A Census of Star Formation in the Outer Galaxy: The SMOG Field

Elaine Winston1, Joseph Hora1, Robert Gutermuth2, and Volker Tolls1

1 Center for Astrophysics Harvard & Smithsonian, 60 Garden St., Cambridge MA 02138, USA; elaine.winston@cfa.harvard.edu
2 Department of Astronomy, University of Massachusetts, Amherst, MA 01003, USA

Received 2019 April 1; revised 2019 May 29; accepted 2019 June 4; published 2019 July 18

Abstract

In this paper we undertake a study of the 21 deg2 SMOG field, a Spitzer cryogenic mission Legacy program to map a region of the outer Milky Way toward the Perseus and outer spiral arms with the IRAC and MIPS instruments. We identify 4648 YSOs across the field. Using the DBSCAN method, we identify 68 clusters or aggregations of YSOs in the region, having eight or more members. We identify 1197 Class I objects, 2632 Class II objects, and 819 Class III objects, of which 45 are candidate transition disk objects, utilizing the MIPS 24 photometry. The ratio of YSOs identified as members of clusters was 2872/4648, or 62%. The ratios of Class I to Class II YSOs in the clusters are broadly consistent with those found in the inner Galactic and nearby Gould Belt young star formation regions. The clustering properties indicate that the protostars may be more tightly bound to their natal sites than the Class II YSOs, and the Class III YSOs are generally widely distributed. We further perform an analysis of the WISE data of the SMOG field to determine how the lower resolution and sensitivity of WISE affects the identification of YSOs as compared to Spitzer: we identify 931 YSOs using combined WISE and 2MASS photometry, or 20% (931/4648) of the total number identified with Spitzer. Performing the same clustering analysis finds 31 clusters that reliably trace the larger associations identified with the Spitzer data. Twelve of the clusters identified have previously measured distances from the WISE H II survey. SEDFitter modeling of these YSOs is reported, leading to an estimation of the initial mass function in the aggregate of these clusters that approximates that found in the inner Galaxy, implying that the processes behind stellar mass distribution during star formation are not widely affected by the lower density and metallicity of the outer Galaxy.

Key words: circumstellar matter – infrared: stars – stars: formation – stars: pre-main sequence – stars: protostars

Supporting material: machine-readable tables

1. Introduction

The study of star formation has been revolutionized with the launch of the Spitzer Space Telescope (Spitzer; Werner et al. 2004). To date, most of the published studies have focused on star-forming regions (SFRs) in the inner part of our own Galaxy: those regions located within ≈1 kpc of the Sun or toward the Galactic center/bar. Regions such as Orion or the Gould Belt clusters cover sites of star formation from low-mass clusters with hundreds of members to high-mass regions with tens of thousands of YSOs. They are within 1 kpc of the Sun and can be studied in high resolution to substellar masses (Dunham et al. 2015; Megeath et al. 2016). More distant SFRs located toward the Galactic center, such as W51 and W43 (Saral et al. 2017, d ≈ 6 kpc) or RCW 38 (Winston et al. 2011, d ≈ 1.7 kpc), may be the precursors for young massive clusters (YMCs; Bressert et al. 2012; Ginsburg et al. 2012). Spitzer observations of these regions can detect YSOs down to ~0.5−1 M☉ and are more similar to what we can achieve toward the outer Galaxy. Though the initial environmental conditions can vary greatly across these regions, properties such as the initial mass function (IMF) appear to be universal across all formation size scales. In calculating the star formation rates in external galaxies, it is generally assumed that this universality holds. However, when we look at extragalactic star formation, we are studying star formation not just in these inner Galactic regions but in the entire integrated flux of the galaxy, which includes the equivalent of our own outer Galaxy, where we have not yet established the universality of the IMF slope.

We expect that the efficiency with which a molecular cloud forms stars is dependent on its density, temperature, and chemical abundances (e.g., Evans 1999). The outer Galaxy is thought to be a distinctly different environment from that of the inner Galaxy, with conditions less likely to efficiently form stars. The metallicity of the Milky Way has been shown to decline as a function of galactocentric radius (e.g., Rudolph et al. 1997). A recent paper by Huang et al. (2015) used red clump stars to show that [Fe/H] is decreasing in the midplane with radial distance from R$_{GC}$ of 7−11.5 kpc at 0.082 ± 0.003 dex kpc$^{-1}$, with [Fe/H] ≈ 0 in the solar neighborhood. There is less evidence for a decrease in the outer Galaxy (11.5 ≤ R$_{GC}$ ≤ 14 kpc) with [Fe/H] ~ −0.2 to −0.4 and Δ[Fe/H] of −0.017 ± 0.023 dex kpc$^{-1}$. They ascribe this difference to different evolutionary histories in the inner and outer disk of the galaxy.

Average temperatures in molecular clouds are found to be lower (Mead & Kutner 1988), and the cosmic-ray flux is also thought to be lower (Bloemen et al. 1984). Further, the volume density of molecular clouds in the outer Galaxy is lower, and so interaction rates and incidence of spiral arm crossings will be lower over the star-forming lifetime of the cloud. With the advent of all-sky surveys we can now turn our attention to the outer Galaxy to identify new star-forming clusters in these more remote regions to determine whether the outer Galaxy is as amenable to the formation of stars as the inner Milky Way. For our purposes here, the term “outer Galaxy” refers to clusters more than ≈2 kpc from the Sun in directions away from the Galactic center from roughly 90° < l < 270°.

The Spitzer SMOG survey (Carey et al. 2008) was designed to provide deep coverage of a representative field in the outer Galaxy. It was observed during the cold mission, making it the largest field in the Galactic plane of the outer Galaxy with coverage across all IRAC and MIPS bands. We undertook an
initial survey of the outer Galaxy with this region, first, to make use of the longer wavelength coverage, and second, to then compare the YSOs identified with Spitzer and Wide-field Infrared Survey Explorer (WISE). Our future work will expand on this study by applying the techniques outlined in this paper to the outer Galaxy covered by the GLIMPSE 360 Galactic plane survey (warm mission, IRAC bands 1 and 2 only) and WISE.

With this work we identify YSOs across the field from their excess emission in the infrared, locate new regions of star formation, determine the evolutionary class of the YSOs to analyze the protostellar ratio of the clusters, and make a preliminary assessment of the IMF across clusters with known distances in the outer Galaxy. Figure 1 shows an IRAC three-color mosaicked image of the entire SMOG field, with the 8.0 μm (red) highlighting a number of young stellar clusters, and the warm dust across the field, 4.5 μm (green), tracing shocked hydrogen emission, and 3.6 μm in blue.

In this paper we first discuss, in Section 2, the origins of the infrared catalogs. In Section 3 we discuss the contamination removal process, the identification of the YSOs, and their evolutionary classification. We then discuss the spatial distribution of the young stars and the identification of stellar clusters in the outer Galactic fields in Section 4. We discuss the fits to the SEDs of YSOs in clusters with known distances in Section 5. We then compare the WISE catalog with the IRAC observations in Section 7. Finally, a brief summary is presented in Section 8.

2. Observations and Data Reduction

2.1. SMOG Survey

The Spitzer Mapping of the Outer Galaxy (SMOG; Carey et al. 2008, Prog. ID 50398) survey was a Spitzer Cryogenic Cycle 5 program to map a 3° × 7° region of the outer Milky Way with IRAC (Fazio et al. 2004) and MIPS (Rieke et al. 2004). The field covers 21 deg² over 102° < l < 109° and 0° < b < 3°, covering the FWHM of the gas disk at R ≈ 20 kpc, and tracing the warp of the Galactic plane in the second quadrant (Wouterloot et al. 1990). The field was selected to contain sections of two of the galaxy’s spiral arms: the Perseus arm and the outer spiral arm. The Perseus arm lies at an estimated distance of d ≈ 3–5 kpc or galactocentric distance of R ≈ 9 kpc, while the outer spiral arm is located at d ≈ 7–10 kpc or R ≈ 12 kpc (Xu et al. 2006; Churchwell et al. 2009).

The SMOG data products, including mosaics and point-source archive, were downloaded from the Infrared Science Archive (IRSA). The IRAC observations at 3.6, 4.5, 5.8, and 8.0 μm were obtained in High Dynamic Range mode, which obtained 0.6 and 12 s integrations for four dithered images at each mosaic position. These data products were processed using the IRAC SSC S18.5.0 pipeline. The data products consist of the SMOG Catalog, with 2,512,099 high-reliability point sources, and the SMOG Archive, with 2,836,618 sources with less stringent selection criteria. The Archive has lower signal-to-noise ratio thresholds and fluxes, and sources with separations as close as 0.5″ are retained. In this study we utilize the larger Archive (referred to as “the SMOG catalog” going forward) to include as many faint YSOs as possible. The mid-IR photometry was supplemented by J-, H-, and K-band photometry from the Two Micron All Sky Survey (2MASS) point-source catalog (Skrutskie et al. 2006). The photometric catalogs were merged using a 1″ matching radius. Documentation describing the SMOG survey reduction is available on the IRSA website. The data products also include the SMOG Image Atlas, consisting of 3°×3°7 mosaics in each of the four IRAC bands, with a 1″2 pixel size, which were merged using Montage to create the image in Figure 1.

