Astronomy & Astrophysics

A census of young stellar objects in two line-of-sight star-forming regions toward IRAS 22147+5948 in the outer Galaxy

Agata Karska[1] , Maciej Koprowski[1] , Aleksandra Solarz[2], Ryszard Szczerba[3] , Marta Sewiło[4,5,6],
Natasza Siódmiak[3] , Davide Elia[7] , Marcin Gawroński[1] , Konrad Grzesiak[1], Bosco H. K. Yung[3] ,
William J. Fischer[8], and Lars E. Kristensen[9]

[1] Institute of Astronomy, Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University, ul. Grudziądzka 5, 87-100 Toruń, Poland
e-mail: agata.karska@umk.pl
[2] European Southern Observatory, Alonso de Cordova 3107, Casilla 19001, Santiago, Chile
[3] Nicolaus Copernicus Astronomical Center, ul. Rabiańska 8, 87-100 Toruń, Poland
[4] Exoplanets and Stellar Astrophysics Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
[5] Department of Astronomy, University of Maryland, College Park, MD 20742, USA
[6] Center for Research and Exploration in Space Science and Technology, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
[7] Istituto di Astrofisica e Planetologia Spaziali – INAF, Via Fosso del Cavaliere 100, I 00133 Roma, Italy
[8] Space Telescope Science Institute, 3700 San Martin Dr., Baltimore, MD 21218, USA
[9] Centre for Star and Planet Formation, Niels Bohr Institute, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen K, Denmark

ABSTRACT

Context. Star formation in the outer Galaxy, namely, outside of the Solar circle, has not been extensively studied in part due to the low CO brightness of the molecular clouds linked with the negative metallicity gradient. Recent infrared surveys provide an overview of dust emission in large sections of the Galaxy, but they suffer from cloud confusion and poor spatial resolution at far-infrared wavelengths.
Aims. We aim to develop a methodology to identify and classify young stellar objects (YSOs) in star-forming regions in the outer Galaxy and use it to resolve a long-standing disparity in terms of the distance and evolutionary status of IRAS 22147+5948.
Methods. We used a support vector machine learning algorithm to complement standard color–color and color–magnitude diagrams in our search for YSOs in the IRAS 22147 region, based on publicly available data from the *Spitzer* Mapping of the Outer Galaxy survey. The agglomerative hierarchical clustering algorithm was used to identify clusters. Then the physical properties of individual YSOs were calculated. The distances were determined using CO 1–0 from the Five College Radio Astronomy Observatory survey.
Results. We identified 13 Class I and 13 Class II YSO candidates using the color–color diagrams, along with an additional 2 and 21 sources, respectively, using the applied machine learning techniques. The spectral energy distributions of 23 sources were modeled with a star and a passive disk, corresponding to Class II objects. The models of three sources include envelopes that are typical for Class I objects. The objects were grouped into two clusters located at a distance of ∼2.2 kpc and 5 clusters at ∼5.6 kpc. The spatial extent of CO, radio continuum, and dust emission confirms the origin of YSOs in two distinct star-forming regions along a similar line of sight.
Conclusions. The outer Galaxy may serve as a unique laboratory for exploring star formation across environments, on the condition that complementary methods and ancillary data are used to properly account for cloud confusion and distance uncertainties.

Key words. stars: formation – ISM: molecules – ISM: individual objects: IRAS 22147+5948

1. Introduction

Stars form inside cold and dense cores in molecular clouds. In the Solar neighborhood, about 3000 young stellar objects (YSOs) have been identified in molecular clouds as part of the *Spitzer* Space Telescope Gould Belt survey, providing lifetimes of YSO evolutionary stages and star formation efficiencies in various clouds (Dunham et al. 2015; Kristensen & Dunham 2018). Similar efforts in the outer Galaxy suffer from low angular resolution, sensitivity, and distance uncertainties, and they are often limited to high-mass objects (Urquhart et al. 2008; König et al. 2021). However, due to differences in environmental conditions and a high fraction of atomic-to-molecular gas, the star

formation rate at the outer parts of the Galactic disk is expected to be lower (Kennicutt & Evans 2012).

Galactic surveys using CO and its isotopologues have been used to pinpoint the location and determine physical properties of molecular clouds up to about 20 kpc from the Galactic Centre. A negative metallicity gradient in the Galaxy affects the abundances of dust and molecules (Sodroski et al. 1997), as well as the overall gas and dust cooling budget (Roman-Duval et al. 2010). It causes an increase in the CO-to-H_2 conversion factor (Digel et al. 1990; Pohl et al. 2008; Heyer et al. 2009) and a decrease in the gas-to-dust ratio with the Galactocentric radius (Giannetti et al. 2017). All of these factors likely contribute to the decrease of the mass surface density of molecular

clouds in the outer Galaxy (for a review, see Heyer & Dame 2015).

Clearly, a proper characterization of physical conditions and chemistry of star-forming regions in multiple locations is necessary for improving our understanding of the impact of environment on star formation and its efficiency in the outer Galaxy. The initial step, however, is to obtain a census of YSOs in multiple molecular clouds and build a sample for follow-up spectroscopy surveys. Subsequently, it will be possible to obtain physical and chemical properties of YSOs and link them with the large-scale properties of molecular clouds in the Milky Way.

Recent infrared sky surveys with the *Spitzer* Space Telescope (*Spitzer*, Werner et al. 2004) and the *Herschel* Space Observatory (*Herschel*, Pilbratt et al. 2010) provide a complementary picture of ongoing star formation in uncharted regions of the Galaxy. The *Spitzer* Mapping of the Outer Galaxy (SMOG; PI: S. Carey) survey is specifically aimed to characterize YSOs in sparsely studied star-forming regions around the galactic longitude $l \sim 105°$ and encompassing Perseus and Outer spiral arms at Galactocentric radii of about 9 and 12 kpc, respectively (Carey et al. 2008). A recent study by Winston et al. (2019) identified ~4650 YSO candidates located in 68 clusters and stressed the similarities in star formation properties with the inner Galaxy.

Similar exploratory studies around $l \sim 220°$ were performed by combining near-IR photometry with longer wavelength data, which are also sensitive to more deeply-embedded YSOs. Sewiło et al. (2019) characterized a very active star formation site CMa−*l*224 consisting of ~290 Class I/II YSOs, using data from GLIMPSE360: Completing the *Spitzer* Galactic Plane Survey (PI: B. Whitney) and the *Herschel* infrared Galactic Plane Survey (Hi-GAL; Molinari et al. 2010, see also Elia et al. 2013). The region resembles low-mass star-forming regions in the Gould Belt and possesses the largest concentrations of YSOs and their clusters in a 100 deg^2 field centered on the Canis Major star-forming region toward $l \sim 220°$ (Fischer et al. 2016).

In this paper, we aim to investigate one specific star-forming region located in the $l \sim 105°$ area, which has been subject to contradictory classifications and distance estimates. Here, we assess to what extent various methods of YSO identification provide consistent results. IRAS 22147+5948 was first detected in CO 1–0 with IRAM 30 m as part of a follow-up observing campaign of the star-forming region candidates among IRAS sources in the outer Galaxy, and assigned a distance of 6.48 kpc ($v_{LSR} = -59.05$ km s^{-1}, Wouterloot & Brand 1989). In a parallel survey with the Millimeter-Wave Observatory, the distance of 2.8 kpc was determined based on CO 2–1 observations ($v_{LSR} = -59.5$ km s^{-1}, Wilking et al. 1989). The Dominion Radio Astrophysical Observatory survey at $l \sim 105°$ classified IRAS 22147+5948 as a likely supernova remnant due to a slightly non-thermal spectral index (SNR G104.7+2.8, Joncas & Higgs 1990; Green & Joncas 1994). Subsequent observations from the Canadian Galactic Plane Survey showed that radio spectra are rather flat and this object is most likely an H II region (Taylor et al. 2003; Kothes et al. 2006), without, however, any maser detections (Molinari et al. 1996; Fontani et al. 2010). Infrared observations with 2MASS and WISE provided further evidence that IRAS 22147+5948 is a star-formation site, with stellar surface density enhancement typical for embedded cluster and a blob or shell morphology (Kumar et al. 2006; Lundquist et al. 2014).

Here, we address the most prevalent questions, such as the measurement of the distance to IRAS 22147+5948 and the census of YSOs obtained with standard methods using infrared colors. We also consider whether machine learning techniques

utilizing the AllWISE catalog are equipped to provide new YSO identifications. We also ask whether they are grouped in clusters and whether the distribution of gas and dust emission is consistent with the picture of star formation in IRAS 22147+5948. Finally, we examine whether the spectral energy distributions of YSO candidates are consistent with the models from Robitaille (2017).

The paper is organized as follows. Section 2 describes the multi-wavelength data used in this work, including our maser survey. In Sect. 3, we perform an analysis involving the determination of the kinematic distances to the SMOG sources (Sect. 3.1), an initial identification of YSOs using color–color cuts (Sect. 3.2) and machine-learning algorithms (Sect. 3.3), a comparison to the results from Winston et al. 2019 (Sect. 3.4), an investigation of the clustering of YSOs (Sect. 3.5) and associations with the interstellar medium (Sect. 3.6), finding the best-fit spectral energy distributions with the corresponding physical parameters (Sect. 3.7). Section 4 provides the discussion of the results in the context of previous studies and Sect. 5 presents our summary and conclusions.

2. Data

2.1. Spitzer SMOG survey

The primary data set used in this study consists of the images and point source catalogs from the SMOG survey (Carey et al. 2008). The SMOG survey covers an area of ~21 square degrees in the outer Galactic plane with the Galactic longitude (*l*) and Galactic latitude (*b*) ranges of roughly (102°, 109°) and (0°, 3°), respectively, with the Infrared Array Camera (IRAC, 3.6–8.0 µm; Fazio et al. 2004) and Multiband Imaging Photometer for *Spitzer* (MIPS, 24–160 µm; Rieke et al. 2004) instruments. In this study, we focus on a small region around IRAS 22147+5948 (104°.55 < *l* < 105°) and (2°.55 < *b* < 2°.95), which we refer to as the "IRAS 22147 region".

