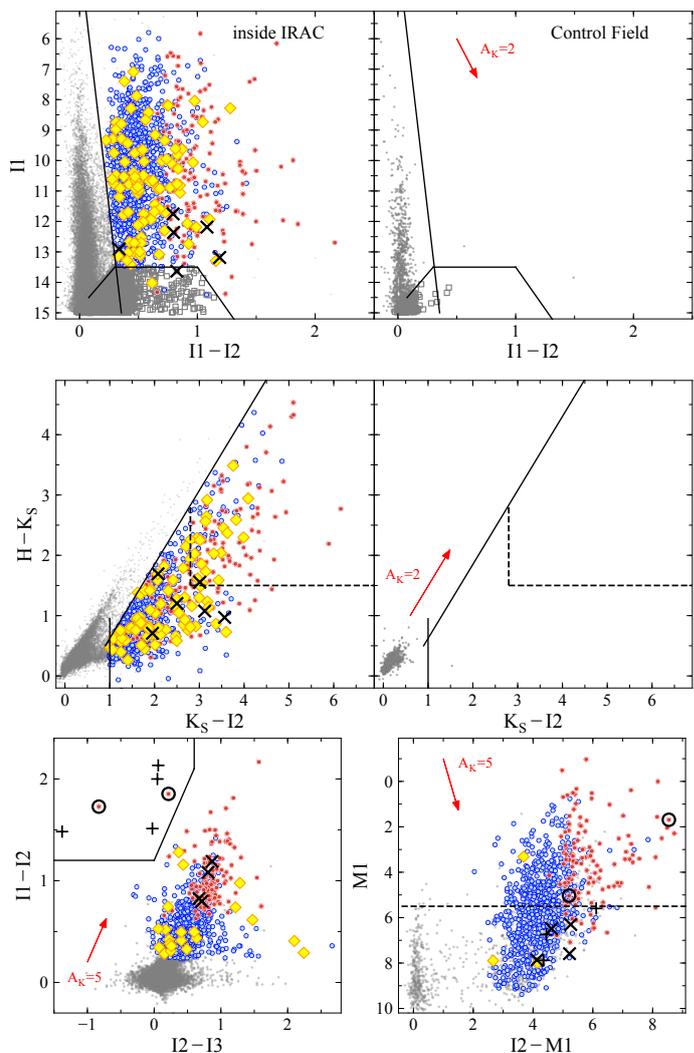
### Appendix B.1: Selections using VISTA and Spitzer

Here we present selections based on VISTA and *Spitzer* photometry, to add new YSO candidates inside and also outside the IRAC region (including  $M1$ ). Sources might have been missed previously, due to coverage, resolution and/or sensitivity issues in the NIR, or due to different selection criteria. To estimate the background contamination we use the control fields of IRAC that are overlapping with VISTA (see the small rectangular IRAC fields in Fig. 1). These contain in total much less sources compared to the IRAC science field ( $\sim 4\%$ ), therefore we also use the knowledge of the color spaces of known objects to decide the selection conditions.

#### Appendix B.1.1: VISTA and IRAC1-2

First, we use the higher sensitivity of VISTA and  $I1-2$ , compared to 2MASS and the other *Spitzer* bands. These bands are less affected by saturation and contamination near the ONC. However, YSO candidates are not as well separated when using this band combination compared to longer MIR bands. Sources with too little IR-excess will be missed, or are excluded by a reddening cut (parallel to extinction vector above the MS). We use the CMD  $I1 - I2$  vs  $I1$  and the CCD  $K_S - I2$  vs  $H - K_S$  (HKI12-selection, Fig. B.1), and apply the following error cuts:

$$\begin{aligned} Herr, K_S err, I1 err, I2 err &< 0.1 \text{ mag} \\ I1, I2 &< 14.5 \text{ mag} \\ \text{ClassSex} &> 0.1 \end{aligned} \quad (\text{B.1})$$



**Fig. B.1.** CCDs and CMDs showing the HKI12-selection. Top:  $I1$  vs  $I1 - I2$  color-magnitude diagram for the science and control field. Middle:  $H - K_S$  vs  $K_S - I2$  color-color diagram for the science and control field. The science fields (left) show the YSO selection inside the IRAC coverage (yellow filled diamonds), while the recovered sources are highlighted as well (disks - blue, protostars - red). Black crosses mark contamination or uncertain sources. The control fields (right) are used to exclude background contamination. The extinction vector is shown in red with a length of  $A_K = 2$  mag. Bottom: Shock emission exclusion for sources with valid  $I3$  measurements for the HKI12-selection. Symbols are the same as above in the science field plots. Additionally we show black plus symbols, marking sources that are caused by shock emission, and therefore are excluded. The open black circles mark sources which fall in the color region of shock emission, but which are likely YSO candidates, by showing a significant  $M1$  measurement.

The latter condition excludes extended objects, mostly galaxies but also sources surrounded by prominent outflows. This means we also exclude some protostar candidates with this criteria. However, we do not expect that there are still such sources left undiscovered, since protostars with prominent outflows are already selected by previous works. Furthermore, such prominent objects are eye-catchers, and are mostly also found when visually inspecting the images. After applying basic error cuts there are 21,641 and 953 sources left in the science and control field, respectively. Since the control field only contains about 4% compared to the science field, we also infer on the color spaces of known YSOs to decide the selection conditions. To exclude fur-

ther galaxies and faint unclear candidates we apply the following conditions (AGN-cut, see CMD in Fig. B.1 top row, lower borders). The sources beyond this region (gray boxes) are excluded in the following from the CCD.<sup>20</sup>

$$\begin{aligned} I1 &\leq 13.5 \vee \\ I1 &\leq -4.5 \times (I1 - I2 - 0.4) + 13.5 \vee \\ I1 &\leq 5.5 \times (I1 - I2 - 1) + 13.5 \end{aligned} \quad (\text{B.2})$$

To exclude MS stars we apply a cut in the CMD, and also in the CCD, where we cut parallel to the extinction vector.

$$\begin{aligned} I1 &\leq 33 \times (I1 - I2 + 0.1) \wedge \\ H - K_S &\leq 0.546/0.445 \times (K_S - I2 - 0.9) + 0.5 \wedge \\ K - I2 &\geq 1 \end{aligned} \quad (\text{B.3})$$

The selection, until this point, contains contaminating sources due to shock emission, which can be identified including  $I3$  in the CCD  $I1 - I2$  vs  $I2 - I3$ , similar to Gutermuth et al. (2009). Shock emission is located in the upper left corner of this diagram (see bottom row in Fig. B.1). We apply the following conditions, but only for sources with  $I3$  not NULL, and which are not in  $M1$  or with  $M1 > 5.5$  mag, to keep YSO candidates included, that show a significant  $M1$  measurement (bottom right plot in Fig. B.1).

$$\begin{aligned} I1 - I2 &\leq 1.5 \times (I1 - I3) + 1.2 \vee \\ I1 - I2 &\leq 1.2 \vee \\ I2 - I3 &\geq 0.6 \end{aligned} \quad (\text{B.4})$$

Finally, we add sources in the top right region of the HKI12 CCD (Fig. B.1 middle row, black dashed lines). This is a color region, that does not suffer from contamination. Therefore, we do not apply the AGN-cut here (Equ. B.2). This additional cut adds faint YSO candidates, that are located in the galaxy region of the CMD.

$$\begin{aligned} H - K_S &\geq 1.5 \wedge \\ K_S - I2 &\geq 2.8 \end{aligned} \quad (\text{B.5})$$

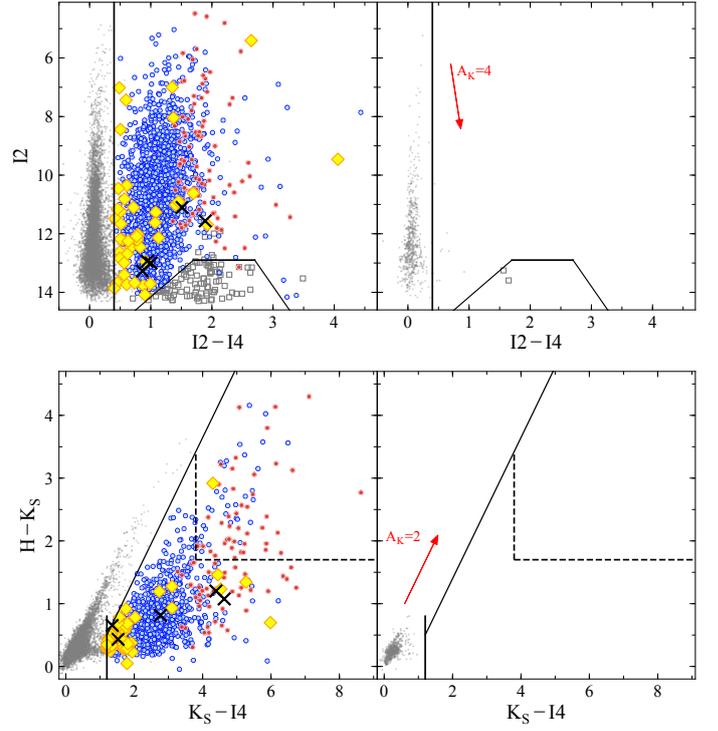
With the HKI12-selection we select 1,365 sources inside the IRAC region of which 1,270 are known YSO candidates from previous works, and 89 are new candidates (see map, yellow filled diamonds in Fig. B.10). The rest 6 are contaminating or uncertain sources. The HKI12-selection recovers  $\sim 47\%$  of the known YSO population.