The MIPS mosaics (24, 70, 160 μm) were taken at medium scan rate with full-width scan stepping. Due to the high background and low angular resolution of the 70 and 160 μm bands, point-source extraction was only applied to the 24 μm mosaic. The MIPS 24 μm observations were reduced following Gutermuth et al. (2008, 2009) and yielded a point-source

---

Figure 1. Spitzer IRAC three-color image of the full 21 deg² SMOG field, with blue for 3.6 μm, green for 4.5 μm, and red for 8.0 μm. Dust emission dominates the field, with the larger, brighter regions corresponding to areas of star formation activity.
catalog containing 15,298 objects. We merged the MIPS catalog with the SMOG Archive, requiring a 1″ matching radius, consistent with the positional accuracy of the data sets and to minimize false matches. A match to an IRAC source was found for 14,695 (96.1%) of the MIPS sources.

2.2. UKIRT Infrared Deep Sky Survey (UKIDSS)

While 2MASS covers the full sky, it only does so to a depth of 16.5 mag at J band. In order to improve our ability to detect fainter and more embedded YSOs and to add photometric coverage for the SEDFitter modeling, we have included the deeper UKIDSS near-IR J-, H-, and Ks-band photometry, with a depth of ~19.8 mag at J band and an FWHM of 0.9″. The UKIDSS project is defined in Lawrence et al. (2007). UKIDSS uses the UKIRT Wide Field Camera (WFCAM; Casali et al. 2007). The photometric system is described in Hewett et al. (2006), and the calibration is described in Hodgkin et al. (2009).

The pipeline processing and science archive are described in Hambly et al. (2008). The UKIDSS survey does not cover the entire SMOG field, instead covering approximately half of the survey area from $l = 102°$ to $l = 105°$. The data were downloaded from the WFCAM Science Archive (WSA) as part of DR10. The catalog of sources covering the SMOG field contained 163,785 point sources.

2.3. WISE

WISE (Wright et al. 2010) provides mid-IR photometry at 3.4, 4.6, 12, and 22 μm. The AllWISE catalog is an all-sky survey combining the cryogenic WISE All Sky survey and the initial release of the NEOWISE post-cryogenic survey (Mainzer et al. 2011). The catalog is available via the NASA/IPAC IRSA archive. A selection covering the entire SMOG field was made, resulting in a regional catalog of 581,705 point sources. The WISE satellite has considerably lower spatial resolution when compared to Spitzer, with a highest resolution of ~6″ at 3.5 μm. However, the astrometric accuracy and matching to the 2MASS catalog are to within 1″, and so we performed catalog matching to the SMOG catalog at 1″ also.

3. YSO Identification and Classification

YSOs are frequently identified by their excess emission at IR wavelengths. This emission arises from reprocessed stellar radiation in the dusty material of their natal envelopes or circumstellar disks. The infrared identification of YSOs is carried out by identifying sources that possess colors indicative of IR excess and distinguishing them from reddened and/or cool stars (Allen et al. 2004; Gutermuth et al. 2004; Winston et al. 2007).

A full description of the criteria for contaminant removal and source identification for each data set is given in the appendices. The following subsections outline the YSO selection methods for each data set, the combined YSO catalog, and the evolutionary classification of the YSOs based on their excess IR emission.

3.1. IRAC and 2MASS

3.1.1. Contamination

The SMOG field points toward the outer Galaxy, where the background extinction from the Galactic bar is negligible. It is expected that many of the point sources detected with excess emission will, in fact, be active galactic nuclei (AGNs) or star-forming galaxies (PAH galaxies). Knots of emission in the structure of molecular clouds may also be mistaken for YSOs. A further source of contamination in the YSO sample comes from contamination of the apertures by polycyclic aromatic hydrocarbon (PAH) emission. These sources were identified in the SMOG catalog and removed before selection of YSOs was carried out.

Figure 2 shows the contaminant removal diagrams. In the first three panels the full catalog is displayed in gray, while the various contaminants are highlighted in red. The final panel shows the IRAC color–color diagram (CCD) with the cleaned catalog in gray, overlaid on the full catalog in red. A total of 10,300 sources were removed from the catalog as contaminants. This cleaned catalog was then used to select the YSOs across the SMOG field.

3.1.2. YSO Selection

YSOs were selected using a combination of CCDs with IRAC and 2MASS+IRAC colors, as described in Appendix A. Photometric uncertainties of <0.2 mag and saturation magnitude limits were required in all the bands used for a particular CCD to select YSOs. Sources with colors and magnitudes consistent with YSOs were classified as Class 0/Class I (for deeply embedded protostars, Class II for disk-bearing pre-main-sequence objects, and Class III for objects with weak/anemic disks.

There were 2,826,318 sources in the cleaned SMOG catalog, of which 3,835 were identified as YSOs using the combined 2MASS and IRAC photometry. Figure 3 shows three of the source selection CCDs used in the identification of YSOs, while the fourth shows a color–magnitude diagram of the cleaned catalog and the identified YSOs. Some contaminants that are not accounted for here include foreground and background stars, such as asymptotic giant branch (AGB) stars and highly/unsukely reddened field stars, which may be confused for Class III objects. Such objects are expected to be scattered randomly over the field, and some of the brighter YSO candidates may be foreground AGB stars. A more detailed discussion is given in Section 4.1 on large-scale clustering.

3.2. MIPS

The MIPS 24 μm mosaic yielded 15,298 sources. Of the 14,695 matches to an IRAC counterpart, 14,226 (96.8%) had matches once the contaminant removal had been performed on the SMOG catalog. The remaining 603 unmatched sources, or 3.94% of all the 24 μm detections, are likely to be either background extragalactic sources or perhaps deeply embedded protostellar objects. There were no sources with only 8 μm and 24 μm detections, and only 17 detected only at 5.8, 8.0, and 24 μm. The YSO selection criteria were thus based on two complimentary methods, one including only the three longest bandpasses (see Figure 4), and the second including the shorter IRAC bands and MIPS 24 μm; see Appendix B. These methods identified 2109 and 1819 sources, respectively, for a combined total of 2140 YSOs identified with 24 μm excess.

An attempt was made to identify any embedded protostellar objects from the list of 24 μm only sources. This was done by examining their spatial distribution across the field in relation to other identified YSOs. The objects were not found to be
spatially coincident with the other YSOs, instead being randomly distributed, and so no further attempt is made to classify them here, though the possibility remains that some are deeply embedded protostars.

3.3. IRAC and UKIDSS

The UKIDSS near-IR photometry was merged with the cleaned SMOG catalog using a 1″ matching radius. The rms positional uncertainty for IRAC is <0″/3 after pointing refinement, so matching at 1″ allows for a 3σ positional offset while incorporating positional uncertainties in the UKIDSS photometry. If more than one object falls within the 1″ radius, the closest is taken to be the match. Of the 163,785 UKIDSS sources, 163,485 were matched to a mid-IR detection. No further contaminant removal was applied to the SMOG data. The UKIDSS photometry is saturated for $K < 10.5$, and so it was necessary to apply photometric cuts to the matched sources to avoid selecting saturated sources. We expect all brighter sources to have been previously detected with 2MASS. Figure 5 shows one of the selection criteria used to identify YSOs with the combined IRAC and UKIDSS data set. There were 1182 and 979 sources identified, respectively, with the two criteria; see Appendix C. A combined total of 1352 YSOs were selected. As previously mentioned, the UKIDSS coverage consisted of roughly half of the SMOG field. For this reason our survey is deeper over one-half of the field, and YSO completeness is higher for young objects with fainter near-IR emission in this region.

3.4. Combined YSO Catalog

The three sets of YSO selections, 2MASS+IRAC (3835 sources), IRAC+MIPS (2140 sources), and UKIRT+IRAC (1352 sources), were combined using source identification tagging. All three catalogs required a detection in the SMOG catalog, which uniquely identifies each source. These were combined and overlaps removed. The full list of YSOs contained 4648 objects across the SMOG field. Table 1 lists the column identifiers of the photometry table for the full list of identified YSOs. The full data table for all 4648 YSOs is available online in electronic format. The portion of the SIMBAD database covering the SMOG field was downloaded and the sources compared to the locations of the 4648 identified YSOs. Table 2 provides a sample listing of the YSOs with matches within 1″ of a SIMBAD source, the source identification, the object type, and the spectral type. The full
table is available online in electronic format and includes further selected information pertaining to the SIMBAD objects.

3.4.1. Evolutionary Classification

Young stars evolve through a number of broad stages from the embedded core phase, through the protostellar phase where stellar accretion is still dominant, to the circumstellar disk phase where the envelope has dissipated and processing of disk material is ongoing, to the weak disk regime where the disk has dissipated and any planets will have formed.