The SMOG IRAC data were processed by the Wisconsin GLIMPSE IRAC pipeline and the data products are available at the Infrared Science Archive (IRSA; Meade et al. 2016). The data products include the IRAC point source catalog containing the highest reliability sources and the more complete IRAC point source archive containing sources with less stringent selection criteria than the catalog. Here, we use the IRAC archive, which is more complete, although it requires the identification and removal of unreliable detections during our analysis. The angular resolution of the *Spitzer* IRAC 3.6–8.0 µm observations is ~2″.

Spitzer data is supplemented with near-IR (JHK_s) photometry from the Two Micron All Sky Survey (2MASS, Skrutskie et al. 2006), which is included in the final IRAC catalogs at IRSA.

Figure 1 shows the three-color composite image of the IRAS 22147 region using IRAC and MIPS mosaics. The morphology of the region reveals an embedded cluster corresponding to the position of IRAS 22147+5948 and an arc of emission to the north. An extended structure toward south-east corresponds to the position of IRAS 22159+5948.

2.2. AllWISE catalog

The AllWISE program combines the data from the Wide-field Infrared Survey Explorer (WISE; Wright et al. 2010) cryogenic and NEOWISE (Mainzer et al. 2011) post-cryogenic missions. The AllWISE catalog provides the photometry at 3.4 (W1),
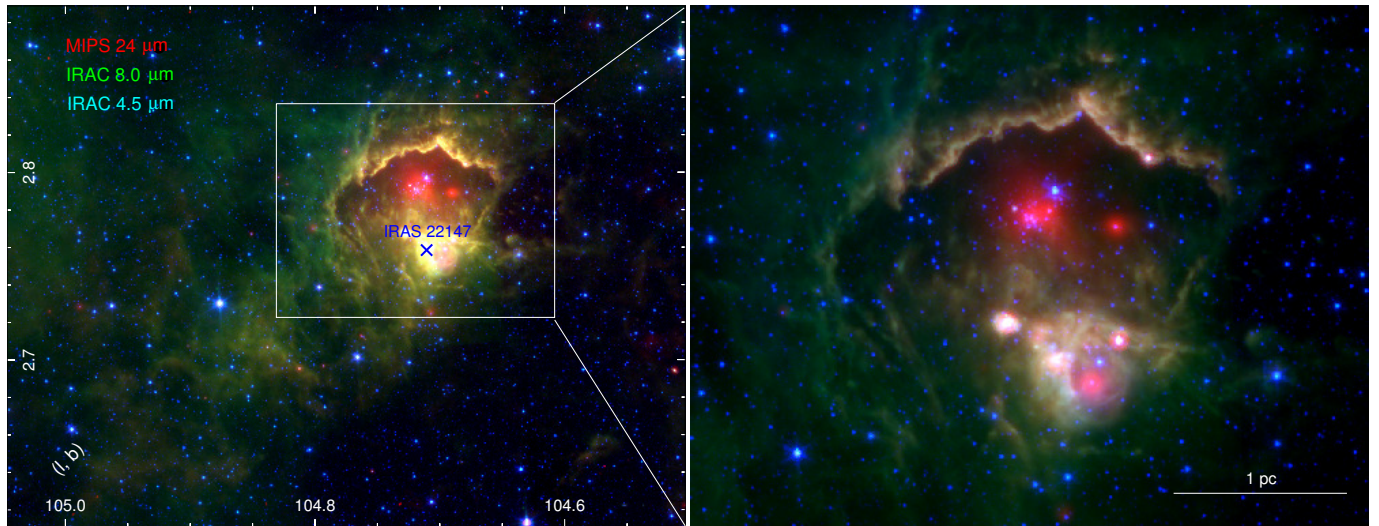
Fig. 1. Three-color composite image of the IRAS 22147 region combining the *Spitzer*/MIPS 24 μm (red) and *Spitzer*/IRAC 8.0 μm (green) and 4.5 μm (blue) mosaics. The position of IRAS 22147+5948 is indicated with an × symbol. The right panel shows a zoom-in on the most active region.

4.6 (W2), 12 (W3), and 22 μm (W4), with the resolutions of 6.″1, 6.″4, 6.″5, and 12″, respectively. To convert AllWISE 3.4, 4.6, 12 and 22 μm magnitudes to fluxes, we use the zero magnitude fluxes of 309.54, 171.787, 31.674, and 8.363 Jy (Jarrett et al. 2011), respectively.

2.3. Spitzer MIPS catalog

We use complementary MIPS 24 μm point source catalog provided by the SMOG team, with the angular resolution of ∼6″. In order to determine the most optimal search radius between the *Spitzer*/SMOG IRAC and MIPS catalogs, we performed simple Monte Carlo simulations. In each of 1000 realizations, an artificial SMOG catalog was constructed by shifting all the individual sources in the original catalog in random directions over a distance of 10″. The artificial catalog is then matched with the actual MIPS data, giving the average number of false matches as a function of the search radius. In order to determine the percentage of incorrect associations in a given sample, the number of false matches is divided by the number of counterparts found using the original catalog. For the *Spitzer* MIPS 24 μm catalog, we adopted a search radius of 1.8″, giving a false association rate of 10%, from which 244 counterparts were found.

2.4. Herschel EPoS survey

The long-wavelength measurements are taken from the catalog and original images from the Earliest Phases of Star formation (EPoS) *Herschel* key program (Ragan et al. 2012). The maps of the region were obtained with the Photodetector Array Camera and Spectrometer (PACS; Poglitsch et al. 2010) at 70, 100, and 160 μm, and the Spectral and Photometric Imaging REceiver (SPIRE; Griffin et al. 2010) at 250, 350, and 500 μm. IRAS 22147+5948 corresponds the EPoS target ISOSS J22164+6003 and photometry for five point-sources is provided in Ragan et al. (2012). Here, we calculate upper limits for additional sources detected in the IRAS 22147 region using the level 3 processed images from the *Herschel* Science Archive. We identify point sources as emission peaks and perform photometry by fitting the Gaussian to the source brightness

profile using the Curvature Threshold Extractor package (CuTEx; Molinari et al. 2011). The details of the procedure are described in Molinari et al. (2010) and Elia et al. (2010, 2013).

2.5. CO surveys

We used CO 1–3 data cubes from the Canadian Galactic Plane Survey (CGPS, Taylor et al. 2003) based on observations from the Five College Radio Astronomy Observatory (FCRAO) CO Survey of the Outer Galaxy (Heyer et al. 1998). The CGPS reduction included initial re-processing at DRAO, followed by data resampling and smoothing to the Nyquist resolution limit prior to spatial regridding (see Kerton & Brunt 2003, for details). The spatial resolution is 100.44″ instead of the FCRAO beamsize of 45″. The antenna temperature scale was corrected for forward scattering and spillover losses using a scaling factor of 0.7.

2.6. Maser surveys

There have been several prior attempts to detect masers towards the IRAS 22147 region, all with negative results. Palla et al. (1991) surveyed the occurrence of the H_2O 22 GHz maser emission from a sample of bright IR sources in different star-forming regions. A sample of 260 sources was selected from the IRAS Point Source Catalog, with no detection for IRAS 22147+5948. Kalenskij et al. (1992) searched for the 7_0-6_1 A^+CH_3OH line emission at 44 GHz with non-detections for IRAS 22147+5948 at the 10 Jy level. Wouterloot et al. (1993) used 100 m Effelsberg and 32 m Medicina radiotelescopes to search for H_2O maser emission at 22 GHz towards 1143 IRAS sources – again, with no success. Other failed attempts include non-detections of masing lines of SiO (Jiang et al. 1996), NH_3 (Molinari et al. 1996), OH (Edris et al. 2007) and CH_3OH (Fontani et al. 2010).

In addition, we conducted our own survey at 22 GHz using Toruń 32 m radio telescope. Two series of observations were performed in May 2018 and February 2019, with no detections of 22 GHz H_2O maser emission. We used two 4096 channel correlator parts with 8 MHz bandwidth each, which provided a velocity coverage from $-87\,\mathrm{km\,s^{-1}}$ to -33 km/s with respect to
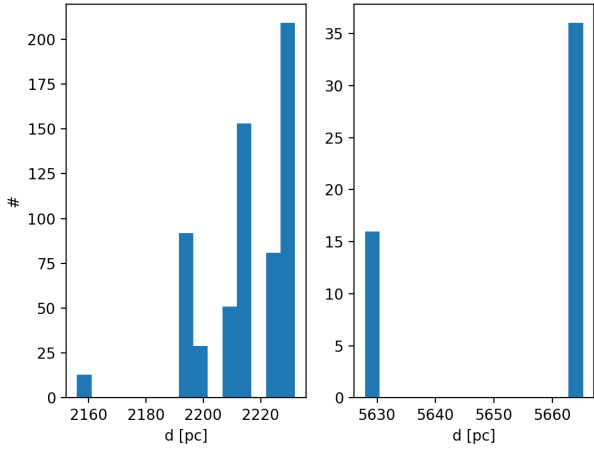
Fig. 2. Histogram of distances for sources in the IRAS 22147 region with highly reliable IRAC photometry ($\sigma < 0.2$ mag).

the local standard of rest. The system temperature was ∼75 K and ∼115 K during first and second part of our survey, which resulted in 3σ detection limits ∼3.5 Jy and ∼5 Jy, respectively.

3. Results and analysis

In this section, we identify YSO candidates in the IRAS 22147 region using color–color and color–magnitude diagrams, as well as the machine-learning techniques. We search for YSO candidate clusters and compare the positions of YSO candidates with the spatial extent of the CO 1–0 emission and dust continuum. Finally, we verify the physical properties of the YSO candidates using a modeling of the spectral energy distributions (SEDs).