#### Appendix B.1.2: VISTA and IRAC2-4

Next we combine VISTA and *Spitzer*/ $I2-4$ , to primarily look for possible missed disk candidates, by using different color spaces compared to previous works. We use the CMD  $I2 - I4$  vs  $I2$  and the CCD  $K_S - I4$  vs  $H - K_S$  (HKI24-selection, Fig. B.2) with a similar procedure as for to the HKI12-selection above. We apply the following basic error cuts:

$$\begin{aligned} Herr, K_Serr, I2err, I4err &< 0.1 \text{ mag} \\ I1, I2 &< 14.3 \text{ mag} \\ \text{ClassSex} &> 0.1 \\ \text{ClassCog} &= 1 \end{aligned} \quad (\text{B.6})$$

With this there are 9316 sources left in the science field and only 539 in the control field ( $\sim 6\%$ ). To exclude galaxies and faint



**Fig. B.2.** CCDs and CMDs showing the HKI14-selection. Top:  $I2$  vs  $I2 - I4$  color-magnitude diagram for science and control field. Bottom:  $H - K_S$  vs  $K_S - I4$  color-color diagram for science and control field. Symbols as in Fig. B.1.

unclear candidates we apply the following conditions (AGN-cut, CMD in Fig. B.2 top row, lower trapezoidal borders).

$$\begin{aligned} K_S &< 15.5 \wedge \\ (I2 &\leq 12.9 \vee \\ I2 &\leq -1.8 \times (I2 - I4 - 1.7) + 12.9 \vee \\ I2 &\leq 3 \times (I2 - I4 - 2.7) + 12.9) \end{aligned} \quad (\text{B.7})$$

The sources beyond these borders (gray boxes) are excluded from the CCD. To exclude MS stars we apply cuts in the CMD and in the CCD, where we cut parallel to the extinction vector.

$$\begin{aligned} I2 - I4 &> 0.4 \wedge K_S - I4 \geq 1.2 \wedge \\ H - K_S &\leq 0.546/0.486 \times (K_S - I4 - 1.2) + 0.5 \end{aligned} \quad (\text{B.8})$$

Again, we add sources in the top right region of the HKI24 CCD (Fig. B.2 bottom row, black dashed lines), by not applying the AGN-cut (Equ. B.7).

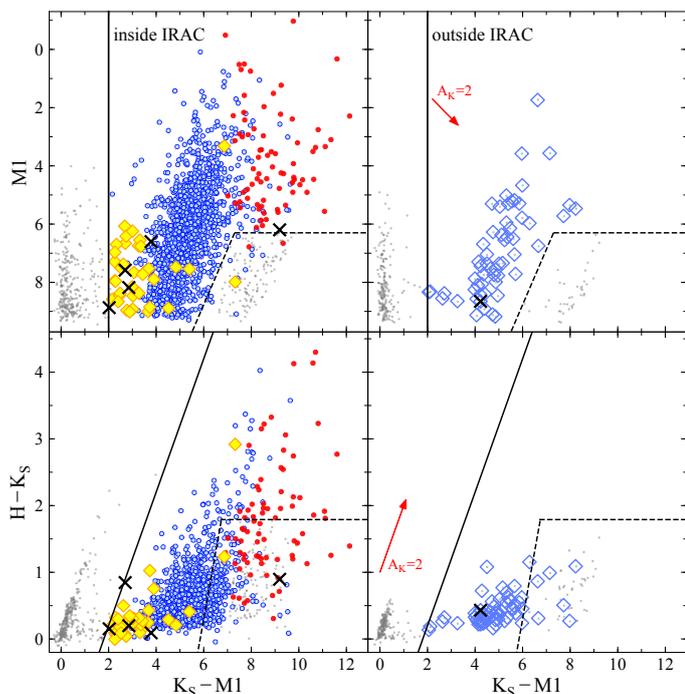
$$H - K_S \geq 1.7 \wedge K_S - I4 \geq 3.8 \quad (\text{B.9})$$

We select 1,675 sources, of which 1,626 are known YSO candidates from previous works, and 44 are new candidates (see map, yellow filled triangles in Fig. B.10). The rest 5 were found to be false positives or are uncertain candidates. Out of the 44 only 9 were already selected by HKI12. The 35 additional candidates add up to 124 new candidates inside IRAC until this point (HKI12-HKI24-selection). The HKI24-selection alone recovers  $\sim 60\%$  of the known YSO population.

#### Appendix B.1.3: VISTA and $M1$

We finally combine VISTA and *Spitzer*/ $M1$  to select new YSO candidates inside and also outside the IRAC region. This band

<sup>20</sup> The symbols  $\wedge$  and  $\vee$  stand for the logical AND and OR, respectively.



**Fig. B.3.** Top: CMD  $M1$  vs  $K_S - M1$ . Bottom: CCD  $H - K_S$  vs  $K_S - M1$ . Left: Selection of YSOs inside the IRAC coverage, showing the recovered sources (disks - blue, protostars - red). Yellow filled diamonds are new YSO candidates and the black crosses are contaminants or uncertain sources. Right: Selected new YSO candidates outside the IRAC region (open blue diamonds). The extinction vector is shown for both diagrams in red with a length of  $A_K = 2$  mag. The black solid lines represent cuts to exclude MS stars, and the black dashed lines exclude extra-galactic sources.

combination is sensitive to anemic and transition disks, which often show only an IR-excess long-wards of about 10 to 20  $\mu\text{m}$ . We use the CMD  $K_S - M1$  vs  $M1$  and the CCD  $K_S - M1$  vs  $H - K_S$  (HKM-selection). There is no control field for this band combination, therefore, we decide the selection conditions solely based on known objects. We apply the following basic error cuts:

$$\begin{aligned} K_S \text{err}, H \text{err} &< 0.1 \text{ mag} \\ M1 \text{err} &< 0.15 \text{ mag} \\ K_S &< 17 \text{ mag} \\ \text{ClassSex} &> 0.1 \\ \text{ClassCog} &= 1 \end{aligned} \quad (\text{B.10})$$

This leaves 1,793 sources inside and 271 outside the IRAC region. To exclude MS stars we apply rather conservative cuts in both the CMD and CCD (black solid lines in Fig. B.3), to exclude possible fake  $M1$  detections that give some sources a false IR-excess. Also AGBs are particularly influencing this region.

$$\begin{aligned} K_S - M1 &> 2 \wedge \\ H - K_S &\leq 0.55/0.48 \times (K_S - M1 - 1.3) - 0.5 \end{aligned} \quad (\text{B.11})$$

Next we cut extra-galactic contamination in the lower right corner of the CMD, shown by the black dashed lines in the upper panels of Fig. B.3.