A broad classification may be carried out by measuring the slope, $\alpha$, of the spectral energy distribution (SED) across the mid-IR bandpasses (Lada & Wilking 1984). Two values of the slope were calculated by taking the least-squares polynomial fit to the data from 3.6 to 8.0 $\mu$m (IRAC) and from 3.6 to 24.0 $\mu$m (IRAC+MIPS1). Protostellar objects (Class 0 and I) have a rising slope, $\alpha > 0$, Class II sources are characterized by decreasing slopes between $-1.6 < \alpha < -0.3$, and Class III sources lack optically thick emission from a disk and possess decreasing slopes $\alpha < -1.6$, consistent with a weak emission above a stellar photosphere. We note a distinction in nomenclature between Class III sources as defined here and X-ray-detected Class III YSOs (Winston et al. 2010, 2011). Here Class III sources show weak IR emission consistent with a thin disk or weak-line T Tauri star, while X-ray-detected diskless Class III sources do not show any excess IR emission in the IRAC and MIPS bands and would have $\alpha \approx -2$ and be classified as field stars by our method.

The $\alpha_{\text{IRAC}}$ classification was chosen, as it can be applied to the entire combined YSO list simultaneously, regardless of how the YSO was selected. The YSOs were classified as follows: 1197 Class 0/I sources, 2632 Class II sources, and 819 identified as anemic disk Class III stars. Of the 819 Class III sources, 45 were further identified as candidate transition disk objects based on their possessing rising slopes between 8 and 24 $\mu$m.

3.4.2. Incompleteness of the YSO Population

An important topic to address in a survey aimed at identifying a specific population is the subject of incompleteness. In this case, we can address the incompleteness in terms of both total YSO membership and evolutionary class.

The primary difficulty in discussing incompleteness in the total cluster populations, in the context of this study, is that the field contains clusters spanning a wide range of distances, from 1 to 10 kpc, through two of the Galaxy’s spiral arms. Thus, we...
cannot make a meaningful estimate of the minimum mass of YSO detected without referring to a specific cluster; we can only discuss the incompleteness in terms of the photometric cuts applied as selection criteria. As we do not have distances to all of the identified clusters, and the available distances are not confirmed, we make no further assessment of the completeness of the catalog of identified YSOs in this study.

In discussing the incompleteness by evolutionary class we can say that we are more sensitive to Class II sources, which are less embedded and are dominated by photospheric emission and therefore tend to be brighter in the shorter IRAC bands than the protostellar Class I sources. The more deeply embedded Class I sources may be detected by MIPS, but here the lower resolution and source crowding will reduce the selection of Class I sources in dense and/or more distant clusters. To discuss the incompleteness of the Class III objects, the distinction between truly diskless Class III YSOs and those with weak/anemic disks must be taken into consideration. Those with weak disks, as identified here, will have a similar completeness to the Class II objects. Class III objects without disks will be completely missed, as they cannot be distinguished from field stars at IRAC wavelengths. Spectroscopic or X-ray observations would better allow us to identify the full Class III populations of these regions.

4. Cluster Identification

4.1. Large Scale

The SMOG survey covers a 21 deg$^2$ field toward two of the outer spiral arms of the Milky Way. A cursory visual

| Table 1 | SMOG Field YSOs: Photometry Table Description |
|---------|------------------------------------------------|
| Column  | Column ID | Description                      |
| Number  |           |                                |
| 1       | SF ID     | SMOG Field ID                   |
| 2       | GLIMPSE ID| GLIMPSE ID                      |
| 3       | 2MASS ID  | 2MASS ID                        |
| 4       | R.A.      | Right ascension                 |
| 5       | Decl.     | Declination                     |
| 6       | IRAC det. | Selected by IRAC photometry     |
| 7       | MIPS det. | Selected by MIPS photometry     |
| 8       | UKIRT det.| Selected by UKIRT photometry    |
| 9       | $J_{2M}$  | 2MASS $J$ band                  |
| 10      | $\epsilon J_{2M}$ | 2MASS $J$-band uncertainty     |
| 11      | $H_{2M}$  | 2MASS $H$ band                  |
| 12      | $\epsilon H_{2M}$ | 2MASS $H$-band uncertainty    |
| 13      | $K_{2M}$  | 2MASS $K_s$ band                |
| 14      | $\epsilon K_{2M}$ | 2MASS $K_s$-band uncertainty  |
| 15      | $J_{UK}$  | UKIRT $J$ band                  |
| 16      | $\epsilon J_{UK}$ | UKIRT $J$-band uncertainty    |
| 17      | $H_{UK}$  | UKIRT $H$ band                  |
| 18      | $\epsilon H_{UK}$ | UKIRT $H$-band uncertainty   |
| 19      | $K_{UK}$  | UKIRT $K$ band                  |
| 20      | $\epsilon K_{UK}$ | UKIRT $K$-band uncertainty   |
| 21      | 3.6       | IRAC band 1                     |
| 22      | $\epsilon 3.6$ | IRAC band 1 uncertainty       |
| 23      | 4.5       | IRAC band 2                     |
| 24      | $\epsilon 4.5$ | IRAC band 2 uncertainty       |
| 25      | 5.8       | IRAC band 3                     |
| 26      | $\epsilon 5.8$ | IRAC band 3 uncertainty       |
| 27      | 8.0       | IRAC band 4                     |
| 28      | $\epsilon 8.0$ | IRAC band 4 uncertainty       |
| 29      | 24        | MIPS band 1                     |
| 30      | $\epsilon 24$ | MIPS band 1 uncertainty      |
| 31      | $\alpha_{IRAC}$ | SED slope 3.6-8.0 \(\mu m\) |
| 32      | $\alpha_{MIPS}$ | SED slope 3.6-24 \(\mu m\)   |
| 33      | Class IRAC| Evolutionary class: IRAC bands  |
| 34      | Class MIPS | Evolutionary class: IRAC and MIPS bands |
| 35      | Class Trans| Evolutionary class: transition disk candidate |
| 36      | Cl #      | Number of clusters containing YSO |
| 37      | $sCl$ #   | Number of subclusters containing YSO |

(This table is available in its entirety in machine-readable form.)
examination of the spatial distribution of the YSOs shows evidence of clustering/clumping, as can be seen in Figure 6. The spatial distribution of all identified YSOs is shown in Figure 6 (a); the subsequent three panels show the spatial distributions of the Class I (panel (b)), Class II (panel (c)), and Class III (panel (d)) objects with candidate transition disks, respectively. The Class II sources most clearly show the structures of their underlying cluster distributions. The Class I YSOs show a similar structure, though they are less numerous and somewhat more tightly associated than the Class II sources. The Class III sources do not show as strong a trend toward location in the clustered regions as the younger objects, though in one exception there is a knot entirely consisting of Class III sources.

In order to identify overdensities in the spatial distribution of the identiﬁed YSOs, and thus pick out “clusters” or “regions” of star formation, we elected to use a method called “Density-based spatial clustering of applications with noise” (DBSCAN) as described by Xu et al. (1997). This is a density-based clustering algorithm, where points with many nearby neighbors are selected to form a group and those with only more distant neighbors are ﬂagged as outliers. The algorithm used was the Python SKLEARN package’s implementation of DBSCAN. The method requires only two input parameters: $\epsilon$, the scaling size for clustering, and MinPts, the minimum number of points required to deﬁne a dense region. A detailed description of DBSCAN and its utility in ﬁnding clusters of YSOs in SFRs is presented by Joncour et al. (2018) in their paper on Taurus. Here, the value of $\epsilon$ was chosen by plotting the cumulative distribution of nearest neighbor distances for all the YSOs and determining the turnoff point of the distribution. A value of $\epsilon_{\text{fp}} = 0.11$ was selected from this plot. The minimum number

| SF ID | Glimpse ID | SIMBAD ID | OTYPE | SP_TYPE |
|------|-----------|-----------|-------|---------|
| SRC27 | SSTSMOGA G106.1258+01.0513 | IRAS 22314+5904 | Star | ... |
| SRC52 | SSTSMOGA G107.9079-02.2809 | IRAS 22390+6101 | Star | ... |
| SRC118 | SSTSMOGA G106.0932+01.1394 | IRAS 22308+5908 | Star | ... |
| SRC210 | SSTSMOGA G106.1866+02.3247 | V* V675 Cep | Candidate_LP | ... |
| SRC217 | SSTSMOGA G108.5792+02.5484 | IRAS 22429+6134 | Star | ... |
| SRC254 | SSTSMOGA G108.5944+00.4895 | 2MASS J22523871+6000445 | IR | ... |
| SRC561 | SSTSMOGA G106.5735+00.5228 | [D75b] Em$^*$ 22-073 | Em$^*$ | ... |
| SRC603 | SSTSMOGA G103.4831+01.9991 | IRAS 22103+5828 | IR | ... |
| SRC618 | SSTSMOGA G103.4890+00.5787 | EM$^*$ GGR 56 | Em$^*$ | B |
| SRC707 | SSTSMOGA G102.4061+00.8939 | LS III +56 27 | Star | OB- |

(This table is available in its entirety in machine-readable form.)
density of group members was set at MinPts = 8, to minimize the number of false clusters detected.