3.1. Kinematic distances

To find a distance to a given source, we searched for its spatial association with the gas clumps, as traced by the CO 1–0 emission, where we used the data collected as part of CGPS (Taylor et al. 2003) and FCRAO surveys (Heyer et al. 1998). We searched for the spatially and kinematically coherent structures, namely, the CO 1–0 clumps, using the cloud decomposition techniques, that is, the CLUMPFIND algorithm in the CLOUDPROPS package (Williams et al. 1994; Rosolowsky & Leroy 2006). The algorithm considers all the spectral channels along the line of sight of a given source, which are associated with a CO 1–0 clump and assigns a distance to the source based on the clump with the closest centroid (Elia et al. 2013), adopting the Milky Way rotation curve of Brunthaler et al. (2011).

Figure 2 shows a distribution of distances associated with sources with IRAC photometry with $\sigma < 0.2$ identified in the IRAS 22147 region. The histogram shows a bimodal shape with distances clustering around 2.2 kpc and 5.6 kpc.

The distance of ∼5.6 kpc associated with YSO candidates in the central cluster is consistent with the determination by Ragan et al. (2012). The 5% difference is due to the adopted rotational curve (see Sect. 4.1). Appendix A shows a comparison of individual YSO kinematic distances to those obtained using the *Gaia* DR3 catalog (Bailer-Jones et al. 2021). The distances for two thirds of the sources show good agreement; for the remaining sources, Appendix A shows the impact of the distance determination on the analysis.

3.2. Color–color and color–magnitude diagrams

To identify YSO candidates using infrared colors, we followed the procedures described in Gutermuth et al. (2009), where 2548 YSO candidates have been identified inside 36 young, nearby, star-forming clusters, situated at distances between 0.14 and 1.7 kpc.

This approach adopts various color–color (CC) and color–magnitude (CM) cuts and it is, therefore, most reliable when applied to objects with known distances. We note here that Winston et al. (2019) adopted a more liberal version of the Gutermuth et al. (2009) algorithms in order to extend the method to objects without known distances. However, since we do require distances for the later stages of our analysis, namely, the SED fitting and spatial clustering, we strictly follow the more conservative, original approach of Gutermuth et al. (2009), where we rescaled all the magnitudes to the distance of 1 kpc.

The starting point of the analysis is the SMOG catalog with 18 017 sources lying within the IRAS 22147 field, of which 7 635 (∼42%) have known distances (see Sect. 3.1). We divide our YSO identification process into three phases, whereby we classify sources based on a number of color–color and color–magnitude cuts. The procedure, described in detail in the appendix of Gutermuth et al. (2009), can be summarized as follows.

In phase 1, YSOs were identified based on four *Spitzer*/IRAC bands, where we only considered objects with detections in all four bands with photometric uncertainties $\sigma < 0.2$ mag, resulting in a sample of 681 sources (∼9%). First, we separated out sources with the excess IR emission that are not YSOs. The active star-forming galaxies were eliminated based on their strong PAH-features with red 5.8 and 8.0 μm colors. Since these are quite distinct from the YSOs, the contamination from star-forming galaxies in the final sample is expected to be negligible. The broad-line AGNs, with mid-IR colors largely consistent with YSOs (Stern et al. 2005), were removed utilizing the [4.5] vs. [4.5]−[8.0] color–magnitude diagram. Finally, the unresolved knots of shock emission, often detected in all IRAC bands, as well as resolved structured PAH-emission producing excess emission in the 5.8 and 8.0 μm bandpasses, were eliminated. The YSOs were then selected from the cleaned sample, where we found 14 Class I and 9 Class II sources (Fig. 3).

Phase 2 was applied to sources lacking 5.8 or 8.0 μm detections, but having high quality ($\sigma < 0.1$ mag) 2MASS data in H and K_s (J can also be used when present). Out of 280 sources passing those criteria, only 3 are located within the Class II region of the diagram (Fig. 4).

Finally, in phase 3, we utilized the MIPS 24 μm data in order to re-examine the original catalog. In this step we find one of our Class I YSOs from Phase 1 to lack adequate 24 μm excess, which reclassifies it as a heavily reddened Class II YSO (green diamond in Fig. 3). We therefore ended up with 26 YSO candidates, namely, 13 of Class I and 13 of Class II.

In order to maximize the YSO selection process by including the available AllWISE data, we utilized our YSO candidate sample as a training set for a machine learning algorithm, as described in the following section.

3.3. Machine learning technique

Machine learning algorithms are a branch of artificial intelligence whereby the algorithm is designed to learn from data, identify patterns or structures within the data, and then make predictions with minimal human involvement. Here, we used
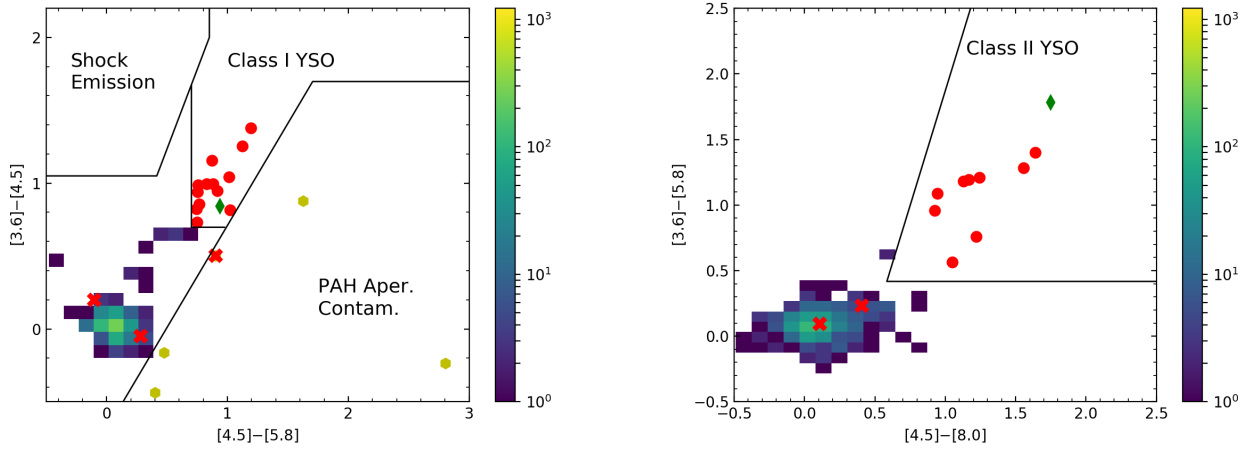
Fig. 3. Color-color diagrams using the photometry from *Spitzer*/IRAC. *Left:* [3.6]–[4.5] vs. [4.5]–[5.8] diagram for the IRAS 22147 region, where all 681 sources with distances and IRAC errors $\sigma < 0.2$ mag are shown. Normal stars are depicted as a Hess diagram (color bar), and excess PAH emission contaminants are shown with light green hexagons and Class I YSOs with red circles. The dark green diamond represents the source initially classified as a Class I YSO, but later re-classified as a heavily reddened Class II source, based on its 24 μm flux. Red × symbols indicate three objects identified as YSOs by Winston et al. (2019) but failing to pass our selection criteria (see Sect. 3.4). *Right:* [3.6]–[5.8] vs. [4.5]–[8.0] diagram for the IRAS 22147 region, with Class I YSOs and excess PAH emission contaminants from the left panel removed. Symbols are the same as in the left panel.
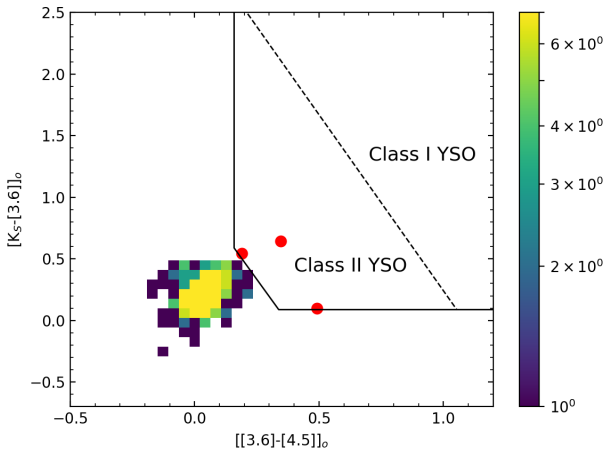

Fig. 4. $[K_s–[3.6]]_0$ vs. $[[3.6]–[4.5]]_0$ diagram for the IRAS 22147 region, with stars and Class II YSOs depicted as Hess diagram and red circles, respectively.

a combination of the machine learning anomaly detection and supervised algorithms to identify candidate YSOs in the IRAS 22147 region. We followed procedures developed specifically for the analysis of the AllWISE catalog, which led to the identification of, for example, heavily reddened AGN and quasar candidates that were originally missing from source catalogs created using standard cuts on CC diagrams (Solarz et al. 2017). The process of selection is applied to the entire SMOG region, but the output catalog discussed here includes only the IRAS 22147 region.

As a first step, we distinguished sources with YSO signatures from the rest of the Galactic sources. For that purpose, we used the anomaly detection algorithm called the one-class support vector machines (OCSVM, Schölkopf et al. 1999) with the R interface for libsvm (R Core Team 2018; Meyer et al. 2020). In particular, OCSVM is designed to select outliers in the data based on the input data, where all known objects have a single label: "known." As the training objects, we used all the YSOs selected through the *Spitzer*'s color–color diagram

method (Sect. 3.2) that have a counterpart in the AllWISE data (151 Class I and 346 Class II objects).

The training process of the OCSVM starts with the creation of a feature vector associated with each training point, composed of N quantities that describe the discriminating properties of a given object. In case of the AllWISE data, we use 4 AllWISE fluxes and three color combinations ($W1–W2$, $W2 − W3$, and $W3 − W4$), so, in total, a 7D input parameter space. Subsequently, these numbers would act as coordinates for each training object (vector) in 7D feature space. Using the pre-defined training sets of sources, the algorithms used the kernel functions to map the input parameter space into a higher dimensional feature space (Shawe-Taylor & Cristianini 2004). The algorithm will search for a hyperplane, which separates the training points and creates a "normality" model. Once this process is complete, the new data, which does not match the "known" object patterns, are flagged as outliers.