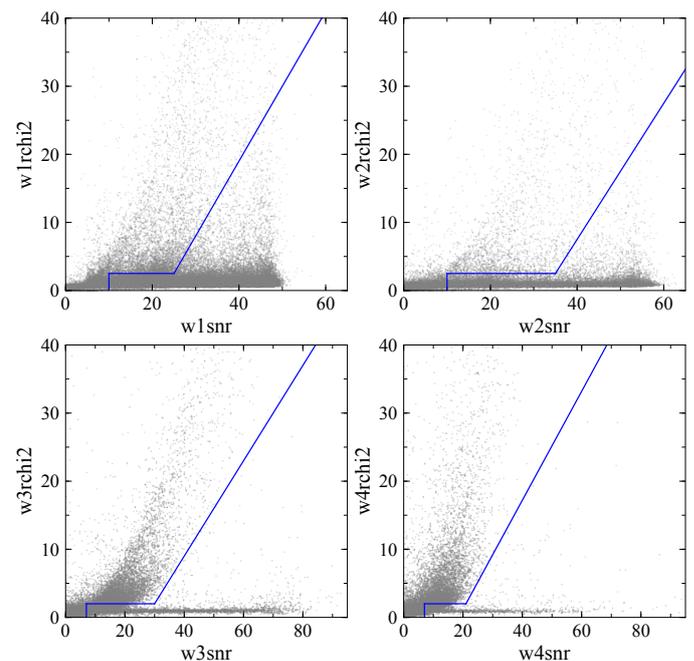
$$\begin{aligned} M1 &\leq 6.25 \vee \\ M1 &\leq -1.9 \times (K_S - M1 - 7.3) + 6.3 \end{aligned} \quad (\text{B.12})$$

The remaining sources are selected as YSO candidates. We add further candidates in the CCD but exclude sources right to the dashed lines in the lower panels of Fig. B.3, which is again contaminated by extra-galactic sources.

$$\begin{aligned} H - K_S &> 1.79 \vee \\ H - K_S &\leq 2 \times (K_S - M1 - 5.6) - 0.5 \end{aligned} \quad (\text{B.13})$$

We select 1,387 sources inside the IRAC region, of which 1,346 are known YSO candidates from previous works, and 36 are new candidates (see map, cyan open diamonds in Fig. B.10). The rest 5 were found to be false positives or uncertain candidates. Out of the 36 new disk candidates 8 are overlaps with HKI12-HKI24. The 28 extra sources were not picked up previously, likely because they show no or very little IR-excess in the shorter wavelength bands. Out of the 36 there are 31 anemic disk candidates. The combined HKI12-HKI24-HKM-selections give a total of 152 new YSO candidates inside the IRAC region. Outside the IRAC region we select 56 new YSO candidates using HKM1 (see map, cyan filled diamonds in Fig. B.10), including 8 transition disks and 5 anemic disks. There is only one uncertain source selected outside IRAC (black cross), which lies at the border of the MIPS1 coverage. The HKM-selection recovers about 50% of the known YSO population.

## Appendix B.2: Selections using VISTA and WISE



**Fig. B.4.** SNR versus reduced chi square for each WISE band. The blue lines show the cuts (SNR-RCHI2-cut) to reduce fake point sources (similar to Koenig & Leisawitz 2014). Sources to the left of the blue borders are excluded, as given in Eqs. B.14 to B.17.

We use WISE MIR data to add additional YSO candidates outside the IRAC region, keeping the MIPS region included. To estimate the background we use the VISTA control field cross-matched with WISE data. Koenig & Leisawitz (2014) presented a detailed analysis of different possible cleaning processes for AllWISE photometry in the galactic plane, to mitigate the high contamination due to fake point sources. They find that a combination of signal to noise ratio (SNR) and reduced chi square

(RCHI2) gives the best separation between fake and true point sources. We adopt their approach but modify it to be less conservative, and to recover more YSO candidates. The following conditions are applied for *WISE* bands if they are used in one of the selections presented below. Fig. B.4 shows the SNR-RCHI2-cut as blue solid lines, while sources to the left of these borders are excluded.

$$w1snr > 10 \wedge (w1rchi2 < 2.5 \vee w1rchi2 < 1.1 \times (w1snr - 25) + 2.5) \quad (\text{B.14})$$

$$w2snr > 10 \wedge (w2rchi2 < 2.5 \vee w2rchi2 < 1 \times (w2snr - 35) + 2.5) \quad (\text{B.15})$$

$$w3snr > 7 \wedge (w3rchi2 < 2 \vee w3rchi2 < 0.5 \times (w3snr + 28) - 2) \quad (\text{B.16})$$

$$w4snr > 7 \wedge (w4rchi2 < 2 \vee w4rchi2 < 0.8 \times (w4snr + 21) - 2) \quad (\text{B.17})$$

The SNR is directly correlated to the magnitude error ( $w?sigmpro$ ), hence, an additional error-cut with the latter is not necessary. Still, we require that these errors should not be NULL values.

$$w?sigmpro \text{ not NULL} \quad (\text{B.18})$$

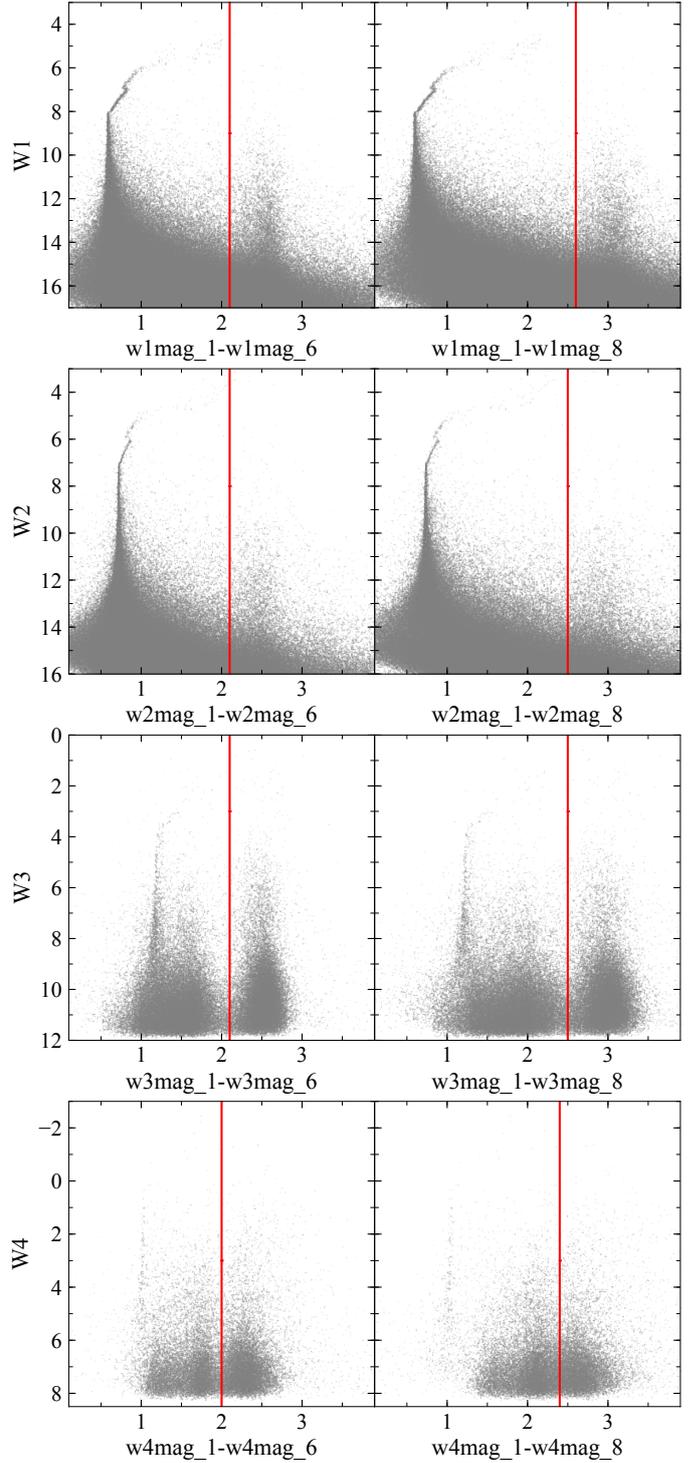
To mitigate further contamination due to extended emission we use *WISE* aperture photometry, which is provided by the AllWISE catalog for eight different aperture sizes (Cutri et al. 2013). Extended emission in the MIR is mainly caused by nebulosities, that can give an erroneous IR-excess. Especially the *W3* band shows high contamination due to PAH emission. The aperture cuts were chosen using information from figure 2 in Meisner & Finkbeiner (2014), where they show the curve-of-growth for the first six apertures of the *W3* band. They use test sources of known type (extended or point like) and get a separation at about 2 when using  $w3mag_1 - w3mag_6$ . We use the 6th and also the 8th (the largest) aperture, while the latter is not used by Meisner & Finkbeiner (2014). To decide the final separation we plot  $w?mag_1 - w?mag_6$  and  $w?mag_1 - w?mag_8$  versus the magnitude of each band, shown in Fig. B.5. In most of these diagrams, one can see an over-density of brighter sources, beyond  $\sim 2$ , best visible for the *W3* band, caused by bright nebulous structures. The *W1* and *W2* bands show a less prominent tip. Sources that do not satisfy the following conditions will be excluded if the *WISE* band is used in a selection. The conditions and are shown as red solid lines in Fig. B.5.