Using this method, 68 clusters or groupings were identified. Figure 7 plots the spatial distribution of all identified YSOs in black, with the colored circles representing the identified clusters. The smallest identified cluster contained eight members, and the largest cluster contained 429 members, with a median size of 17 members. Eight clusters had 80 or more members: clusters #0, #2, #10, #11, #22, #30, #59, and #66, with 155, 318, 428, 198, 221, 174, 80, and 104 members, respectively. A point to note is that the addition of the UKIDSS near-IR data over one-half of the field leads to more YSOs detected in that area, which in turn leads to an increase in the number of small-sized clusters detected on that half of the field that would not be detected otherwise. Of the 68 clusters identified, 9 had 8 members, 45 had 25 members or less, and 60 had 80 members or less.

To determine the statistical likelihood that the identified groupings are bona fide stellar clusters, we repeated the DBSCAN analysis for 1000 iterations with distributions of randomly generated points covering the same area as the SMOG field. The size of the distribution was tied to the value of MinPts so that it equaled the number of YSOs in the field and in clusters at and below that value. We found that setting the MinPts number density to 8 returned one false cluster 0.1% of the time. For a MinPts number density of 6–7, fake clusters were detected ≈20% of the time, rising to ≈50% when that value drops to 5. Therefore, we assume that all identified clusters of eight members or more are real, and hence all 68 clusters identified in the SMOG field are likely to be bona fide stellar associations. The choice of value for $\epsilon_p$ was tested with a range in values around 0.9 deg < $\epsilon$ < 0.13 deg to assess the reliability of the cluster identification method. The number of clusters identified was [74, 70, 70, 68, 71, 70, 65] for values of $\epsilon$ of [0.09, 0.1, 0.105, 0.11, 0.115, 0.12, 0.13], respectively. In all cases, the central cores of the clusters identified remain the same, while some YSOs are added or removed. As $\epsilon$ decreases, the larger clusters fragment into multiple smaller clusters that are identified as subclusters; see Section 4.2. As $\epsilon$ increases, regions merge and more diffuse regions are identified as clusters, eventually leading to the merger of all clusters in the field as $\epsilon \gg \epsilon_p$. From this analysis, we have selected $\epsilon_p = 0.11$ deg as the turnoff point in the cumulative distribution.

The total number of YSOs identified as belonging to a cluster was 2888 out of 4648, or 62%, or conversely, that 1776 out of 4648, or 38%, of the YSOs were outliers. It is possible that some of these outliers are associated with clusters/associations located nearby to them but fainter YSOs connecting them are not detectable with this data set. They may also be contaminants: foreground or background galactic field dwarfs or AGB stars with either a high extinction or a small amount of dusty material surrounding them that leads to an excess of flux in the mid-IR. Of the 1776 outliers, 512 were Class I, 748 were Class II, and 516 were Class III objects.

Two approaches were taken to estimating the AGB contaminant population: using the Besançon galactic population synthesis models (Robin et al. 2003, 2014), we obtained the modeled population in 1 deg$^2$ toward $l = 105$ and $b = 1$ and found two AGBs, which when scaled to 21 deg$^2$ yields a population of roughly 40 AGB stars along the line of sight of the SMOG field. A second method of estimating the AGB population was described in Robitaille et al. (2008), by modeling the Galactic population of AGB stars based on that found in the LMC (Srinivasan et al. 2009). By applying a color cut of [8.0–24] < 2.5 to the YSOs with a 24 $\mu$m detection, we estimate a population of approximately 400 AGB stars in the SMOG field. We can consider these two values as upper and lower limits to the AGB contamination in the field. Thus, it is possible that a fraction (23%) of the nonclustered candidates identified are not YSOs, but are in fact Galactic AGB stars.

The approximate area of each cluster was quantified by measuring the convex hull of the associated cluster members. The convex hull of a set of points in two dimensions is the minimum area polygon that contains those points such that all internal angles between adjacent edges are less than 180°. Figure 8 shows the 8.0 $\mu$m gray-scale image of the SMOG field, with the convex hulls of the clusters overlaid. For the eight largest identified clusters, an expanded view of the 8.0 $\mu$m three-color and gray scale is shown in Figures 9–11 with the YSOs overplotted.

Table 3 lists the properties of the clusters found in the SMOG field, the coordinates of their central point, the number of YSOs and subclusters, the WISE HII counterpart, and the
circular radius based on the separation of the most distant YSOs. The complete version of the table is available in the electronic version. The online version also includes a full listing of the SIMBAD objects located within the convex hull of each cluster. These sources are assumed to be associated with the cluster, though no attempt has been made to filter the lists or to match them to the YSOs. They are provided as a reference for more in-depth studies.

4.2. Subclustering

For each of the 68 clusters, the spatial distribution of the YSOs was plotted with the convex hull delineating the region; examples of three of the largest clusters (clusters #0, #2, and #10) are shown in Figure 9. As can be seen from this figure, the regions found by the DBSCAN algorithm are not all uniformly spherically symmetric, with obvious substructure and asymmetries visible in the three clusters shown.

To perform a quantitative search for substructure within the 68 identified clustered regions, a minimum spanning tree (MST) graph of the YSOs was calculated for each of the regions. This graph connects the YSOs in such a way as to connect each point without loops while minimizing the total length of the branches. This was done using the Python MST clustering routine written by VanderPlas (2016). Subclusters were identified as those points connected by branches with lengths less than the characteristic branch length (Gutermuth et al. 2009). Figure 12 shows the MST graphs for clusters #0, #2, and #10. The purple circles indicate the objects that were found not to be in subclusters, and the brighter colored circles indicate the identified subclusters in each cluster.

The value of the characteristic branch length was determined separately for each region and is defined as the knee point in the cumulative distribution of branch lengths. Subclusters were identified in 56 of the 68 of the clusters, or ∼82%. There were 12 clusters without subclustering, all of which contained 20 members or less. For the eight largest clusters with more than 80 members, clusters #0, #2, #10, #11, #22, #30, #59, and #66 were found to contain 6, 13, 18, 8, 15, 3, 3, and 6 subclusters, respectively.

The result is that the larger the identified clusters/regions, the higher the number of subclusters. This is intuitive and indicates that the larger regions are likely to be comparable to the nearby Orion complex in terms of extended structure and multiple formation sites surrounding the massive central core. Further high-resolution near-IR observations of the eight largest regions are underway to better determine their high-mass stellar content and to identify the lower-mass and diskless population more completely. This will provide a more accurate assessment of the subclustering and determine whether certain clusters are older/more evolved or whether triggering has occurred.

Having determined that a number of the larger clusters contained subclusters, an attempt was made to determine whether the protostellar fraction varied within each cluster. The protostellar fraction, as defined here, represents the ratio of the Class I to Class II YSOs in an area. We include only the Class I and Class II sources in this ratio, as these two classes will be similarly complete based on our data and selection methodology. In this case, for each subcluster and for the dispersed population with each cluster, the ratio was calculated, and the results were plotted for the three clusters #0, #2, and #10 in Figure 13. The circles represent the YSOs, and the colors represent the protostellar fraction for each subcluster. In a number of cases the dispersed population showed a higher ratio than many of the subclusters, with knots of mainly Class II sources showing very low ratios (as can be seen in cluster #0). In other cases, as exemplified by cluster #10, the subclusters show slightly higher ratios than the dispersed population, with some regions composed of almost 50% Class I objects. There was no trend in protostellar fraction with spatial distribution obviously visible in any of the 68 clusters examined. In the majority of cases, the variation in protostellar fraction was small, a factor of ∼2–3 across the region. In a few cases, a factor of ∼5–10 was present in

Figure 8. Gray scale showing the IRAC 8.0 μm mosaic of the SMOG field. The cyan contours overlaid on the image represent the calculated convex hulls for each cluster, providing a rough spatial outline of each cluster, annotated with the cluster number. The area to the right (lower longitudes) of the dashed white line is covered by the UKIDSS photometry.
regions where small knots of both Class I and II sources were present. Though incomplete without the inclusion of a confirmed population of Class III young stars to provide accurate disk fractions as a measure of cluster age, this study nevertheless provides a useful indication of variation in age and possibly environment across the identified regions.

Figure 9. First three of the eight largest clusters identified in the SMOG field by cluster membership from largest to smallest: #10, #2, #22 from top to bottom, in three-color (r: 8.0; g: 4.6; b: 3.6 μm) and IRAC 8.0 μm gray scale. The contours overlaid on the gray scale represent the calculated convex hulls for each cluster, providing a rough spatial outline of each cluster. The symbols represent the YSOs: Class I (red circle), Class II (green square), Class III (cyan star), transition disk (blue triangle).
5. SED Model Fitting

The SED has previously been used to broadly classify each of the YSOs into an evolutionary stage based on its slope across the IRAC bands. The Python SEDFitter package of Robitaille et al. (2007) uses a sample grid of YSO model SEDs with varying age, mass, inclination, etc., to compare to the input photometry, with the scale factor $S$ (dependent on source distance and luminosity) and the extinction $A_V$ as free parameters. The code returns a sample of best-fit models and their associated parameters.