In the IRAS 22147 region, we focus on sources which display the behavior similar to known YSOs. Therefore, we use the OCSVM to perform a "reversed" anomaly detection, that is, we train the algorithm on the known YSOs and we reject all those sources which do not fit the model created based on the "known" YSOs. Using the sample of general YSO candidates selected by the OCSVM, we proceeded to classify the sources as either Class I or II. For that purpose, we used the classic support vector machine algorithm, which is designed to recognize two (or more) types of objects based on the training examples provided by the supervisor (SVM, Vapnik 1995).

The process is similar to the OCSVM, except that the algorithm searches for the most optimal hyperplane separating the examples of the two categories, instead of searching for the most optimal enclosure for these sources. The hyperplane separates the examples from two categories with the largest possible margin. We used the same training sample as for the anomaly detection, but divided into separate Class I and II groups.

The number of Class I and II sources detected in the IRAS 22147 region using the machine learning techniques is 2 and 21, respectively. Out of these sources, only 9 Class II sources are assigned kinematic distances. The left panel of Fig. 5 shows the YSO candidates in the $W1–W2$ vs. $W2–W3$ diagram. All
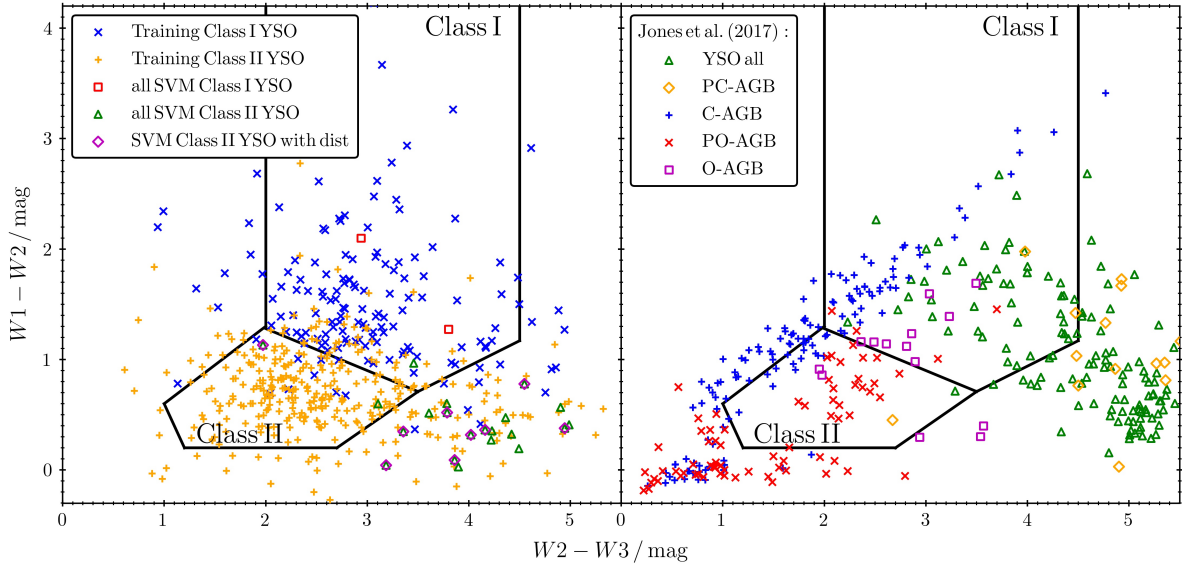
Fig. 5. Color–color diagrams using the AllWISE catalog. *Left:* $W1$–$W2$ vs. $W2$–$W3$ diagram with SVM selected YSO Class I (red squares) and Class II (green triangles) candidates. Pink diamonds mark the position of SVM-selected YSO candidates with measured distances for Class II. Blue × and orange + symbols show the position of the training YSO I (151 objects) and II (346 objects), respectively, based on the *Spitzer* selection in the entire SMOG survey area. Black lines mark the boundaries of color spaces within which Class I and Class II are expected to lie following the work of Fischer et al. (2016). *Right:* $W1$–$W2$ vs. $W2$–$W3$ diagram showing the positions of spectroscopically confirmed YSOs and AGB stars from Jones et al. (2017). Green triangles mark all YSOs, yellow diamonds mark the carbon-rich post AGB stars (PC-AGB), blue crosses signs mark the carbon-rich AGB stars (C-AGB), magenta × symbols mark oxygen-rich post AGB stars (PO-AGB), and pink × symbols mark the oxygen-rich AGB stars (O-AGB).

Class I and II candidates identified with the machine-learning algorithms are located in the parts of the diagram, where the training sample obtained using the color–color and color–magnitude diagrams is also located (Sect. 3.2). The right panel of Fig. 5 shows the WISE colors of different types of dusty objects from Large Magellanic Cloud (LMC), classified by Jones et al. (2017) based on the *Spitzer*/IRS spectroscopic observations. The Jones et al. (2017) catalog includes evolved objects: carbon and oxygen rich asymptotic giant branch (AGB) and post-AGB stars (C-AGB, O-AGB, PC-AGB and PO-AGB, respectively), as well as different types of YSOs. We can see that the majority of the spectroscopically observed YSOs in LMC lie outside of the Fischer et al. (2016) YSO selection lines.

The SVM method selects YSO candidates, which lie in the vicinity of the border proposed for Class II YSOs, with the $W2$–$W3$ color reddened by up to 2 mag (Fischer et al. 2016). A similar area of the CCD is occupied by confirmed YSOs from Taurus (Koenig & Leisawitz 2014) and star-forming regions in the LMC (Jones et al. 2017); however, it might be to some extent contaminated by star-forming galaxies. Thus, for the sake of catalog reliability, the region with numerous galaxies is not recommended for the YSO classification based solely on the IR photometry (Koenig & Leisawitz 2014; Fischer et al. 2016). Here, we find that all the SVM selected candidate YSOs are associated with CO emission (see Sect. 3.6), which implicates that they are likely to be well-identified by the algorithm. However, a contamination from the PC-AGB stars could be still present.

3.4. Comparison to Winston et al. (2019) and completeness

Adopting color–color and color–magnitude cuts, as well as machine learning techniques, we end up with a sample of 49 YSO candidates, 15 Class I and 34 Class II (see Appendix B for their photometry). Winston et al. (2019), within the IRAS 22147 region, identified 49 YSOs, 16 Class I, 29 Class II, and 4 Class III. In this section, we discuss some of the differences between both samples and completeness of the final sample.

There are 24 YSOs, which were identified in both samples. Thirteen Class I and 11 Class II objects have been identified in this work, while Winston et al. (2019) finds 11 Class I and 13 Class II sources (see Appendix C for a detailed comparison). The small discrepancies in the classification of the identified YSOs can be attributed to the subtle differences in the color–color and color–magnitude cuts applied in both works.

Twenty five sources identified in Winston et al. (2019) do not appear in our sample: 9 of them have errors in *Spitzer*/IRAC bands larger than 0.2 mag, while 13 were not assigned distances, which immediately removes them from our sample. One was classified as a resolved structured PAH-emission source, while the remaining 2, again, were not identified in this work due to differences in cuts used. We show them as red crosses in Fig. 3.

Finally, 25 sources have been classified as YSOs in this work but do not appear in Winston et al. (2019). Eighteen of them have been classified using the AllWISE data. While Winston et al. (2019) uses color–color cuts in 4 WISE bands from Koenig & Leisawitz (2014) and Fischer et al. (2016) (black lines in Fig. 5), we adopted the machine learning algorithms (Sect. 3.3) based on the training sample of candidate YSOs selected using IRAC color–color and color–magnitude diagrams. In the left panel of Fig. 5 we can see that out of 21 sources classified as YSO Class II using ML techniques (green triangles), 18 fall outside the region of the color–color diagram used by Winston et al. (2019). The remaining 7 YSOs that do not appear in Winston et al. (2019) have been identified in this work using the IRAC CC/CM diagrams. As above, this is caused by the Winston et al. (2019) adjusting the original Gutermuth et al. (2009) cuts to objects at larger distances, which resulted in subtle cut differences.

Apart from the differences in the identification methods, the number of YSO detections depends on the sensitivity of the instruments and their capability to detect point-sources at a given distance. In the case of Class 0/I protostars, an empirical relation between the sensitivity in luminosity and the distance of the object has been proposed by Dunham et al. (2006, 2008), and subsequently confirmed for objects in the outer Galaxy by Ragan et al. (2012). Accordingly, the limiting luminosities of 1.0 and 6.4 $L_\odot$ are expected for Class 0/I sources at ~2.2 kpc and ~5.6 kpc, respectively, in agreement with bolometric luminosities (Appendix D).

Additional limitation on the detection of the most embedded, namely, the least evolved, YSOs stem from the adopted CC cuts, which require that the target is detected in all four *Spitzer*/IRAC bands. This could be the case for Class 0 YSOs, whose SEDs are dominated by the cold, dense envelopes with the peak around 100 μm, and very little or no near-IR emission is observed (Andre et al. 1993; Evans et al. 2009; Karska et al. 2018). The *Herschel*/PACS spectral maps allowed the identification of five deeply-embedded YSOs in the IRAS 22147 region with $L_{\rm bol}$ from 24 to 260 $L_\odot$ (Ragan et al. 2012). Two sources, G104.6895+02.7945 and G104.7270+2.8003, are identified as Class I YSOs using the CC diagram method presented here; however, the three remaining sources – including IRAS 22147+5948 – lack the near-IR photometry in the *Spitzer* point source catalog. Similarly, two IRAS sources associated with an arc of emission to the north of IRAS 22147+5948 also lack the *Spitzer*/IRAC emission and are not classified here.

3.5. Spatial clustering of selected YSOs

The spatial distribution of YSO candidates provides support for their physical association and likely a shared origin. For the case of the Outer Galaxy sources, it is yet another means of avoiding the cloud confusion and confirming the YSO identification.

To identify clusters in the IRAS 22147 region, we used the agglomerative hierarchical clustering algorithm (Everitt et al. 2011), which belongs to the unsupervised family of machine learning. Similar to all clustering algorithms, it serves to classify a set of objects into subgroups that display similar characteristics. The definition of the similarity measure among objects depends on the metrics definition.