$$w1mag_1 - w1mag_6 < 2.1 \wedge w1mag_1 - w1mag_8 < 2.6 \quad (\text{B.19})$$

$$w2mag_1 - w2mag_6 < 2.1 \wedge w2mag_1 - w2mag_8 < 2.5 \quad (\text{B.20})$$

$$w3mag_1 - w3mag_6 < 2.1 \wedge w3mag_1 - w3mag_8 < 2.5 \quad (\text{B.21})$$

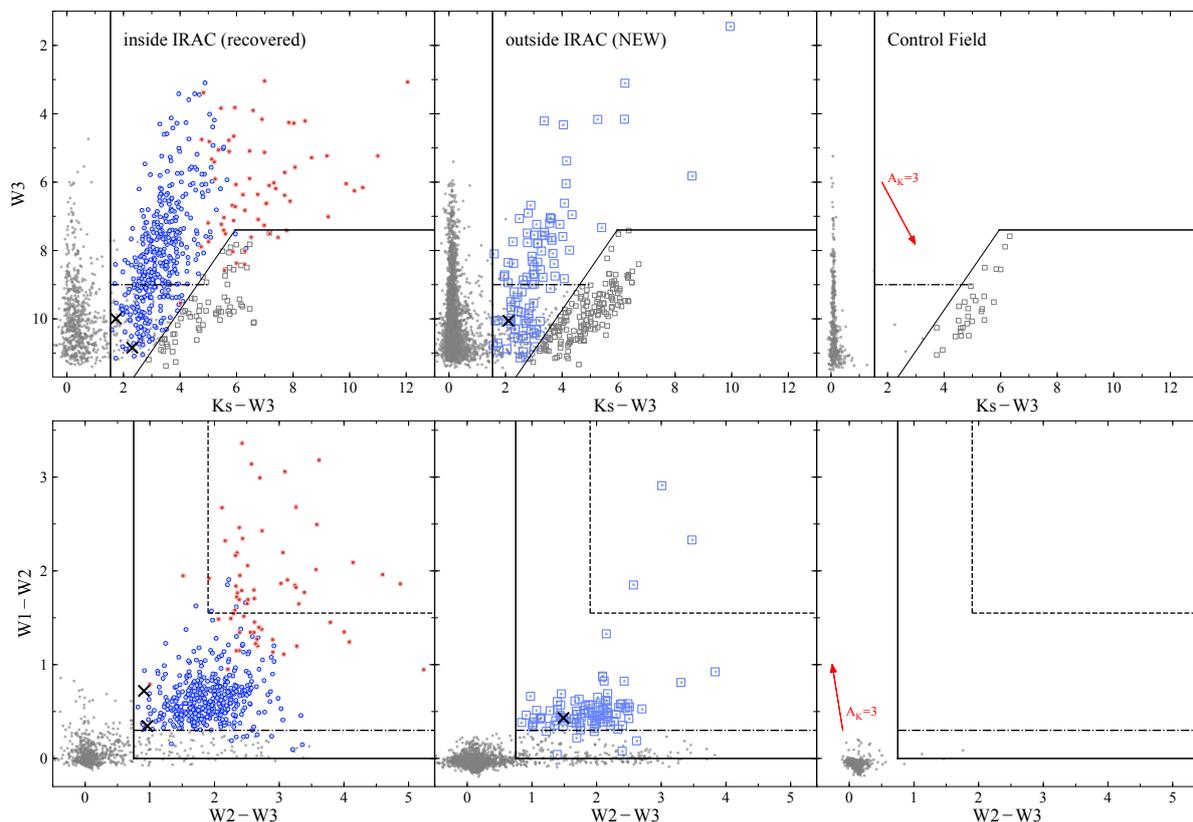
$$w4mag_1 - w4mag_6 < 2.0 \wedge w4mag_1 - w4mag_8 < 2.4 \quad (\text{B.22})$$



**Fig. B.5.** *WISE* aperture photometry cuts to reduce extended source contamination. Gray dots are all *WISE* sources after applying basic error-cuts:  $w?snr > 5$  and  $w?sig < 0.5$ . The red solid lines show the criteria to separate point sources (left) from extended sources (right), given in Eqs. B.19 to B.22.

Furthermore, we check if the sources are reliable detections by using the AllWISE cataloged number of individual exposures for a given band, where a profile-fit measurement of the source was possible ( $w?m$ ). We combine this with  $w?nm$ , which gives the number of times it was detected with a SNR  $> 3$ . We require this should be at least 20% when compared to  $w?m$ :

$$w?nm/w?m \geq 0.2 \quad (\text{B.23})$$



**Fig. B.6.** W123-selection. Left: Orion A science field inside the IRAC coverage for L1641, showing recovered YSO candidates (blue - disks, red - protostars). Center: Orion A science field outside the IRAC coverage showing new YSO candidates (blue open squares). Right: Control field showing background and MS star contamination. Top:  $K_S - W3$  vs  $W3$  CMD. To clean the sample of extra-galactic contamination we exclude sources below and right to the black lines (AGN-cut, gray open boxes). The left border reduces contaminated  $W3$  photometry and eliminates MS stars. Bottom:  $W2 - W3$  vs  $W1 - W2$  CCD. The black lines show the selection conditions. Sources from the AGN-cut are already removed. The slope on the left separates MS stars. The cut at the bottom reduces contaminated photometry. The dashed black lines in the upper right corner show an additional selection of sources where the AGN-cut was not applied due to the fact that this region is free of any contamination. The horizontal black dashed lines in both diagrams show the exclusion condition for faint sources ( $W3 > 9$  mag) which at the same time fall below  $W1 - W2 < 0.4$ .

This cut excludes mainly faint sources which are mostly already excluded by the SNR-RCHI2-cut. Finally, we exclude sources listed as contaminated or confused by artifacts (`cc_flags`<sup>21</sup>), allowing only 0 or lower-case letters (d, p, h, o). In the following *WISE* based selections we apply the discussed cuts only to those bands used for a selection.

Further we over-plot known objects from the already analyzed IRAC region to get an idea of the color-spaces of YSOs and contaminating objects in the diagrams (see left plots in Figs. B.6, B.7, B.9), as it was done for the *Spitzer* based selections above. However, we use only known objects from the L1641 region ( $l > 210^\circ$ ), which is a region less influenced by fuzzy nebulous contamination. Otherwise, if including the ONC region, we would need to apply more conservative error cuts.

#### Appendix B.2.1: VISTA and W123

First we use a combination of the first three *WISE* bands and VISTA  $K_S$  to investigate the CCD  $W2 - W3$  vs  $W1 - W2$  and the

CMD  $K_S - W3$  vs  $W3$  (W123-selection, Fig. B.6). Including the  $K_S$  band reduces contamination due to fake *WISE* point-sources. Additional to the *WISE* error-cuts presented above we further apply the following:

$$\begin{aligned} K_S \text{err} &< 0.1 \text{ mag} \\ W1, W2 &< 14 \text{ mag} \end{aligned} \quad (\text{B.24})$$