An initial run was performed for all identified YSOs, separated into clusters and the field population, where the distance was allowed to range from 1 to 8 kpc and the extinction to range from 0 to 40$A_V$. It was found that within each cluster the range in distances for the best-fit models

Figure 10. Similar to Figure 9, but for clusters #11, #30, and #0 from top to bottom.
covered the entire provided distance range and that therefore the results could not be used in any meaningful way. It was therefore decided that only those clusters with reliable estimates for their distances would be included in the SEDFitter modeling. Reliable distances were found for 12 of the 68 clusters in the SMOG field. The WISE H II survey was consulted for known objects in the field (Anderson et al. 2014). These were overplotted on the clusters and radial distance matching as well as spatial coincidence of the convex hulls with the H II region determined. A final visual assessment indicated an association between H II regions with distance estimates and 12 of the clusters. The associated H II regions are

**Table 3**

SMOG Field Young Stellar Objects: Cluster Descriptions

| Cluster ID | Cluster Name | YSO # | Subcluster | Central R.A. (J2000) | Central Decl. (J2000) | Circular Radius (deg) | Branch Length (deg) | WISE H II |
|------------|--------------|-------|------------|----------------------|-----------------------|----------------------|---------------------|-----------|
| 0          | "G106.48+1.0" | 155   | 5          | 339.03101153         | 59.4870041152         | 0.429                | 0.015               | ["G106.241+00.957", "G106.499+00.925"] |
| 1          | "G106.33+0.4" | 40    | 3          | 339.642358587        | 58.558101249          | 0.123                | 0.013               | []        |
| 2          | "G108.57+0.3" | 321   | 13         | 343.177203295        | 59.8416664739         | 0.809                | 0.028               | ["G108.603+00.494"] |
| 3          | "G106.08+1.0" | 11    | 0          | 338.307555636        | 59.2724963636         | 0.089                | 0.031               | []        |
| 4          | "G106.41+3.1" | 27    | 1          | 336.675116071        | 61.235122011          | 0.163                | 0.034               | []        |
| 5          | "G106.36+0.45" | 8     | 0          | 339.308996875        | 58.936327375          | 0.073                | 0.031               | []        |
| 6          | "G106.18+0.14" | 30   | 2          | 339.30691536         | 58.576359             | 0.183                | 0.032               | ["G106.142+00.129"] |
| 7          | "G108.51+2.44" | 8    | 0          | 341.185721           | 61.7164695            | 0.093                | 0.047               | []        |
| 8          | "G107.97+0.49" | 11   | 1          | 341.9835166          | 59.7509637            | 0.130                | 0.077               | []        |
| 9          | "G108.69+1.63" | 17   | 1          | 342.296550864        | 61.0738400643         | 0.172                | 0.039               | []        |

(This table is available in its entirety in machine-readable form.)

**Figure 11.** Similar to Figure 9, but for clusters #66 and #59 from top to bottom.
listed in Table 3. The clusters with identified distances were clusters #0, #2, #10, #11, #25, #30, #32, #39, #41, #60, #64, and #66. The distances ranged from 2.5 to 7.8 kpc and are listed in Table 4. Cooper et al. (2013) provide distances to three of these clusters from their spectral survey of high-mass YSOs: clusters #10, #39, and #41. The distances to clusters #39 and #41 matched exactly, and their distance to cluster #10 was 2.9 kpc, whereas the WISE distance was 5.7 kpc. We elected to use the WISE distance in order to have a consistent distance metric for all the modeled clusters. These distances indicate that the SFRs are distributed over a wide section of the outer Galaxy and are located close to both the Perseus arm and the outer spiral arm, as well as falling at intermediate distances, indicating that they may be in the interarm region.

The second release of the Gaia catalog was also searched for matches between known YSOs. For two clusters, a previously identified YSO within the convex hull was matched to a Gaia source within an estimated distance of 1″. The Gaia parallaxes did not provide distances within the ranges of the spiral arms. Therefore, considering that there was a high likelihood that the matches were erroneous, we did not include them in the subsequent analysis.

SEDfitter was rerun using the distance range fixed to the known distance, while the $A_V$ was still allowed to vary from 0 to 40$A_V$, for each cluster separately. Not all of the YSOs could be fit; those lacking photometry across a sufficient number of bands were not fit. Table 5 lists a sample of the results of the SEDfitter routine for the YSOs in the 12 clusters for which distances were known. A number of parameters are presented for each model fit, including the best fit to the $\chi^2$-squared, object mass, disk mass, age, $A_V$, central temperature, disk accretion rate, etc. The upper and lower limit model fit parameters are also supplied in each case. The full data table is available online in electronic format.

The individual ages and masses derived from the model fits may not be entirely reliable and will not be discussed here.

---

Table 4

| Cluster | #0 | #2 | #10 | #11 | #25 | #30 | #32 | #39 | #41 | #60 | #64 | #66 |
|---------|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Distance (kpc) | 6.0 | 2.5 | 5.7 | 2.7 | 7.8 | 3.1 | 7.4 | 2.6 | 7.3 | 7.2 | 4.4 | 5.5 |

---

Figure 12. Examples of the subclusters of the YSOs as identified using the MST method for three of the larger identified clusters, #0, #2, and #13. The black lines show the MST branches. The purple circles indicate the YSOs that were not assigned to a subcluster. The various other colors code for the identified subclusters with the larger region.

Figure 13. Examples of the variation in protostellar fraction, Class I/(Class I+Class II), across the subclusters and field with each cluster. The clusters are the same as in Figure 12. The color bar shows how the color scaling relates to protostellar fraction in each cluster, with red indicating more protostars and blue indicating more pre-main-sequence YSOs. It should be noted that the scale is different in each cluster in order to best show variation within its subclusters.
However, cumulatively by cluster, they can provide an insight into the relative ages and masses of the YSOs and thus provide an estimate of the age and mass range of the cluster as a whole. Figure 14 shows the age and relative mass of the stars in three clusters by their spatial distribution. The size of the circles indicates the mass of the star relative to the most massive object in that cluster. The color indicates the age: the range in ages is shown by the color bar for each plot. The modeled age of the YSO, with the bluer objects being younger and the redder objects being older.

Figure 14. Examples of the SEDfitter results for spatial distributions of ages and masses within each cluster. The clusters are the same as in Figure 12. The relative sizes of each symbol indicate how massive the YSO is with respect to the most massive object in that cluster. The color bar shows how the color scaling relates to the calculated age of the YSO, with the bluer objects being younger and the redder objects being older.

### Table 5

| SF ID | Cl # | Dist (kpc) | $\chi^2$ | $A_V$ (mag) | $M_*$ ($M_\odot$) | Age (yr) | $M_{	ext{disk}}$ ($M_\odot$) | $M_{	ext{envl}}$ ($M_\odot$) | $M$ ($M_\odot$ yr$^{-1}$) | $T_\text{f}$ (K) | $L_\text{e}$ ($L_\odot$) |
|-------|------|------------|----------|-------------|-----------------|----------|------------------|------------------|-----------------|--------|----------------|
| SRC6  | 0    | 6.0        | 0.073    | 36.580      | 2.596           | 2.766e+06 | 8.765e−08        | 3.217e+00        | 0.000e+00       | 5.752e+03 | 1.958e+01 |
| SRC16 | 0    | 6.0        | 0.613    | 32.110      | 1.993           | 4.062e+05 | 1.754e−02        | 3.725e+00        | 4.882e−06       | 4.583e+03 | 2.606e+01 |
| SRC60 | 0    | 6.0        | 2.006    | 35.130      | 1.443           | 4.462e+05 | 2.603e−02        | 4.149e+00        | 1.073e−07       | 4.447e+00 | 6.725e+00 |
| SRC66 | 0    | 6.0        | 2.687    | 19.690      | 0.745           | 1.965e+05 | 6.353e−03        | 9.685e+00        | 4.222e−06       | 4.007e+00 | 5.877e+00 |
| SRC74 | 0    | 6.0        | 0.139    | 22.330      | 2.993           | 1.273e+06 | 6.148e−04        | 1.599e+00        | 0.000e+00       | 5.125e+03 | 1.383e+01 |
| SRC79 | 0    | 6.0        | 0.656    | 18.900      | 1.523           | 3.219e+05 | 6.096e−05        | 4.805e+00        | 4.153e−07       | 4.425e+00 | 9.841e+00 |
| SRC91 | 0    | 6.0        | 0.005    | 33.580      | 2.086           | 5.322e+05 | 9.047e−04        | 3.475e+00        | 1.629e−06       | 4.646e+03 | 9.658e+00 |
| SRC93 | 0    | 6.0        | 2.776    | 4.278       | 2.030           | 4.709e+04 | 1.068e−03        | 8.707e+00        | 1.673e−05       | 4.312e+03 | 4.010e+01 |
| SRC99 | 0    | 6.0        | 0.287    | 8.046       | 0.414           | 1.604e+05 | 1.781e−03        | 5.948e+00        | 3.171e−05       | 3.625e+03 | 1.906e+00 |
| SRC100| 0    | 6.0        | 2.512    | 39.450      | 2.803           | 1.983e+06 | 2.536e−03        | 4.349e+00        | 0.000e+00       | 5.436e+03 | 1.717e+01 |

(This table is available in its entirety in machine-readable form.)
6. Eight Largest Identified Regions

The DBSCAN method identified eight regions of ongoing star formation with more than 80 identified members from the combined IR catalogs. These regions were clusters #0, #2, #10, #11, #22, #30, #59, and #66, with 155, 318, 428, 198, 221, 174, 80, and 104 members, respectively. Each of these regions is associated with a region of bright emission in the 8.0 μm image, generally associated with emission from dust grains heated by the stellar irradiation of massive stars. Each of the regions shows evidence of subclustering within the convex hull. An expanded view of the 8.0 μm gray scale around each of the eight clusters is shown in Figures 9–11 with the YSOs and convex hulls overplotted, alongside a similarly scaled three-color image of the cluster at 4.5, 8.0, and 24 μm.