As a first step, the algorithm assigns each data point in its own cluster. Then, it identifies the closest two clusters and combines them into one. This process is re-iterated until all data points reside within a single cluster. After the YSOs have been combined into a single cluster, the user chooses a critical length and a minimum membership number. The process can be represented by a dendrogram structure showing a hierarchical relationship between objects. The advantage of the adopted clustering algorithm is the fact that the user does not have to define the number of classes a priori, but can choose a fitting number once dendrograms are calculated. Following Gutermuth et al. (2009), we employed single-linkage agglomerative clustering where links longer than the critical length are cut, and the remaining clusters with at least a minimum number of members are preserved.

We calculated the projected distances in a standard way using source coordinates, but separately for sources at distances of ~2.2 and 5.6 kpc. The exact distance regimes are chosen based on the distance histogram (Fig. 2), with sources at the distance from 1.2 to 2.4 kpc referred to as 2.2 kpc, and those with distances above 4 kpc referred to as 5.6 kpc (Fig. 6). Due to a visual inspection of dendrograms, a critical length of 170
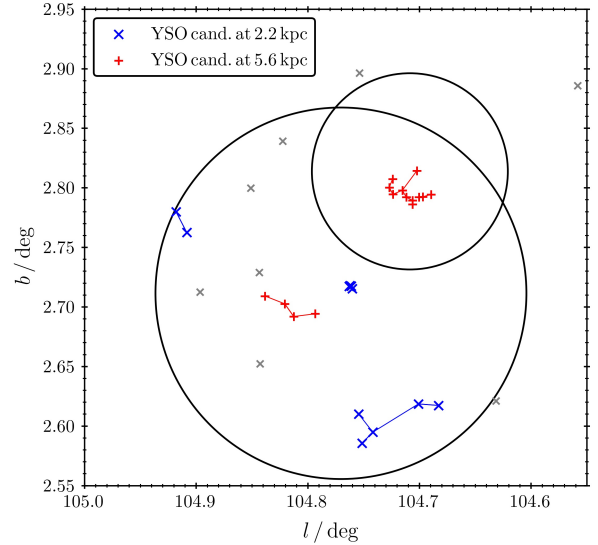

Fig. 6. Minimum spanning tree for YSO candidates in the IRAS 22147 region. Blue and red × symbols show objects belonging to clusters located at $d \sim 2.2$ and $d \sim 5.6$ kpc, respectively; the isolated objects are shown as gray × symbols. Blue and red lines correspond to branches connecting objects from the same clusters. Black circles correspond to the two clusters identified in the region by Winston et al. (2019).

Table 1. Clusters of YSOs in the IRAS 22147 region.

d (kpc)	l (°)	b (°)	$N_{\rm I}$	$N_{\rm II}$	$N_{\rm II}/N_{\rm I}$
2.207	104.726	2.605	1	4	4.00
2.215	104.762	2.717	2	2	1.00
2.223	104.913	2.771	2	0	–
5.628	104.709	2.797	3	8	2.67
5.665	104.816	2.699	3	1	3.00

Notes. Coordinates (l, b) correspond to the location of the center of the clusters. $N_{\rm I}$ and $N_{\rm II}$ are numbers of Class I and Class II YSO candidates in the clusters, respectively.

and 100 arcsec was set for objects at roughtly 2.2. and 5.6 kpc, respectively.

Figure 6 shows the minimum spanning tree for clusters associated with IRAS 22147+5948. Five clusters containing between 2 and 11 members, and a total of 26 YSO candidates, are identified (see also Table 1). Eight sources are not associated with any of the clusters. We note that there are three clusters at ~2.2 kpc and one cluster at ~5.6 kpc covering the area of a single cluster identified in Winston et al. (2019), using the Density-Based Spatial Clustering and Application with Noise (DBSCAN, Ester et al. 1996), see Sect. 4.2. In their analysis, no distance information were available, causing the line-of-sight confusion (see Fig. 6).

Fischer et al. (2016) note that clusters with fewer than five members should be discounted as unrealistic chains and may be produced by the single-linkage method. Accordingly, we would have two fully reliable clusters corresponding to two distance regimes with four and with eight members. Yet, there are clearly two separate clusters at 5.6 kpc in the considered region (Fig. 1), corresponding to IRAS 22147+5948 in N–W and IRAS 22159+5948 in S–E. These two clusters are likely physically linked, but nevertheless, we chose to keep them separate in our identification.
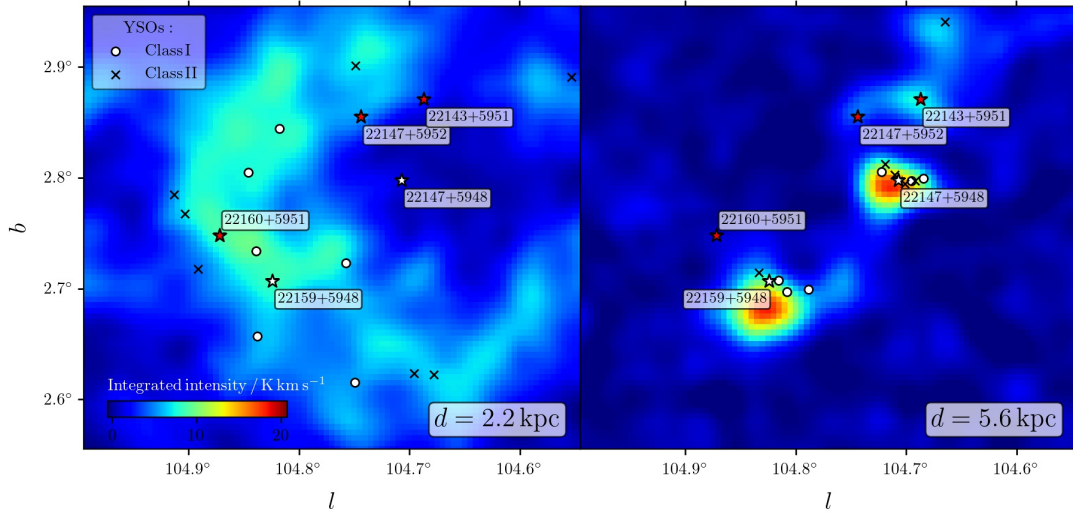
Fig. 7. Integrated intensity CO 1–0 maps from FCRAO using velocity ranges from $-25.4\,\mathrm{km\,s^{-1}}$ to $-23.7\,\mathrm{km\,s^{-1}}$ (*left*) and $-60.8\,\mathrm{km\,s^{-1}}$ to $-58.4\,\mathrm{km\,s^{-1}}$ (*right*). The positions of YSOs, as classified via CCD/ML methods (Sects. 3.2 and 3.3), are shown in circles (Class I) and × symbols (Class II). Positions of IRAS 22147+5948 and IRAS 22159+5948 are shown in white stars, and those of other IRAS sources in the region as red stars.

3.6. Spatial extent of CO gas and dust

Young stellar objects at early evolutionary stages are associated with gas and dust from their nascent molecular cloud cores. Extended CO 1–0 emission pin-points gas in the envelopes and outflows from Class 0/I protostars. Dust emission traces star forming filaments, cores and envelopes. Appendix E shows images of the IRAS region in individual bands from near-IR to radio wavelengths.

Figure 7 shows the spatial distribution of CO 1–0 emission, separately for the two line components which correspond to distances of ∼2.2 and 5.6 kpc (see Sect. 3.1). The CO gas at ∼2.2 kpc forms an extended arc to the east and south from IRAS 22147+5948. The component at ∼5.6 kpc shows two gas concentrations which corresponds to the positions of IRAS 22147+5948 and IRAS 22159+5948. Additionally, a weak arc of emission is detected to the north of IRAS 22147+5948 and a bridge in between the two IRAS sources.

The positions of YSO candidates identified in Sect. 3.2 and 3.3 closely follow the distribution of CO 1–0, supporting their YSO classification; YSOs at ∼2.2 kpc are consistent with a few small clusters which are rather loosely distributed (see Sect. 3.5). The objects at ∼5.6 kpc form well-defined clusters, consisting of Class 0/I candidates, as opposed to the Class II and III populations found at 2.2 kpc.

The spatial resolution of the CGPS CO 1–0 cube of the order of 100″ does not allow us to isolate emission from individual objects within the cluster associated with IRAS 22147+5948. Yet, in locations corresponding to clusters of YSOs, emission in the CO 1–0 line wings has been detected, which is a signature of outflows and the young evolutionary status of the driving sources (Class 0/I). Higher spectral and angular resolution observations are necessary to characterize outflows and envelopes from individual YSOs.

Figure 8 shows the spatial distribution of the dust continuum emission in the far-infrared, its temperature, and a comparison with the CO 1–0 emission (see also Fig. 7). Observations available from the EPoS survey cover only the west part of the region containing the cluster associated with IRAS 22147+5948.

Emission from the dust is co-spatial with CO gas detected at ∼5.6 kpc, strongly suggesting that the most prominent structures seen in Fig. 1 all belong to IRAS 22147+5948 and not the star-forming region at ∼2.2 kpc. Consequently, the dust also follows the positions of YSO candidates detected in the cluster, as well as the northern arc of material.

The dust temperatures were calculated using the photometry from *Herschel* at 70 and 160 μm, assuming a power-law dust emissivity (Eq. (1)), where B_ν is the Planck function and β is the emissivity index:

$$I_\nu \propto \lambda^{-\beta} B_\nu(T_{\mathrm{dust}}). \tag{1}$$

We adopted a constant value of β equal to 1.7, which is typical for envelopes of Class 0 and I YSOs (Kristensen et al. 2012; Goicoechea et al. 2012). The dust temperatures range between 15 and 30 K, in agreement with typical temperatures in sites of low-mass star formation (Kristensen et al. 2012). Temperature peaks of ∼25 K typically omit the positions of YSO candidates with one exception, corresponding to evolved Class II and III candidates at the far-east of the cluster. The temperature increase is most prominent at the northern edge of the cluster, which is devoid of YSO candidates found in Sects. 3.2 and 3.3. This region corresponds to the bright MIPS 24 μm emission in Fig. 1, which contains several sources that were not identified as YSOs.