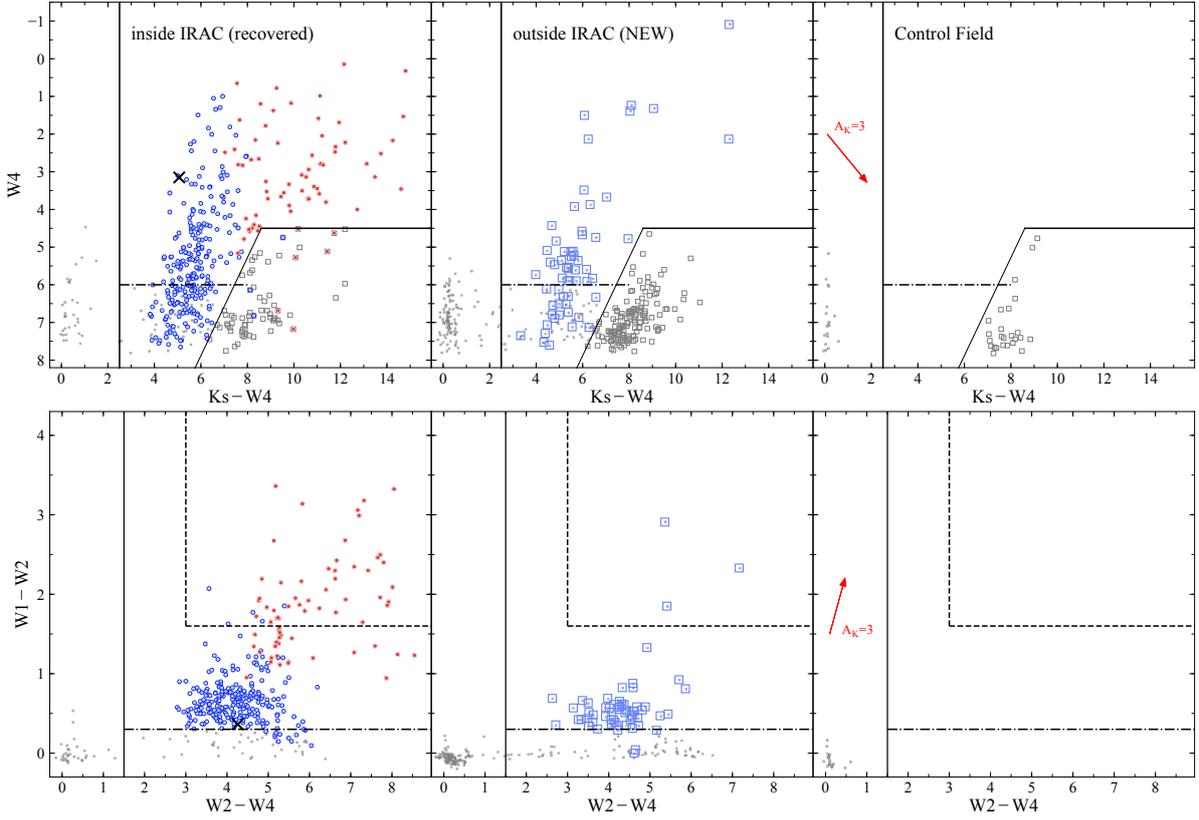
With this there are 1,910 sources left in the IRAC region at L1641, 2,276 outside the IRAC region, and 350 in the control field. To eliminate extra-galactic contamination we exclude sources in the CMD, where AGNs and galaxies with PAH emission are located at the lower right of this diagram at about  $K_S - W3 \approx 5$  (see control field, AGN-cut, Fig. B.6, gray boxes). Therefore all source that do not fulfill the following conditions are removed from the CCD.

$$\begin{aligned} W3 &\leq 7.4 \text{ mag} \vee \\ W3 &\leq -1.2 \cdot (K_S - W3 - 5.95) + 7.4 \end{aligned} \quad (\text{B.25})$$

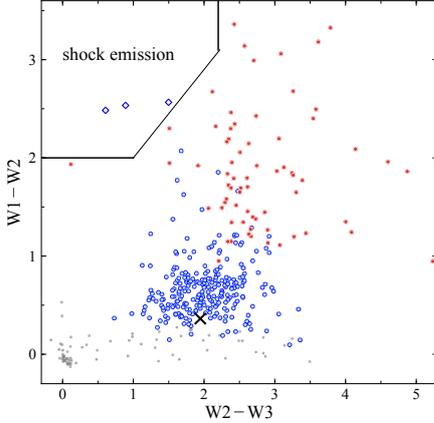
To exclude MS stars and reddened sources we apply the following conditions, which are the solid vertical lines on the left in the CMD and CCD plots,

$$\begin{aligned} K_S - W3 &\geq 1.54 \wedge \\ W1 - W2 &\geq 0.75 \end{aligned} \quad (\text{B.26})$$

<sup>21</sup> Contamination and confusion flag due to proximity to an image artifact indicated by four flags: diffraction spike (d, D), persistence of a short-term latent image (p, P), halo (h, H), and optical ghost (o, O). 0 indicates not-confused photometry. Upper-case letters denote spurious detection of an artefact and lower-case letters denote that the measurement may be real but might be contaminated by the artefact.



**Fig. B.7.** W124-selection. See caption of Fig. B.6 for more explanations.



**Fig. B.8.** W123 CCD showing the cut to exclude shock emission from the W124 selection.

and we cut sources at the very bottom of the CCD, which is a color region highly contaminated by fake  $W3$  point-sources,

$$W1 - W2 > 0. \quad (\text{B.27})$$

To further mitigate such contaminants we apply a special cut for faint sources, shown by the two horizontal dash-dotted lines in the CMD and CCD. For faint sources with  $W3 > 9$  mag we require  $W1 - W2 \geq 0.3$ , since this is a color region less influence by contaminated photometry. The black dashed borders at the top right of the CCD are used to re-add candidates. Similar to the HK12-selection, this region is found not to be effected by galaxy contamination, and the following conditions are applied

without the AGN-cut (Equ. B.25):

$$\begin{aligned} W1 - W2 &\geq 1.55 \wedge \\ W2 - W3 &\geq 1.9 \end{aligned} \quad (\text{B.28})$$

With these conditions we select 97 new YSO candidates outside the IRAC region, as shown by the blue open boxes in the middle plots of Figure B.6. Inside the IRAC region at L1641 (left plots) we select 456 YSO candidates, of which 454 are recovered known YSOs, and two are contaminants due to extended emission in the  $W3$  band. In the whole Orion A IRAC region the W123-selection recovers about 28% of known YSO candidates, while in L1641 the recovery rate lies at about 51%.

#### Appendix B.2.2: VISTA and W124

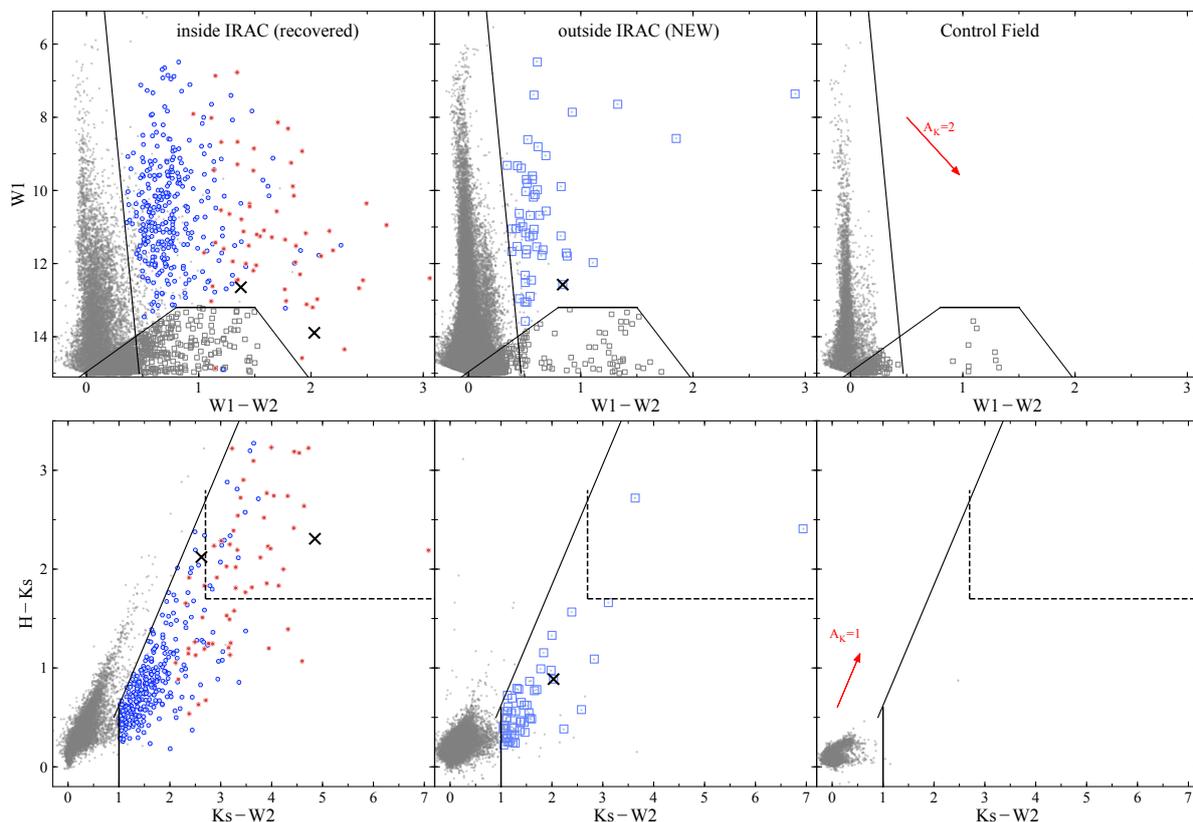
Next we add sources by investigating the CCD  $W2 - W4$  vs  $W1 - W2$  and the CMD  $K_S - W4$  vs  $W4$  (W124-selection, Fig. B.7), with a very similar approach to the W123-selection. Beside the main basic error cuts we require:

$$\begin{aligned} K_S \text{err} &< 0.1 \text{ mag} \\ W1, W2 &< 15 \text{ mag} \end{aligned} \quad (\text{B.29})$$