For six of the eight clusters (all but clusters #22 and #59) estimates of their distances have been obtained by matching to the WISE H II survey catalog. This allows us to estimate their two-dimensional physical spatial extent. We used two values for the cluster radius: the estimated circular radius and the estimated effective hull radius of each cluster. The calculated spatial scales of the seven clusters for \([R_{\text{circ}}, R_{\text{eff-hull}}]\) were approximately [90, 87], [71, 48], [129, 104], [48, 45], [34, 45], and [61, 70], in parsecs, for clusters #0, #2, #10, #11, #30, and #66, respectively. These size scales are comparable to massive star-forming complexes such as Orion, with a span of ~80 pc across the Orion A and B clouds (Megeath et al. 2016), and RCW 38, with a span of ~20 pc across the immediate cloud complex surrounding IRS 2 (Winston et al. 2011). The most massive YSOs identified by SEDFitter in the six clusters with distance estimates were 9.5, 10.5, 11.1, 7.0, 5.7, and 13.6 M⊙ for clusters #0, #2, #10, #11, #30, and #66, respectively. These masses correspond to spectral types from early B to late O-type stars.

Further, by applying the modified Kroupa power-law distribution for the IMF (Weisz et al. 2015) discussed in Section 5, we can estimate the total population above \(\sim 0.5 \, M_\odot\) and compare them to similar regions in the inner Galaxy. We made the assumption that the bins for stars with \(2-5 \, M_\odot\) were complete and used this value of \(N\) to calculate the value of \(\epsilon_0\) for the two mass regimes:

\[
N = (\epsilon_0/\Gamma) \times [M_1^{-\Gamma} - M_2^{-\Gamma}],
\]

where \(M_1 = 2 \, M_\odot\) and \(M_2 = 5 \, M_\odot\). The calculated values of \(\epsilon_0\) were then used to obtain estimates of cluster membership for 0.5–1 \(M_\odot\) and 1–100 \(M_\odot\) and then summed to give the total population estimate for the cluster above \(\sim 0.5 \, M_\odot\). The estimated memberships were 460, 265, 1320, 205, 195, and 285 members for clusters #0, #2, #10, #11, #30, and #66, respectively. These were compared to values of 1000–1500 for Orion (estimated from Megeath et al. 2016) and ~650 for RCW 38 (Winston et al. 2011). The Orion cluster lies at a distance of 414 pc, and we can thus identify young stellar members to well below the hydrogen-burning limit. The RCW 38 complex lies at a distance of 1.7 kpc and thus may more closely reflect the detection rate for intermediate-mass stars with IRAC. All clusters show estimates similar to or lower than the RCW 38 complex, which has been estimated to have a total stellar population of \(10^4\) YSOs. This would imply that at least six of the eight regions are medium-sized star-forming complexes similar in extent and stellar density to those found in the inner Galaxy.

7. WISE Comparison with IRAC

The WISE catalog of the SMOG field contained 581,705 point sources, or ~20% of the number of detections in the SMOG catalog. A search for YSOs was performed using the four-band WISE data combined with 2MASS near-IR photometry; a detailed discussion is given in Appendix D. After the removal of spurious sources and contaminants, there were...
16,689 point sources remaining in the catalog. Figure 16 shows three of the plots used to select YSOs, with the gray points indicating the cleaned catalog and the red points indicating the selected YSO candidates. Of the 16,689 point sources, a total of 931 were identified as YSOs. This number is roughly one-fifth (931/4648, or 20%) of the YSOs detected using the IRAC methods. The evolutionary classifications of these YSOs were determined to be 271 Class I sources, 641 Class II sources, and 19 candidate transition disks.

The WISE full and cleaned catalogs were matched to the SMOG full and cleaned catalogs with both a 1″ and 1.5″ matching radius. The catalogs of detected YSOs were also compared. The numbers of sources matched between catalogs are reported in Table 6. Of the 581,705 sources in the full catalog, 456,530 (78%) were matched to the full SMOG catalog. The percentage is higher at 89% for the two cleaned catalogs, suggesting that a number of spurious detections have been removed from the WISE catalog. There was an overlap of 68% between the two YSO catalogs, indicating that 32% of the WISE YSOs may be misidentifications or contaminants.

Table 7 lists column information for the data table of WISE-identified YSOs, including the WISE photometry, IRAC identifier, and whether or not the object was selected as a YSO by IRAC. The full data table is available online in electronic format.

We compared the IRAC band 1 and 2 (3.6 and 4.5 μm) photometry to WISE bands 1 and 2 (3.4 and 4.5 μm) for those sources matched at 1″ in the cleaned catalogs. The expected linear relationship between the magnitudes was found with increasing scatter toward fainter magnitudes, which is expected owing to source confusion and differences in background subtraction. The histograms of the differences in magnitude for all matched sources and the YSOs were also examined. In all instances, the peak of distribution is very close to zero, indicating relatively good photometric precision between the two instruments. The median values for [I1−W1] were −0.013, −0.127, and −0.153 mag for all sources, Class I, and Class II, respectively. The median values for [I2−W2] were −0.019, 0.237, and 0.049 mag for all sources, Class I, and Class II, respectively.

The spatial distribution of the WISE-identified YSOs was examined in the same manner as with the large-scale clustering of the IRAC sources, using the DBSCAN algorithm. In this case, the separation length was set to ε = 0.2 based on the turnover in the cumulative distribution of nearest neighbors. A total of 31 clusters were identified, as shown in Figure 17. These clusters are broadly consistent with the locations of the IRAC-identified clusters, though a number are formed from

![Figure 16. Color–color and color–magnitude diagrams showing the WISE YSO selection criteria. The red/pink circles in each plot indicate the selected YSOs. The gray points indicate the cleaned WISE photometric catalog. Left: [3.4−4.6] vs. [4.6−12.0]: three-band WISE selection criteria, useful for objects lacking 22 μm photometry, most similar to IRAC selection criteria. Middle: [3.4−4.6] vs. [12.0−22.0]: sources with full four-band WISE photometry, useful for selecting candidate transition disk objects. Right: [H−K] vs. [3.4−4.6]: combined 2MASS and WISE bands 1 and 2, for sources lacking WISE longer-wavelength detections.](image)
agglomerations of smaller regions or do not extend to the same area as the IRAC-identified clusters owing to the detection of fewer, more distributed YSOs. Figure 18 shows an overlay of the spatial distribution of both the WISE clusters and the IRAC clusters, as outlined by the convex hulls of the sources in each cluster, with the blue circles representing the IRAC-identified YSOs. From this figure, we suggest that the WISE data are, for the greater part, locating the same star formation sites as the IRAC data but with a lower spatial resolution or fewer YSOs, and that further studies with WISE alone would be sufficient to locate areas of massive star formation especially.

An investigation was undertaken to see whether there would be an advantage to combining Spitzer warm mission observations with the longer-wavelength WISE data to extend the color space available to identify disks. A similar comparison was carried out by Sewilo et al. (2019) in CMa-l224, who compared their Spitzer GLIMPSE 360 and AllWISE YSO selection to the WISE-only selection of Fischer et al. (2016) and found that 20/93 YSOs had been misidentified in the WISE-only sample. For those WISE detections with matching IRAC photometry, the two shortest WISE bands were replaced by the two shortest IRAC bands, and the WISE selection criteria were run again to ascertain whether a similar number of YSOs would be retrieved. It should be noted that the number of objects now available to be searched decreased from 16,689 to 15,199 sources. Using the same selection criteria, the number of YSOs detected on the $[\text{I}_1-\text{I}_2]$ versus $[\text{I}_2-\text{W}_3]$ diagram dropped from 271 Class I and 433 Class II objects to 97 Class I and 397 Class II objects. This represents a loss of 210 sources, or 29% of the YSOs selected with WISE only. Given the close matching in photometry between the IRAC and WISE data and the spatial positions of the unmatched YSOs in the cores of clusters and in close proximity to the borders of the SMOG field, the decrease in detections is attributed to the 1490 sources that were not matched between the catalogs, likely due to sources being at the edge of the IRAC field of view and to crowding in the more densely populated SFRs.

8. Summary

We have undertaken a study of the 21 deg$^2$ SMOG field that was observed with the IRAC and MIPS instruments. We combined the Spitzer data with 2MASS and UKIDSS near-IR photometry. We also compared our results to the WISE catalog.
of the field to assess the usefulness of merging these two data sets when selecting YSOs.

1. We identify 4648 YSOs with IR-excess emission; of these, 1197 were Class I, 2632 Class II, 819 Class III, and 45 were candidate transition disk objects.