To summarize, the spatial extent of CO gas and dust provides important implications for the region and help distinguish star forming regions along the same line-of-sight. The distance calculated from CO emission verifies the clustering in IRAS 22147+5948 and favors the interpretation of two small clusters at 5.6 kpc (Sect. 3.5) instead of the cluster identification proposed by Winston et al. (2019). The CO 1–0 line wings detected toward the IRAS 22147 region confirm the formation of deeply-embedded protostars in the region. The gas and dust distribution together explain the composite mosaics images and imply the physical connection of the observed structures.

3.7. Spectral energy distributions

An additional test of the YSO identification and classification is available via the modeling of their broad-band SEDs. Here, we use the Robitaille (2017) set of YSO model SEDs and dedicated
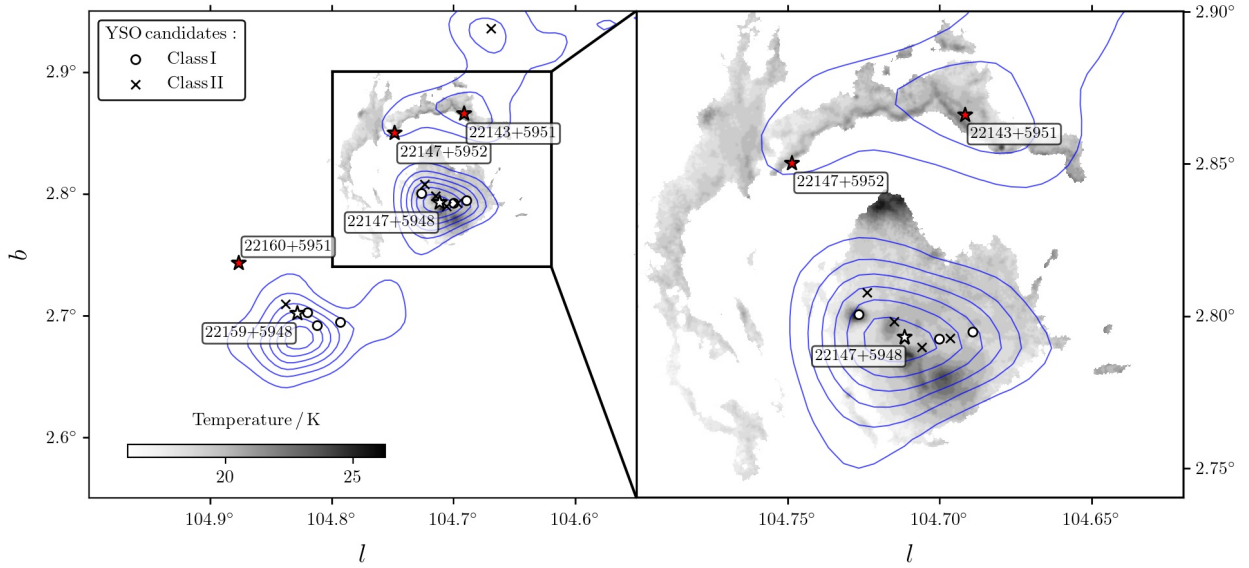
Fig. 8. Dust temperature map (grayscale) and CO 1–0 map (blue contours) for the gas components at the distance of 5.6 kpc. The dust temperature distribution is limited to the central region, covered by the *Herschel* EPoS key program (Sect. 2.4 and Fig. E.1) and, therefore, does not cover the south-east feature. The CO contours start at 2.7 K km s^{-1} (5 σ) and are drawn in steps of 2.7 K km s^{-1}; the range of CO integrated intensities is shown on Fig. 7. The positions of YSOs, as classified via CCD/ML methods (Sects. 3.2 and 3.3), are shown in circles (Class I) and × symbols (Class II).

Table 2. Best-fit models of the spectral energy distribution using the Robitaille (2017) classification.

Model set	# of sources	Components	Group
sp−−s−i	12	star + passive disk; $R_{\text{inner}} = R_{\text{sub}}$	d
sp−−h−i	4	star + passive disk; variable R_{inner}	d
sp−−smi	5	star + passive disk + medium; $R_{\text{inner}} = R_{\text{sub}}$	d
sp−−hmi	2	star + passive disk + medium; variable R_{inner}	d
s−ubsmi	1	star + Ulrich envelope + cavity + medium; $R_{\text{inner}} = R_{\text{sub}}$	e
spu−hmi	1	star + passive disk + Ulrich envelope + medium; variable R_{inner}	$d + e$
spubsmi	1	star + passive disk + Ulrich envelope + cavity + medium; $R_{\text{inner}} = R_{\text{sub}}$	$d + e$

Notes. The model set names are from Robitaille (2017). Seven characters in the model set names indicate which component is present; they are (in order): s: star; p: passive disk, p or u: power-law or Ulrich envelope; b: bipolar cavities; h: inner hole; m: ambient medium; and i: interstellar dust. A dash (−) is used when a component is absent. R_{inner} is the inner radius for the disk, envelope, and the ambient medium – when one or more of these components are present. R_{sub} is the dust sublimation radius.

fitting tool (Robitaille et al. 2007) to reproduce multi-wavelength photometry of YSO candidates.

The Robitaille (2017) SED models are divided into 18 sets consisting of a combination of a number of components: star, disc, infalling envelope, bipolar cavities, and an ambient medium (Table 2). The intention is to span a wide range of parameter space, however, as a result many of the models are unphysical. Therefore, we limit our fitting procedure to the models with positions on the Hertzsprung-Russell (HR) diagram that are consistent with the PARSEC evolutionary tracks produced by the revised Padova code (Bressan et al. 2012; Chen et al. 2014, 2015; Tang et al. 2014). We excluded the YSO models located outside the coverage of the PARSEC pre-main sequence stage tracks.

In the fitting procedure, we set a 10% error on the distance and varied the interstellar extinction, A_V, in the range from 0 to 40, using the Weingartner & Draine (2001) extinction law for $R_V = 4.0$. We model only the sources which have at least five photometric detections, including at least one at wavelength $\gtrsim 12 \, \mu$m. In addition, we use only AllWISE 12 and 22 μm data, since the 3.5 and 4.6 μm wavebands are similar to those

of *Spitzer*, but with significantly lower angular resolution (6″ in WISE vs. 2″ *Spitzer*). In the IRAS 22147 region, a total of 26 sources pass these criteria.

In order to find the best fit, the code initially determines the fit with the lowest value of χ^2 and χ^2_{best}. In each of the model sets, the number of fits with the $\chi^2 < (\chi^2_{\text{best}} + F)$ is counted, where F is a threshold parameter which we set to 3 (Sewiło et al. 2019). The best SED fit is the one with the lowest χ^2 from the model set with the largest number of good fits. The above conditions are adopted to ensure the proper selection of the model and limit the risks stemming from the fact that the SED fitting is degenerate, with some of the model sets including 13 parameters.

Table 2 gives the statistics of the best-fit SED models for the YSO candidates in the IRAS 22147 region, with Fig. 9 showing two example best-fit SEDs. Twenty-three sources are successfully modeled with a star and a passive disk, corresponding to Class II+III sources. The model for one source requires the presence of a protostellar envelope (Class 0), with an additional two features, including envelopes as well as passive disks (Class I).
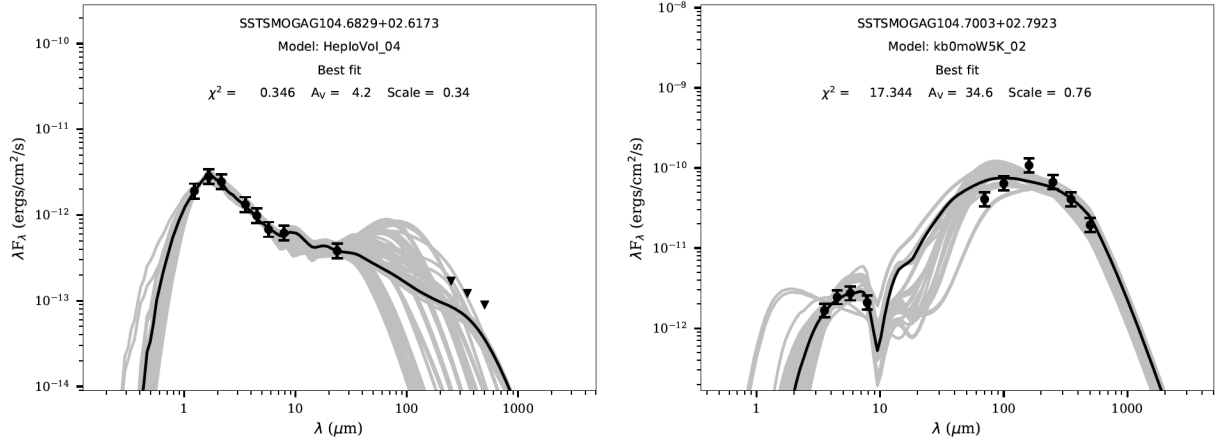
Fig. 9. Example SEDs of candidate YSOs with well-fit Robitaille (2017) YSO models. The best fit model is indicated with the black solid line; gray lines show the YSO models with χ^2 between χ^2_{best} and $\chi^2_{\text{best}}+F$, where F is a threshold parameter which we set to 3 (Sewiło et al. 2019). Filled circles and triangles are valid flux values and flux upper limits, respectively. The values of a reduced χ^2 and interstellar visual extinction for the best-fit model are indicated in the plots. Appendix F shows the SEDs for the remaining YSOs in the IRAS 22147 region.

In order to determine stellar luminosities, masses, and ages for our sample, which are not provided by the Robitaille (2017) SED models, we use the following procedure. First, we find the luminosity using the Stefan-Boltzmann law, adopting the stellar radius and effective temperature from the SED fitting. Then, the mass and age are determined from the closest pre-main sequence (PMS) track, found for each model on the HR diagram (see Appendix F).