With this there are 449 sources left in the IRAC region at L1641, 363 outside the IRAC region, and only 44 in the control field. We apply again an AGN-cut using the CMD (Fig. B.7, gray boxes):

$$\begin{aligned} W4 &\leq 4.5 \vee \\ W4 &\leq -1.6 \times (K_S - W4 - 8.6) + 4.5 \end{aligned} \quad (\text{B.30})$$

as shown by the solid black lines in the bottom right of the CMD. The sources beyond these borders (gray boxes) will be removed



**Fig. B.9.** HKW12-selection. See also caption of Fig. B.6. Top:  $W1 - W2$  vs  $W1$  CMD. To clean the sample of extra-galactic contamination we exclude sources below the black bottom lines, indicating the AGN-cut (gray boxes). Bottom:  $K_S - W2$  vs  $H - K_S$  CCD. The black lines show the selection conditions, and the black dashed lines the additional condition for a region with no AGN contamination. This gives a selection of 52 new YSO candidates outside the IRAC region.

from the CCD. To exclude MS stars and reddened sources due to extinction we apply the following conditions:

$$\begin{aligned} K_S - W4 &\geq 2.5 \wedge \\ W2 - W4 &\geq 1.5 \end{aligned} \quad (\text{B.31})$$

Again we apply a special cut for faint source to mitigate contaminated  $W4$  photometry. We require for faint sources with  $W4 \geq 6$  mag that they should be redder than  $W1 - W2 \geq 0.3$ , shown by the two horizontal dash-dotted lines in the CMD and CCD. Furthermore, we re-add again sources in the top right corner of the CCD (black dashed borders), therefore, the following conditions are applied without the AGN-cut (Equ. B.30):

$$\begin{aligned} W1 - W2 &\geq 1.6 \wedge \\ W2 - W4 &\geq 3 \end{aligned} \quad (\text{B.32})$$

This region is not contaminated by extra-galactic sources, however, there is still some contamination due to shock emission. Similar to the HKI12-selection we require an additional cut to exclude this fuzzy contamination, which can be identified using the  $W123$  CCD (Fig. B.8). Such sources are located in the top left corner of this diagram. We will apply this only to sources with a valid  $W3$  measurement ( $W3$  not NULL).

$$\begin{aligned} W1 - W2 &\leq 2 \vee \\ W1 - W2 &\leq 0.9 \times (W2 - W3 - 1) + 2 \vee \\ (W1 - W2 &\geq 2 \wedge W2 - W3 > 2.2) \end{aligned} \quad (\text{B.33})$$

With the  $W124$ -selection we get 59 new YSO candidates outside IRAC (middle plots). Inside the IRAC region (left plots) we

select 321 YSO candidates, of which 320 are recovered known YSOs. One was found to be contamination due to a double-star, which was erroneously cross-matched. In the whole Orion A region this selection recovers only about 18% of known YSO candidates, while for L1641 the recovery rate is about 36%. The combined  $W123$ - $W124$  selection recovers about 54% in L1641.

### Appendix B.2.3: VISTA and W12

Finally we use the CCD  $K_S - W2$  vs  $H - K_S$  and the CMD  $W1 - W2$  vs  $W1$  (HKW12-selection, Fig. B.9), which is a similar selection as the HKI12-selection. It aims to add sources that might have been missed by the  $W123$  or  $W124$  selections due to the higher contamination in the longer wavelength bands. Additionally to the basic *WISE* cuts we apply the following error-cuts:

$$\begin{aligned} K_S \text{err}, H \text{err} &< 0.1 \\ W1, W2 &< 14.5 \text{ mag} \\ \text{ClassSex} &> 0.1 \end{aligned} \quad (\text{B.34})$$

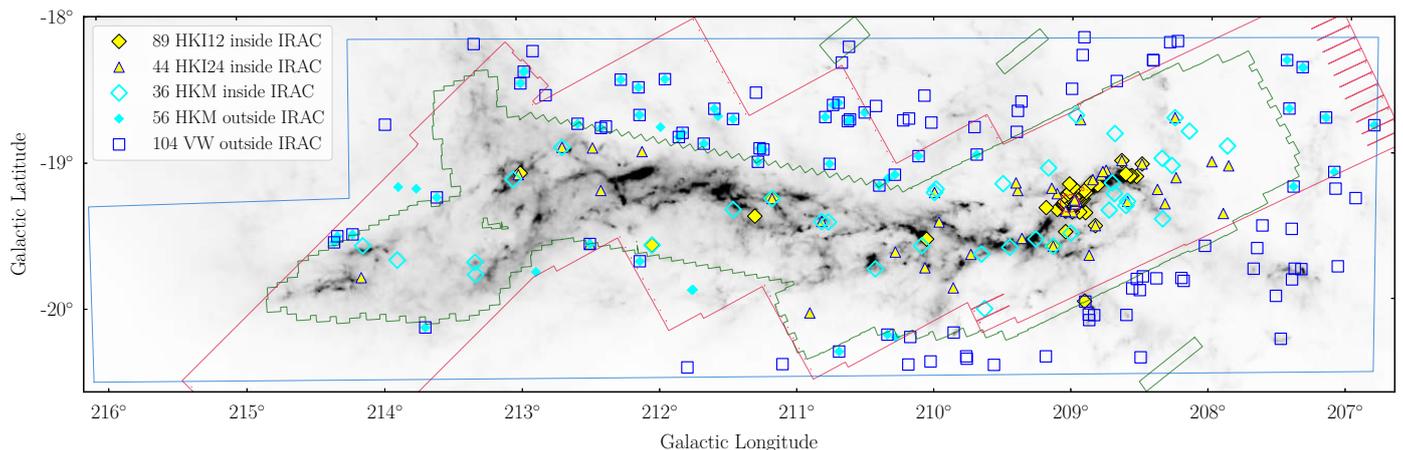
There are 8,325 sources left inside the IRAC region in L1641 (14,641 in whole Orion A), 27,610 outside, and 4,611 in the control field. To get rid of faint uncertain sources and extra-galactic contamination we again apply a cut in the CMD (AGN-cut, Fig. B.9), as shown by black solid lines at the bottom (trapezoidal shape).

$$\begin{aligned} W1 &\leq 13.2 \vee \\ W1 &\leq -2.2 \times (W1 - W2 - 0.8) + 13.2 \vee \\ W1 &\leq 4 \times (W1 - W2 - 1.5) + 13.2 \end{aligned} \quad (\text{B.35})$$

**Table B.1.** Recovery rates for the six individual YSO selection conditions as presented in this Appendix.

Class	L1641 <sup>a</sup>	VISTA/ <i>Spitzer</i> (VS)				VISTA/ <i>WISE</i> (VW)				Total
	known	HKI12	HKI24	HKM	total	W123	W124	HKW12	total	VS+VW
ALL	880	47%	70%	68%	82%	51%	36%	41%	59%	86%
D	665	45%	81%	78%	91%	55%	35%	42%	62%	92%
F	101	81%	63%	61%	83%	51%	51%	65%	69%	88%
P	114	30%	13%	14%	31%	30%	33%	21%	38%	48%

**Notes.** The recovery check is applied only on sources located in L1641 inside the IRAC coverage at  $l > 210^\circ$ , to avoid high contaminated regions close to the ONC. It is given in % compared to known YSO candidates. <sup>(a)</sup> Number of known YSO candidates (revisited MGM and FFA16 sample) in L1641.



**Fig. B.10.** The distribution of the 268 new YSO candidates selected with the VISTA/*WISE*/*Spitzer* methods. The sources found in- and outside the IRAC coverage are shown separately with different symbols (see legend). Outside the IRAC region: blue open box symbols - new VISTA/*WISE* (VW) YSO candidates; cyan filled diamonds - new VISTA/*M1* (HKM) YSO candidates. Inside the IRAC region: yellow diamonds - new VISTA/*I12* (HKI12) YSO candidates; yellow triangles - new VISTA/*I24* (HKI24) YSO candidates; cyan open diamonds - new VISTA/*M1* (HKM) YSO candidates.