2. Using the DBSCAN algorithm, we identify 68 clusters with \( \epsilon_0 = 0.11 \) and \( N_{\text{min}} = 8 \) YSOs across the SMOG field, all of which had eight or more members. The ratio of YSOs identified as members of clusters was 2873/4648, or 62%. The eight largest clusters, clusters \#0, \#2, \#10, \#11, \#22, \#30, \#59, and \#66, had 155, 318, 428, 198, 221, 174, 80, and 104 members, respectively.

3. Subclusters within the 68 regions were found using the MST of the YSOs with a characteristic branch length estimated for each region. Of the 68 clusters, 56 (or \( \sim 82\% \)) were found to have subclusters. The remaining 12 regions had 20 members or less. The eight largest clusters had 6, 13, 18, 8, 15, 3, 3, and 6 subclusters apiece.

4. For the 12 clusters with known distances, the SEDs of the YSOs were fitted using the SEDFitter routine. From the modeled masses, the IMF was constructed for the known clusters in the field. The slope of the combined IMF was found to be \( \Gamma = 2.38 \pm 0.20 \) above 2 \( M_\odot \) and \( \Gamma = 1.77 \pm 0.25 \) above 4 \( M_\odot \). These values are consistent with those obtained in inner galaxy high-mass SFRs and are likely also consistent with a universal Salpeter IMF.

5. With the WISE catalog we identified 931 YSOs across the SMOG field. The differences in the WISE photometry as compared to IRAC were found to be negligible when compared to the infrared excess emission used to select the YSOs. Even so, compared to the number of YSOs detected with IRAC, 931/4648, or 20%, we find that the WISE resolution hampers the detection of individual cluster members. However, the 32 clusters detected compared favorably in location to the larger IRAC-detected clusters, though some individual clusters identified with IRAC are merged in WISE.

We would like to thank the anonymous referee for their helpful comments on the manuscript. This work is based on observations made with the Spitzer Space Telescope (PID 30574), which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under NASA contract 1407. E.W., J.H., and V.T. gratefully acknowledge funding support for this work from NASA ADAP grant NNX16AF37G. This publication makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation. Support for the IRAC instrument was provided by NASA through contract 960541 issued by JPL. This publication makes use of data products from 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. This research made use of Montage. It is funded by the National Science Foundation under grant No. ACI-1440620 and was previously funded by the National Aeronautics and Space Administration’s Earth Science Technology Office, Computation Technologies Project, under Cooperative Agreement No. NCC5-626 between NASA and the California Institute of Technology.

Appendix A

IRAC and 2MASS Source Selection

The removal of contaminating sources and the selection of YSOs in the SMOG field were undertaken following the methods described in Gutermuth et al. (2008) and Gutermuth et al. (2009) with a number of adjustments to the applied criteria to best suit the character of our data set, in particular the lower background extinction and the greater distances of the SFRs. These criteria are described in the following appendix.

A.1. Contamination

Candidate AGN contaminants were selected from their location on the [4.5] versus [4.5 − 8.0] color–magnitute diagram, with 9287 sources selected:

\[
AGN_1 = \begin{cases} 
4.5 - 8.0 \uparrow 1.2 & \text{and } 4.5 > 12.5 \\
4.5 - 8.0 > 1.2 & \text{and } 4.5 > 12.5 
\end{cases} 
\]

\[
AGN_2 = \begin{cases} 
[4.5][AGN_1] > 15 - 0.5(4.5 - 8.0)[AGN_1] - 1 \\
[4.5][AGN_1] > 14 + 0.5(4.5 - 8.0)[AGN_1] - 2 \text{or } [4.5][AGN_1] > 15. 
\end{cases} 
\]

AGNs share similar colors to YSOs, but they are usually fainter and are removed primarily based on magnitude considerations. It is possible that fainter YSOs also fall into this color space; however, a spatial distribution plot shows that the selected sources are randomly distributed across the field and a less severe cut would most likely include a higher level of contamination.

The sources considered to be PAH galaxy contaminants were selected based on their locations on two CCDs and a photometric cutoff at 4.5 \( \mu \)m. There were 761 and 802 sources selected as PAH contaminants with the first and second criteria, respectively:

\[
\begin{align*}
\text{PAH}_1 &= \begin{cases} 
4.5 - 5.8 < (2.5/2)(5.8 - 8.0) - 1 & \text{and } 4.5 - 5.8 < 1.55 \text{ and } 5.8 - 8.0 > 1 \text{ and } 4.5 > 11.5 \\
4.5 - 5.8 < 1.55 & \text{ and } 5.8 - 8.0 > 1 \text{ and } 4.5 > 11.5. 
\end{cases} \\
\text{PAH}_2 &= \begin{cases} 
3.6 - 5.8 < (3.2/3)(4.5 - 8.0) - 1 & \text{and } 3.6 - 5.8 < 2.25 \text{ and } 4.5 - 8.0 > 1 \text{ and } 4.5 > 11.5. 
\end{cases} 
\end{align*} 
\]

Possible knots of shocked emission were identified using IRAC bands 1–3, with 20 sources selected:

\[
\text{KNOT} = \begin{cases} 
3.6 - 4.5 > 1.05 & \text{and } 3.6 - 4.5 > (1.2/0.55)(4.5 - 5.8) - 0.3 + 0.8 \text{ and } 4.5 - 5.8 < 0.85. 
\end{cases} 
\]

Contamination of the aperture by PAH emission was identified using the following criteria, which found 523 sources:

\[
\begin{align*}
\sigma_{12} &= \sqrt{(unc_{4.5}^2 + unc_{3.6}^2)} \\
\sigma_{23} &= \sqrt{(unc_{4.5}^2 + unc_{5.8}^2)}
\end{align*}
\]
The Astrophysical Journal, 880:9 (20pp), 2019 July 20

PA = [3.6 – 4.5] – σ12 ≤ 1.5 * (4.5 – 5.8) – σ23 – 1) and [3.6 – 4.5] – σ12 ≤ 0.4.

(9)

The SMOG field lies in the direction of the outer Galaxy, and thus the extragalactic background is expected to be higher since the level of shielding from the Galactic center from both the stellar population and dust component is reduced. The true level of contaminants remaining in the cleaned catalog is difficult to estimate precisely. Spectral typing of all YSOs would be necessary to confirm their nature.

A.2. YSO Selection

The YSOs were selected from the traditional four-color IRAC diagram using the following criteria:

\[ \sigma_{12} = \sqrt{(unc_{3.6}^2 + unc_{4.5}^2)} \]  
\[ \sigma_{34} = \sqrt{(unc_{4.5}^2 + unc_{5.8}^2)} \]  
\[ [5.8 – 8.0] \geq 0.3 + \sigma_{34} \] and \[ [5.8 – 8.0] \leq 2 – \sigma_{34} \] and \[ [3.6 – 4.5] \geq 0.2 + \sigma_{12} \] or \[ [5.8 – 8.0] \leq 2.5 – \sigma_{34} \] and [3.6 – 4.5] ≥ 0.5 + σ12. \]  

(12)

A total of 3299 YSOs were selected by this method.

A further selection based on the four IRAC bands was also applied:

\[ \sigma_{13} = \sqrt{(unc_{3.6}^2 + unc_{4.5}^2)} \]  
\[ \sigma_{24} = \sqrt{(unc_{4.5}^2 + unc_{5.8}^2)} \]  
\[ [3.6 – 5.8] \geq 0.5 + \sigma_{13} \] and \[ [4.5 – 8.0] \geq 0.5 + \sigma_{24} \] and \[ [3.6 – 4.5] \geq 0.15 + \sigma_{12} \] and \[ [3.6 – 5.8] + \sigma_{13} \leq (0.14/0.04) \times [(4.5 – 8.0) – \sigma_{24} – 0.5] + 0.5. \]  

(15)

There were 3239 sources selected by this method.

For those sources lacking a detection at 8.0 μm, we applied a selection criteria using the shortest three IRAC bands:

\[ [3.6 – 4.5] – \sigma_{12} > 0.3 \] and \[ [4.5 – 5.8] – \sigma_{23} > 0.3. \]  

(16)

This is a less stringent cut than applied in Gutermuth et al. (2009), where only protostellar objects were selected via this method. A total of 2721 were selected using this criteria.

For sources lacking photometry in the mid-IR, a combination of 2MASS and IRAC bands was used to select for less extincted objects. Two cuts were made, one where a detection in J band was required and one where it was not. These methods identified 347 and 353 sources, respectively:

\[ \sigma_{31} = \sqrt{(unc_{K_s}^2 + unc_{3.6}^2)} \]  
\[ [3.6 – 4.5] – \sigma_{12} \geq 0.101 \] and \[ [K_s – 3.6] – \sigma_{31} > 0.5 \] and \[ [K_s – 3.6] – \sigma_{31} \geq -2.85714 \times [(3.6 – 4.5) – \sigma_{12} – 0.101] + 1.5. \]  

(18)

A final selection criterion simply selected for those sources with \( K_s \) and 8.0 μm detections that displayed a large color excess, a method that yielded 449 sources:

\[ \sigma_{4} = \sqrt{(unc_{K_s}^2 + unc_{8.0}^2)} \]  
\[ [K_s – 8.0] – \sigma_{4} > 3. \]  

(20)

Appendix B

MIPS Source Selection

The YSO selection criteria for objects with MIPS 24 μm photometry were adapted from our previous studies (Winston et al. 2007; Gutermuth et al. 2008):

\[ \sigma_{m} = \sqrt{(unc_{8.0}^2 + unc_{24}^2)} \]  
\[ [8.0 – 24] \geq 1.0 + \sigma_{m} \] and \[ [5.8 – 8.0] \geq -0.1 – \sigma_{34} \] or \[ [8.0 – 24] \geq 0.6 + \sigma_{m} \] and \[ [5.8 – 8.0] \geq -0.2 – \sigma_{34} \] \[ [8.0 – 24] \geq 1.0 + \sigma_{m} \] and \[ [3.6 – 4.5] \geq -0.1 – \sigma_{12} \] or \[ [8.0 – 24] \geq 0.6 + \sigma_{m} \] and \[ [3.6 – 4.5] \geq 0.2 – \sigma_{12}. \]  

(23)

These criteria search for embedded objects with only longer-wavelength detections and those with detections across the IRAC bandpasses.