We adopted the stellar masses only for the sources, for which the age from PMS tracks is consistent with the SED fitting value. We use the YSO lifetimes determined by Dunham et al. (2015), based on a sample of ~3000 YSO candidates from nearby star-forming regions, with Class 0+I YSOs lifetimes of 0.46–0.72 Myr (and duration of 0.335 and 0.665 × lifetime for Class 0 and I, respectively), flat-spectrum sources lifetimes of 0.30–0.46 Myr and 3 Myr for Class II+III.

Table 3 shows a comparison of the YSO classification using color–color diagrams and the SED modeling, and the corresponding physical parameters. Fully consistent object classification with the two methods is obtained for 15 out of 26 sources including two Class I sources and 13 Class II and III sources. One of the Class II YSOs from the color–color diagrams is reclassified as Class 0 source using SED modelling. The remaining ten sources originally identified as embedded Class I protostars are identified as more evolved Class II objects based on the SED models. Without far-IR data, it is difficult to distinguish Class I YSOs from reddened Class II YSOs (Kryukova et al. 2012). Two of these ten sources have far-IR photometry that supports their identification as Class II objects. The nature of the other eight sources is more ambiguous due to the large range of far-IR colors among their best-fitting SED models (Ali et al. 2010), but they likely lie near the Class I-II transition. Given these uncertainties, we adopted a classification based on color–color diagrams for the discussion of YSO candidates in the IRAS 22147 region.

Figure 10 shows the Hertzsprung-Russell Diagram with the positions of the YSOs obtained from the SED modeling (Table 3). The spread in YSO age is similar to the one found for the Orion Trapezium Cluster, however, at lower luminosities (Fang et al. 2021). Due to larger distances probed here, the YSOs in the IRAS 22147 region probe higher mass tip of a full mass distribution. The split in the 5.6 kpc group, with about half the stars at young ages and half quite close to the main
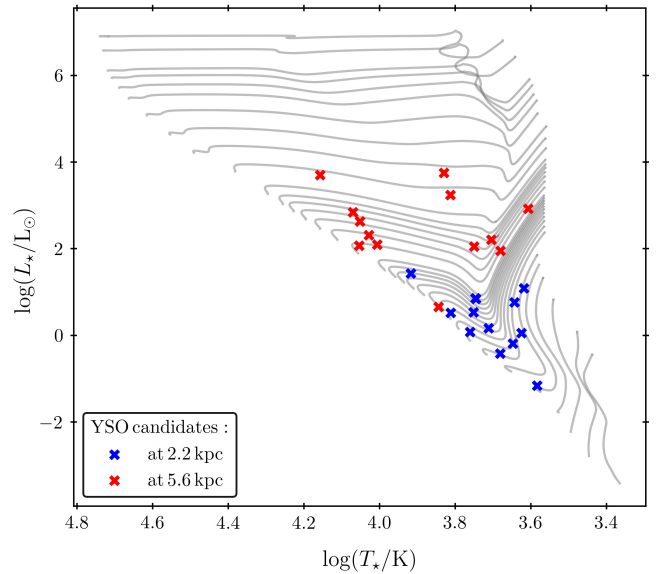

Fig. 10. HR diagram with YSOs in the IRAS 22147 region and the PARSEC evolutionary tracks (Bressan et al. 2012; Chen et al. 2014, 2015; Tang et al. 2014). YSOs located at ~2.2 kpc are in blue, and at ~5.6 kpc are in red. Every third track is drawn for clarity.

sequence, might be an evidence for two populations. Follow-up spectroscopy would be required to confirm such a scenario.

4. Discussion

4.1. Star formation in the IRAS 22147 region

A multi-wavelength analysis of the IRAS 22147 region confirms an on-going star formation associated with IRAS 22147+5948, as well as with the nearby IRAS 22159+5948 (Sect. 3). Altogether, 26 YSOs have been identified and classified as Class 0, I, or II objects using standard color–color diagrams, the OCSVM machine learning algorithm, and the SED modeling. Out of 26 sources, 13 are spatially associated with the CO 1–0 emission at 5.6 kpc and the dust emission seen in Fig. 1. Those sources are grouped in two compact clusters (Fig. 6), while the

Table 3. Physical parameters for a subset of YSO Candidates in the IRAS 22147 region with at least five photometry points, including one at a wavelength longer than 12 μm.

Nr	IRAC designation 'SSTSMOGA'	Model Set[a]	Class[b] init	Class[c] SED	Selection method[d]	$R_\star$ ($R_\odot$)	$T_\star$ (K)	$L_\star^{(e)}$ ($L_\odot$)	$M_\star^{(f)}$ ($M_\odot$)	Age[f] (Myr)
				YSO candidates at ~5.6 kpc						
1	G104.6697+02.9357	sp–smi	II	II+III	ML	2.8	11 340	120	3.0	2.3
2	G104.6895+02.7945	sp–smi	I	II+III	CC	54.5	6760	5600	–	–
3	G104.6968+02.7925	s-ubsmi	II	0	ML	4.2	10 690	210	–	–
4	G104.7003+02.7923	spubsmi	I	I	CC	16.4	5070	160	–	–
5	G104.7062+02.7895[g]	sp–h-i	II	II+III	ML	32.9	6500	1750	–	–
6	G104.7118+02.7921	sp–h-i	II	II+III	ML	58.9	4050	840	–	–
7	G104.7153+02.7979	sp–h-i	II	II+III	ML	5.4	11 280	430	4.4	0.8
8	G104.7241+02.8075	sp–h-i	II	II+III	ML	11.2	5630	110	–	–
9	G104.7270+02.8003	sp–smi	I	II+III	CC	11.5	14 370	5050	–	–
10	G104.7934+02.6942	sp–s-i	I	II+III	CC	3.6	10 150	120	3.0	2.2
11	G104.8126+02.6920	sp–s-i	I	II+III	CC	6.4	11 760	700	–	–
12	G104.8206+02.7026	sp–smi	I	II+III	CC	1.5	6980	5	–	–
13	G104.8383+02.7091	sp–s-i	II	II+III	CC	13.8	4790	90	–	–
				YSO candidates at ~2.2 kpc						
1	G104.5582+02.8857	sp–s-i	II	II+III	CC	2.8	5560	7	2.1	2.4
2	G104.6829+02.6173	sp–s-i	II	II+III	CC	2.0	4220	1	0.7	1.0
3	G104.7009+02.6186	sp–s-i	II	II+III	CC	1.1	5770	1	–	–
4	G104.7538+02.8964[g]	sp–hmi	II	II+III	ML	1.4	6490	3	–	–
5	G104.7544+02.6101	sp–s-i	I	II+III	CC	6.7	4160	12	–	–
6	G104.7622+02.7179	spu-hmi	I	I	CC	4.1	4400	6	0.8	0.2
7	G104.8225+02.8393	sp–s-i	I	II+III	CC	1.5	5150	2	1.4	4.1
8	G104.8429+02.6522[g]	sp–s-i	I	II+III	CC	0.6	3840	0.1	–	–
9	G104.8436+02.7289	sp–s-i	I	II+III	CC	1.4	4450	0.6	–	–
10	G104.8510+02.7998	sp–s-i	I	II+III	CC	2.9	5580	7	2.1	2.3
11	G104.8965+02.7124	sp–smi	II	II+III	ML	1.9	5650	3	–	–
12	G104.9084+02.7625[g]	sp–s-i	II	II+III	CC	2.5	8270	27	–	–
13	G104.9180+02.7798	sp–hmi	II	II+III	ML	0.9	4800	0.4	–	–

Notes. [a]See Table 2 footnotes for the meaning of characters in the model set names. [b]YSO class as initially determined using color–color diagrams and machine learning algorithms (Sects. 3.2 and 3.3, respectively). [c]YSO class as determined from the SED fitting (see Sect. 3.7). [d]Initial selection methods used in this work are the color–color/color–magnitude diagrams (CC; Sect. 3.2) and machine learning algorithms (ML; Sect. 3.3). [e]Stellar luminosities are calculated using the formula $L_\star = 4\pi R_\star^2 \sigma_{SB} T_\star^4$, where σ_{SB} is the Stefan–Boltzmann constant. [f]We provide stellar masses ($M_\star$) and ages (Age) only for sources with the most reliable estimates of these parameters (see Sect. 3.7). [g]A more reliable distance for this source is provided by *Gaia*; see Appendix A for the SED model adopting the *Gaia* DR3 distance from Bailer-Jones et al. (2021).

remaining YSOs form a more loosely distributed, foreground group at 2.2 kpc. Star formation in the smaller cluster toward IRAS 22159+5948 might have been triggered by the earlier formation of stars near the position of IRAS 22147+5948, given that third-fourths of YSOs, there are Class I objects based on the color–color diagrams method with regard to third-ninths of objects in the main cluster.

The main ambiguity in the interpretation of the IRAS 22147 region was due to early observations in the radio regime, showing a steep radio spectrum typical for supernovae remnants (Green & Joncas 1994). However, subsequent measurements at 10.5 GHz[1] were no longer consistent with the non-thermal emission. Subsequent detection of radio recombination lines with the Green Bank Telescope H II Region Discovery Survey (Bania et al. 2010) and significant infrared emission with WISE (Anderson et al. 2015) provided strong arguments in favor of the source classification as the H II region.