Again, the sources beyond these borders (gray boxes) will be removed from the CCD. To exclude MS stars we apply the following conditions:

$$\begin{aligned}
 W1 &\geq 32 \times (W1 - W2) \wedge \\
 K_S - W2 &\geq 1 \wedge \\
 H - K_S &\leq 0.546/0.448 \times (K_S - W2 - 0.9) + 0.5
 \end{aligned}
 \tag{B.36}$$

These conditions are shown by the solid lines on the left in the CMD and CCD. The latter is a slope parallel to the reddening vector in the CCD, excluding reddened MS-stars due to extinction. To re-add YSO candidates which fall in the AGN-cut (Equ. B.35) we again introduce a selection at the top right part of the CCD, indicated by the black dashed lines:

$$\begin{aligned}
 H - K_S &> 1.7 \wedge \\
 K_S - W2 &> 2.7
 \end{aligned}
 \tag{B.37}$$

With the HKW12 selection we select 52 new YSO candidates outside the IRAC region, as shown by the blue open boxes in the middle plots of Figure B.7. Inside the IRAC region (left plots) we select 364 sources, of which 362 are recovered known YSO candidates, and two are uncertain or contaminating sources. Inside the whole Orion A IRAC region we recover about 27% of the known YSO candidates with the HKW12 selection, while the recovery rate for the L1641 region is about 41%. There are 47 overlaps with the W123-W124 selections outside IRAC, therefore we find only 5 new candidates in the surroundings with

the HKW12 selection. However, the recovery rate in L1641 improves to 59% when combining this selection with the two previous ones. In total we are able to add 104 new YSO candidates with the *WISE* based selections (W123-W124-HKW12) outside IRAC (see blue open squares on the map in Fig. B.10.).

### Appendix B.3: Recovery rate of the YSO selections

We test the recovery rate for each individual selection by comparing with known YSO candidates (from Sect. 3.2) in L1641 ( $l > 210^\circ$ ) inside the IRAC region. The low-mass star forming region L1641 is better comparable to regions in the surroundings (outside IRAC), being overall less affected by contamination due to extended emission or crowding. Especially for a *WISE* based selection this comparison makes more sense, since *WISE* shows large saturated regions near the ONC. The individual recovery rates are summarized in Table B.1. We also give individual recovery rates for the three YSO classes as classified in this work.

The combined *Spitzer* based selections (VS-selection) recover about 82% of the known YSO candidates in L1641, or about 80% when including the ONC. The rest 20% are likely missed due to the chosen error cuts and selection conditions. We like to note that these conditions were not designed to redo the selection for the IRAC region, but rather to choose color spaces that can provide additional candidates. The combined *WISE* selections (VW-selection) recover together about 59% in the region of L1641 inside IRAC. Including the ONC the recovery rate drops to about 38%, highlighting the influence of massive

star-forming regions on low resolution MIR data. The whole selection (combining VISTA/*WISE/Spitzer*) recovers about 86% in L1641. When looking at the individual recovery rates for the three YSO classes, we see that our methods better recover disks and flat-spectrum sources, while they recover less than 50% of the protostars. By including the extension flags from VISTA, we are likely losing sources that are connected to outflows or nebulosities. Also, by including NIR data, highly embedded sources can be missed. Nevertheless, such sources are mostly already known, or are added by FIR data or visually (e.g. when there are prominent outflows).

### Appendix C: Table

In Table C.1 we provide the column information for the final table, which contains all sources discussed in this work (3117). The table gives basic information, including RA/Dec J2000 position, the VISION ID and other relevant identification numbers. The VISTA, *Spitzer*, and *WISE* magnitudes are listed with errors, and the spectral indices, which were used to classify the YSO candidates. Columns 87 to 91 give classifications from previous works when available. The latter two classifications from Fang et al. (2009, 2013) or Pillitteri et al. (2013) are not discussed in this work but are listed for completeness ( $C_{\text{Fang}}$ ,  $C_{\text{P13}}$ ). Column 93 (Class\_flag) separates the sources in four categories, labeled with the numbers “1,2,3,4”:

- 1 = revisited YSO candidates which were previously selected by MGM or FFA16 (2704)
- 2 = new YSO candidates (274)
- 3 = rejected candidates (96)
- 4 = uncertain candidates (43)

The number counts of each sample are given in parenthesis. We like to note that the rejected candidates include four sources (92+4) which were actually not listed as YSO candidates previously but as contamination by FFA16. We keep them in this final catalog for completeness.

**Table C.1.** Description of columns of the final catalog containing all sources discussed in this paper. The catalogue is only available in electronic form at the CDS.

Column Number	Column Name	Units	Description
1	RAJ2000	hh:mm:ss	Right Ascension taken from the reference as given in Col. 3.
2	DEJ2000	dd:mm:ss	Declination taken from the reference as given in Col. 3.
3	Ref		The Reference tells the origin of the coordinates. (1) VISTA coordinates, <a href="#">Meingast et al. (2016)</a> ; (2) <i>Spitzer</i> coordinates, <a href="#">Megeath et al. (2012)</a> ; (3) <i>Herschel</i> /PACS point source catalog coordinates, <a href="#">Pilbratt et al. (2010)</a> .
4	VISTA		VISTA identification number
5	WISE		WISE identification number
6	PACS		<i>Herschel</i> /PACS point source catalog (HPPSC) identification number
7	ID		Running identification number from this work
8	MGM		Source index from <a href="#">Megeath et al. (2012)</a> or <a href="#">Megeath et al. (2016)</a> for MGM sources
9	HOPS		Source index from <a href="#">Furlan et al. (2016)</a> for HOPS sources
10	Simbad_Name		Main identification as given in Simbad
11	Otype		Object Type as given in Simbad
12	Jmag	mag	J magnitude from VISTA
13	e_Jmag	mag	1 $\sigma$ error of J magnitude from VISTA
14	Hmag	mag	H magnitude from VISTA
15	e_Hmag	mag	1 $\sigma$ error of H magnitude from VISTA
16	Ksmag	mag	K <sub>S</sub> magnitude from VISTA
17	e_Ksmag	mag	1 $\sigma$ error of K <sub>S</sub> magnitude from VISTA
18	IRAC1	mag	IRAC1 magnitude from <i>Spitzer</i> /IRAC
19	e_IRAC1	mag	error of IRAC1 magnitude from <i>Spitzer</i> /IRAC
20	IRAC2	mag	IRAC2 magnitude from <i>Spitzer</i> /IRAC
21	e_IRAC2	mag	error of IRAC2 magnitude from <i>Spitzer</i> /IRAC
22	IRAC3	mag	IRAC3 magnitude from <i>Spitzer</i> /IRAC
23	e_IRAC3	mag	error of IRAC3 magnitude from <i>Spitzer</i> /IRAC
24	IRAC4	mag	IRAC4 magnitude from <i>Spitzer</i> /IRAC
25	e_IRAC4	mag	error of IRAC4 magnitude from <i>Spitzer</i> /IRAC
26	MIPS1	mag	MIPS1 magnitude from <i>Spitzer</i> /MIPS
27	e_MIPS1	mag	error of MIPS1 magnitude from <i>Spitzer</i> /MIPS
28	W1	mag	W1 magnitude given as <i>w1mpro</i> in the AllWISE catalog
29	e_W1	mag	error of W1 magnitude given as <i>w1sigmpro</i> in the AllWISE catalog
30	W2	mag	W2 magnitude given as <i>w2mpro</i> in the AllWISE catalog
31	e_W2	mag	error of W2 magnitude given as <i>w2sigmpro</i> in the AllWISE catalog
32	W3	mag	W3 magnitude given as <i>w3mpro</i> in the AllWISE catalog
33	e_W3	mag	error of W3 magnitude given as <i>w3sigmpro</i> in the AllWISE catalog
34	W4	mag	W4 magnitude given as <i>w4mpro</i> in the AllWISE catalog
35	e_W4	mag	error of W4 magnitude given as <i>w4sigmpro</i> in the AllWISE catalog
36	alpha_KM		Observed spectral index $\alpha_{KM}$ from 2.15 to 24 $\mu\text{m}$ covering VISTA/K <sub>S</sub> , the four IRAC bands, and MIPS1
37	e_alpha_KM		fitting error of $\alpha_{KM}$
38	alpha_IM		Observed spectral index $\alpha_{IM}$ from 3.6 to 24 $\mu\text{m}$ covering the four IRAC bands and MIPS1
39	e_alpha_IM		fitting error of $\alpha_{IM}$
40	alpha_IRAC		Observed spectral index $\alpha_{IRAC}$ from 3.6 to 8 $\mu\text{m}$ covering the four IRAC bands
41	e_alpha_IRAC		fitting error of $\alpha_{IRAC}$
42	alpha_I2M		Observed spectral index $\alpha_{I2M}$ from 4.5 to 24 $\mu\text{m}$ covering three IRAC bands and MIPS1
43	e_alpha_I2M		fitting error of $\alpha_{I2M}$
44	alpha_I3M		Observed spectral index $\alpha_{I3M}$ from 5.8 to 24 $\mu\text{m}$ covering two IRAC bands and MIPS1
45	e_alpha_I3M		fitting error of $\alpha_{I3M}$