Appendix C

IRAC and UKIDSS Source Selection

The UKIDSS near-IR photometry and IRAC source selection criteria were

\[ [3.6 – 4.5] – \sigma_{12} > 0.101 \] and \[ [K – 3.6] – \sigma_{11} > 0.5 \] and \[ [K – 3.6] – \sigma_{11} \geq -2.85714 \times [(3.6 – 4.5) – \sigma_{12} – 0.101] + 0.5. \]  

(24)

\[ [K – 8.0] – \sigma_{4} > 2 \] and \[ [K] > 12 \] and \[ [K] > 12 + ((K – 8.0) – 5). \]  

(25)

These mirror the criteria used with the 2MASS selection but account for the saturation of the brighter sources.

Appendix D

WISE Source Selection

Following the process laid out by Fischer et al. (2016), spurious detections were cleaned from the catalog. The first step was to remove those sources with flags in bands W1, W2, and W3. Upper limits in bands W1, W2, and W3 are then removed, and a saturation cutoff of W1 > 5 is applied to the data. The remaining catalog contained 184,912 sources, or ~31.8% of the
original catalog. The contaminating background galaxies and source selection criteria for the WISE data were taken from Koenig & Leisawitz (2014) and Fischer et al. (2016) and adapted to the requirements of the SMOG field. We adjusted the criteria for removal of AGN and star-forming galaxy contaminants slightly from those of Koenig & Leisawitz (2014) as follows:

\[
\text{SFG} = [W2 - W3] > 2.3 \text{ and } [W1 - W2] < 1.0 \text{ and } [W1 - W2] < 0.46(W2 - W3) - 0.78 \text{ and } [W1] > 14
\]

\[
\text{AGN} = [W1] > 1.8(W1 - W3) + 4.1 \text{ and } [W1] > 14 \text{ or } [W1] > [W1 - W3] + 10.0.
\]

Of those 184,912 sources, there were 58,845 star-forming galaxies and 167,787 AGNs identified as contaminants, leaving a cleaned catalog of 16,689 sources.

The clean catalog was then searched for YSOs according to criteria taken from the Fischer et al. (2016) and Koenig & Leisawitz (2014). From these the combined 2MASS and WISE diagram returned 664 YSOs, there were 90 transition disk candidates and 165 YSOs identified from the four-band WISE diagram, and 271 Class I and 433 Class II sources were identified from the WISE 3-band diagram.

**ORCID iDs**

Elaine Winston @ https://orcid.org/0000-0001-9065-6633
Joseph Hora @ https://orcid.org/0000-0002-5599-4650
Robert Gutermuth @ https://orcid.org/0000-0002-6447-899X

**References**

Allen, L., Calvet, N., D’Alessio, P., et al. 2004, *ApJS*, 154, 363
Anderson, L. D., Bania, T. M., Balser, D. S., et al. 2014, *ApJS*, 212, 1A
Bastian, N., Covey, K. R., & Meyer, M. R. 2010, *ARA&A*, 48, 339
Bloomen, J. B. G. M., Bennett, K., Bignami, G. F., et al. 1984, *A&A*, 135, 12
Bressert, E., Ginsburg, A., Bully, J., et al. 2012, *ApJ*, 758, L28
Carey, S., Ali, B., Berriman, B., et al. 2008, spitz prop, 50398
Casali, M., Adamson, A., Alves de Oliveira, C., et al. 2007, *A&A*, 467, 777
Churchwell, E., Babler, B. L., Meade, M. R., et al. 2009, *PASP*, 121, 213
Cooper, H. D. B., Lumsden, S. L., oudmaijer, R. D., et al. 2013, *MNRAS*, 430, 1125
Dawson, S. A., & Schroder, K.-P. 2010, *MNRAS*, 404, 917
Dotter, A. 2016, *ApJS*, 222, eD
Dunham, M. M., Allen, L. E., Evans, N. J., II, et al. 2015, *ApJS*, 220, 11
Evans, N. J., II 1999, *ARA&A*, 37, 311
Fazio, G. G., Hora, J. L., Allen, L. E., et al. 2004, *ApJS*, 154, 10
Fischer, W. J., Padgett, D. L., Stapelfeldt, K. L., et al. 2016, *ApJ*, 827, 96
Ginsburg, A., Bressert, E., Bully, J., et al. 2012, *ApJ*, 758, L29
Gutermuth, R., Megeath, S. T., Muzerolle, J., et al. 2004, *ApJS*, 154, 374
Gutermuth, R., Megeath, S. T., Myers, P. C., et al. 2009, *ApJS*, 184, 18G
Gutermuth, R., Myers, P. C., Megeath, S. T., et al. 2008, *ApJ*, 674, 336
Hambly, N. C., Collins, R. S., Cross, N. J. G., et al. 2008, *MNRAS*, 384, 637
Hewett, P. C., Warren, S. J., Leggett, S. K., et al. 2006, *MNRAS*, 367, 454
Hodgkin, S. T., Irwin, M. J., Hewett, P. C., et al. 2009, *MNRAS*, 394, 675
Huang, Y., Liu, X.-W., Zhang, H.-W., et al. 2015, *RAA*, 15, 1240
Joncour, I., Duchêne, G., & Moraux, E. 2018, *A&A*, 620, A27
Koenig, X. P., & Leisawitz, D. T. 2014, *ApJ*, 791, 131
Kroupa, P. 2001, *MNRAS*, 322, 231
Lada, C. J., & Wilking, B. A. 1984, *ApJ*, 287, 610
Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, *MNRAS*, 379, 1599
Mainzer, A., Bauer, J., Grav, T., et al. 2011, *ApJ*, 731, 53
Mead, K. N., & Kutner, M. L. 1988, *ApJ*, 330, 399
Megeath, S. T., Gutermuth, R., Muzerolle, J., et al. 2016, *AJ*, 1151, 5
Miller, G. E., & Scalo, J. M. 1979, *ApJS*, 41, 513
Parravano, A., Hollenbach, D., & McKee, C. F. 2018, *MNRAS*, 480, 2449
Rieke, G. H., Young, E. T., Engelbracht, C. W., et al. 2004, *ApJS*, 154, 25
Robitaille, T. P., Reyle, C., Derriere, S., et al. 2003, *A&A*, 409, 523
Robin, A. C., Reyle, C., Fili, J., et al. 2014, *A&A*, 569, A13
Robitaille, T. P., Meade, M. R., Babler, B. L., et al. 2008, *ApJ*, 136, 2413
Robitaille, T. P., Whitney, B. A., Indebetouw, R., et al. 2006, *ApJ*, 167, 256
Robitaille, T. P., Whitney, B. A., Indebetouw, R., et al. 2007, *ApJS*, 169, 328
Rudolph, A. L., Simpson, J. P., Haas, M. R., et al. 1997, *ApJ*, 489, 94
Salpeter, E. E. 1955, *ApJ*, 121, 161
Saral, G., Hora, J. L., Audard, M., et al. 2017, *ApJ*, 839, 108
Sewilo, M., Whitney, B. A., Yung, B. H. K., et al. 2019, *ApJ*, 240, 26
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, 131, 1163
Srinivasan, S., Meixner, M., Letherer, C., et al. 2009, *AJ*, 137, 4810
VanderPlas, J. 2016, *JOSS*, 1, 1
Weisz, D. R., Clifton Johnson, L., Foreman-Mackey, D., et al. 2015, *ApJ*, 806, 198
Werner, M. W., Roellig, T. L., Low, F. J., et al. 2004, *ApJS*, 153, 1
Winston, E., Megeath, S. T., Wolk, S. J., et al. 2007, *ApJ*, 669, 493
Winston, E., Megeath, S. T., Wolk, S. J., et al. 2010, *AJ*, 140, 266
Winston, E., Wolk, S. J., Megeath, S. T., et al. 2011, *ApJ*, 741, 166
Wouterloot, J. G. A., Brand, J., Burton, W. B., et al. 1990, *A&A*, 230, 21
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, *AJ*, 140, 868
Xu, X., Ester, M., Kriegel, H.-P., & Sander, J. 1997, *VA*, 41, 397
Xu, Y., Reid, M. J., Zheng, X. W., et al. 2006, *Sci*, 311, 54