**Table C.1.** — Continued

Column Number	Column Name	Units	Description
46	alpha_KI3		Observed spectral index $\alpha_{KI3}$ from 2.15 to 5.8 $\mu\text{m}$ covering VISTA/ $K_S$ and three IRAC bands
47	e_alpha_KI3		fitting error of $\alpha_{KI3}$
48	alpha_KW3		Observed spectral index $\alpha_{KW3}$ from 2.15 to 12 $\mu\text{m}$ covering VISTA/ $K_S$ and the first three WISE bands
49	e_alpha_KW3		fitting error of $\alpha_{KW3}$
50	alpha_KW		Observed spectral index $\alpha_{KW}$ from 2.15 to 22 $\mu\text{m}$ covering VISTA/ $K_S$ and all four WISE bands
51	e_alpha_KW		fitting error of $\alpha_{KW}$
52	alpha_W13		Observed spectral index $\alpha_{W13}$ from 3.4 to 12 $\mu\text{m}$ covering the first three WISE bands
53	e_alpha_W13		fitting error of $\alpha_{W13}$
54	alpha_WISE		Observed spectral index $\alpha_{WISE}$ from 3.4 to 22 $\mu\text{m}$ covering the four WISE bands
55	e_alpha_WISE		fitting error of $\alpha_{WISE}$
56	alpha_KW12M		Observed spectral index $\alpha_{KW12M}$ from 2.15 to 24 $\mu\text{m}$ covering VISTA/ $K_S$ , the first two WISE bands, and MIPS1
57	e_alpha_KW12M		fitting error of $\alpha_{KW12M}$
58	alpha_KM_0		De-reddened spectral index $\alpha_{KM}$ from 2.15 to 24 $\mu\text{m}$ covering VISTA/ $K_S$ , the four IRAC bands, and MIPS1
59	e_alpha_KM_0		fitting error of $\alpha_{KM}$
60	alpha_IM_0		De-reddened spectral index $\alpha_{IM}$ from 3.6 to 24 $\mu\text{m}$ covering the four IRAC bands and MIPS1
61	e_alpha_IM_0		fitting error of $\alpha_{IM}$
62	alpha_IRAC_0		De-reddened spectral index $\alpha_{IRAC}$ from 3.6 to 8 $\mu\text{m}$ covering the four IRAC bands
63	e_alpha_IRAC_0		fitting error of $\alpha_{IRAC}$
64	alpha_I2M_0		De-reddened spectral index $\alpha_{I2M}$ from 4.5 to 24 $\mu\text{m}$ covering three IRAC bands and MIPS1
65	e_alpha_I2M_0		fitting error of $\alpha_{I2M}$
66	alpha_I3M_0		De-reddened spectral index $\alpha_{I3M}$ from 5.8 to 24 $\mu\text{m}$ covering two IRAC bands and MIPS1
67	e_alpha_I3M_0		fitting error of $\alpha_{I3M}$
68	alpha_KI3_0		De-reddened spectral index $\alpha_{KI3}$ from 2.15 to 5.8 $\mu\text{m}$ covering VISTA/ $K_S$ and three IRAC bands
69	e_alpha_KI3_0		fitting error of $\alpha_{KI3}$
70	alpha_KW3_0		De-reddened spectral index $\alpha_{KW3}$ from 2.15 to 12 $\mu\text{m}$ covering VISTA/ $K_S$ and the first three WISE bands
71	e_alpha_KW3_0		fitting error of $\alpha_{KW3}$
72	alpha_KW_0		De-reddened spectral index $\alpha_{KW}$ from 2.15 to 22 $\mu\text{m}$ covering VISTA/ $K_S$ and all four WISE bands
73	e_alpha_KW_0		fitting error of $\alpha_{KW}$
74	alpha_W13_0		De-reddened spectral index $\alpha_{W13}$ from 3.4 to 12 $\mu\text{m}$ covering the first three WISE bands
75	e_alpha_W13_0		fitting error of $\alpha_{W13}$
76	alpha_WISE_0		De-reddened spectral index $\alpha_{WISE}$ from 3.4 to 22 $\mu\text{m}$ covering the four WISE bands
77	e_alpha_WISE_0		fitting error of $\alpha_{WISE}$
78	alpha_KW12M_0		De-reddened spectral index $\alpha_{KW12M}$ from 2.15 to 24 $\mu\text{m}$ covering VISTA/ $K_S$ , the first two WISE bands, and MIPS1
79	e_alpha_KW12M_0		fitting error of $\alpha_{KW12M}$
80	AK_Herschel	mag	The extinction extracted from the Herschel map at the position of each source.
81	AK_IR	mag	The line-of-sight foreground extinction towards each source, mainly obtained by the NICER technique using VISTA NIR data. See column “AK_method” for more details.
82	AK_method		The method, which was used to infer the line-of-sight extinction towards each source. The used methods are: NICER (Lombardi & Alves 2001); JH (reddening $E(J - H)$ ); HK (reddening $E(H - K)$ ); JK (reddening $E(J - K)$ ); PNICER (Meingast et al. 2017); HERSCHEL (If the individual line-of-sight extinction is significantly larger than that of the cloud measured by <i>Herschel</i> ( $A_{K,IR} > A_{K,Herschel}$ ). Or when too few bands were available for an extinction calculation, then $A_{K,Herschel}$ was used for an estimate of line-of-sight extinction towards that source, while larger values than $A_K > 9$ mag were not allowed); and LIT(Ref) (Extinction taken from spectral surveys from the Literature with the Reference in brackets: Hillenbrand (1997) (11), Fang et al. (2009, 2013) (12), Furlan et al. (2016) (13)).
83	ClassCog		VISTA extension flag {0,1}. Source morphology derived from variable aperture photometry in combination with machine learning techniques. 0 indicates an extended object, 1 indicates point-like morphology.
84	ClassSex		VISTA extension flag [0,1]. Determined by the source extraction algorithm SExtractor. Values close to 0 indicate an extended object, values close to 1 point-like morphology.

**Table C.1.** — Continued

Column Number	Column Name	Units	Description
85	X		A reference is given if the source was detected as X-ray source: (4) COUP, <a href="#">Getman et al. (2005b)</a> ; (5) SFINCS, <a href="#">Getman et al. (2017)</a> ; (6) XMM-Newton L1641, <a href="#">Pillitteri et al. (2013)</a> ; (7) XMM-Newton $\kappa$ -Ori, <a href="#">Pillitteri et al. (2016)</a> .
86	TTS		Flag if the source was classified as classical (C) or weak-line (W) T-Tauri star, with the reference in brackets: (8) <a href="#">Szege-di-Elek et al. (2013)</a> ; (9) <a href="#">Fang et al. (2009, 2013, 2017)</a> ; (10) <a href="#">Hsu et al. (2012, 2013)</a> .
87	C <sub>MGM</sub>		Classification from MGM including disk sources (D), protostars (P), faint protostar candidates (FP), red protostar candidates (RP).
88	C <sub>FF16</sub>		Classification for HOPS sources from <a href="#">Furlan et al. (2016)</a> including Class 0 (0), Class I (I), flat spectrum (F), Class II (II), galaxies (G), and unclear objects (U).
89	C <sub>LL16</sub>		Classification from <a href="#">Lewis &amp; Lada (2016)</a> for the 44 low- $A_K$ MGM protostars.
90	C <sub>Fang</sub>		Classification from <a href="#">Fang et al. (2009, 2013)</a> for the L1641 region.
91	C <sub>P13</sub>		Classification from <a href="#">Pillitteri et al. (2013)</a> for the L1641 region.
92	Class		Classification as proposed in this work including YSO candidates with IR-excess and rejected sources: Class 0 (0) or Class I protostars (I), flat spectrum sources (F), Class II/III pre-main-sequence stars with disks (D), anemic disks (AD), or transition disks (TD). For rejected candidates the type of contamination (false positive) is given: galaxy (G), nebulosity and fuzzy contamination (fuzz), main-sequence star (star), Class III candidate without IR-excess (III), or photometric contamination like image artifact (C). For uncertain candidates we indicate the suggested uncertain object type as follows: galaxy candidate “UG”, or uncertain YSO candidate: “UP”, “UF”, and “UD” (depending on the spectral index).
93	Class_flag		Flag to distinguish between (1) revisited YSO candidates, (2) new YSO candidates, (3) rejected candidates, and (4) uncertain candidates.