Figure 11 compares the mid-infrared dust emission at 24 μm with the 1.4 GHz emission from Very Large Array (VLA; Green & Joncas 1994), along with the positions of H II regions from the WISE Catalog of Galactic H II regions (Anderson et al. 2015). The radio emission resembles the infrared structures and peaks in the vicinity of the main cluster associated with IRAS 22147+5948 and along the northern arc, suggesting a physical connection. However, according to Anderson et al. (2015) the distance to H II region G104.735+02.859 (at the arc) is 7.8 kpc and to G104.699+02.781 is 6.6 kpc. The distance of 7.8 kpc is clearly inconsistent with the distance of 5.6 kpc determined using CO 1–0 (Sect. 3) and suggests that the radio source is not related to IRAS 22147+5948. On the contrary, the distance of G104.699+02.781 roughly agrees with the CO measurements and the differences may be due to uncertainties in the adopted model of the Galaxy for the outer Galaxy (König et al. 2021). The velocity of hydrogen radio recombination lines of −57.8 km s⁻¹ corresponds to the distance of 5.17 ± 0.6 kpc according to the updated rotational curve from the Bar and Spiral Structure Legacy Survey (BeSSeL; Reid et al. 2014). The

---

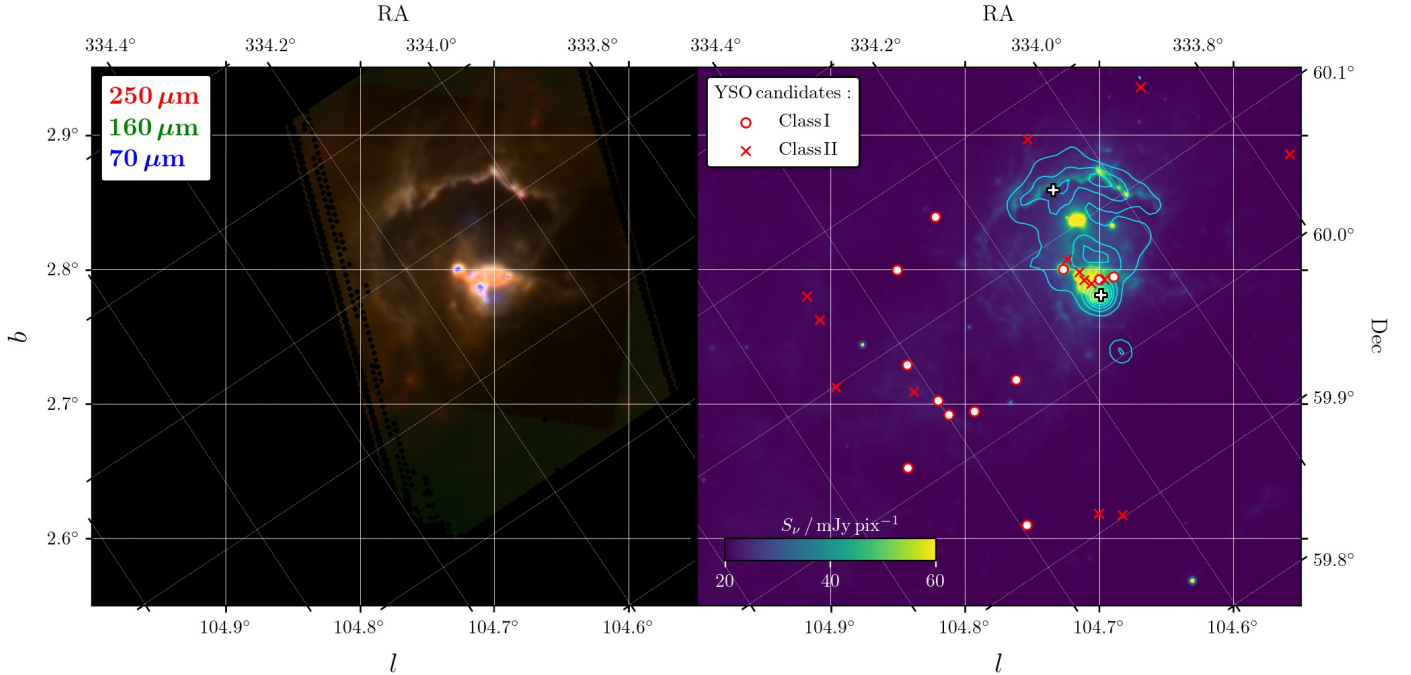[1] See a note added in proof in Green & Joncas (1994).

Fig. 11. Far-IR and radio images of the IRAS 22147 region. *Left:* three-color composite image of the IRAS 22147 region combining the *Herschel*/EPoS SPIRE 250 μm (red), PACS 160 μm (green), and PACS 70 μm (blue) mosaics. *Right: Spitzer*/MIPS 24 μm image with the VLA 1.4 GHz radio continuum contours overlaid (Green & Joncas 1994). Positions of YSO candidates identified in this work are shown in red symbols, as specified in the legend. White xs show central positions of H II regions G104.735+02.859 at 7.8 kpc (position at the arc) and G104.699+02.781 at 6.6 kpc (Anderson et al. 2015).

distance to IRAS 22147+5948 using this slightly updated rotation curve (w.r.t. the one we use in Sect. 3) is 5.33 ± 0.64 and in excellent agreement with the distance of 5.29 ± 0.65 obtained as part of the EPoS survey (Ragan et al. 2012). Thus, the H II region G104.699+02.781 is indeed located in the close vicinity of IRAS 22147+5948 and might have been a driving force triggering lower-mass star formation in the region.

Several YSOs in the IRAS 22147 region are associated with gas at a distance of 2.2 kpc, illustrating the problem with cloud confusion during the interpretation of infrared observations in the outer Galaxy. Kerton & Brunt (2003) developed a method of matching IRAS sources with proper CO clouds based on probability considerations using random lines of sight. Among all IRAS sources covered by the FCRAO survey of the outer Galaxy, they found multiple CO clouds along the line of sight toward 56.5% of IRAS sources and 78% of IRAS sources studied with pointed CO observations by Wouterloot & Brand (1989). These authors also found a single distance to IRAS 22147+5948 consistent with our study, and correctly assigned a higher probability of a similar distance to IRAS 22147+5948. Clearly, the interpretation of star forming regions in the outer Galaxy require a multi-wavelength approach and a combination of state-of-the art methods to properly identify YSOs and the properties of their host clouds and clumps.

4.2. Utility of the machine learning algorithms

Machine learning algorithms are now widely used in various astrophysical applications and clearly provide a useful tool also for the identification of YSOs and their clusters in the outer Galaxy.

In Sect. 3.3, we used the OCSVM algorithm to identify YSOs in the mid-infrared photometry from WISE and obtained

23 new YSO candidates, which accounts for 47% of detections using all adopted methods. Out of 23 objects, 18 are new, unique detections that could not be identified based on the color–color criteria used by Winston et al. (2019). We verified the location of the YSOs selected by the algorithm on the color–color diagrams and found a confirmation of these findings by comparisons with spectroscopically-confirmed YSOs in the LMC (Jones et al. 2017). The spectroscopy of all sources in the IR or submillimeter would be necessary to double-check their status, similar to the LMC efforts, but given the vastness of the outer Galaxy the OCSVM algorithm offers a compelling alternative.

The choice of the machine learning algorithm (and the adopted parameters) influences the final results and shall be well-matched with the specifics of the problem to be solved. In Sect. 3.5, we used the agglomerative hierarchical clustering algorithm to identify YSO clusters in IRAS 22147+5948 (Everitt et al. 2011), whereas Winston et al. (2019) selected the density-based spatial clustering and application with noise method (DBSCAN, Ester et al. 1996), which is a density-based clustering algorithm that can be used to identify any cluster shape in the data set containing noise and outliers. It identifies dense regions, which can be measured by the number of objects close to a given point. When applying the DBSCAN algorithm to our YSO sets within the IRAS 22147 region, we are able to recover only a single cluster located at 5.6 kpc. Clearly, DBSCAN works better with large data-sets and it does not assign every point to a cluster. As opposed to the agglomerative hierarchical clustering algorithm, DBSCAN will not partition the data but, rather, will extract the "dense" clusters and treat sparse background as noise. For this reason, in the case of a relatively small IRAS 22147 region, hierarchical clustering is a more reliable and effective approach.

Various machine learning algorithms have been tested by other authors to identify YSOs on a statistical basis using large-sky surveys. For example, the SVM algorithm used in our analysis has been adopted to an all-sky selection of YSOs from the AllWISE catalog leading to the identification of more than 90% YSOs from older catalogs (Marton et al. 2016). Additional algorithms such as random forests and neural networks were used to refine earlier identifications and benefit from distances from the *Gaia* DR2 catalog (Marton et al. 2019). As a result, 1 129 295 YSO candidates have been identified in the entire sky, including 889 in the IRAS 22147 region alone (assuming probability above 0.6 and the matching radius of 2″). Thus, the results do not seem to be sufficiently selective, which is further evidenced by a significant number of sources at high latitudes.

A recent study of Herbig Ae/Be and classical Be stars with algorithms using neural networks provides more realistic results, likely due to narrowing down the source characteristics and more extended source information (Vioque et al. 2020). Certainly, less evolved, Class 0 or I sources form the largest challenge for such studies given the relatively poor angular resolution of mid- and far-IR data. Thus, a combined analysis of machine learning algorithms with standard methods involving color–color diagrams, dust and gas emission, and CO distances are the most optimal alternatives to spectroscopic surveys.

5. Conclusions

The SMOG survey and complementary far-IR and CO data allow us to identify and characterize a population of YSO candidates in the IRAS 22147 Outer Galaxy region. A combination of standard color–color diagram method and the machine learning techniques, as well as information on gas and dust spatial distribution and SED modeling are necessary to avoid cloud confusion and properly determine the census of YSOs in specific star forming regions. The main conclusions of our study are as follows.

1. The IRAS 22147 region is associated with two distinct star forming regions located at ∼2.2 and 5.6 kpc. The compact, central embedded cluster is the more distant region physically connected with the northern arc of material and a south-west cluster associated with IRAS 22159+5948. The region at 2.2 kpc shows a more dispersed population of YSOs.

2. Using standard methods based on photometry and color–color cuts, we identify 13 Class I and 13 Class II sources, including 6 Class I and 7 Class II associated with the embedded cluster at 5.6 kpc associated with IRAS 22147+5948. A smaller cluster at 5.6 kpc, which is co-spatial with IRAS 22159+5948, contains 3 Class I and 1 Class II YSO candidates, suggesting a later launch of star formation (w.r.t. the IRAS 22147 region).

3. Two Class I sources and 21 Class II sources are identified using the SVM learning algorithm which uses mid-infrared AllWISE catalog. New boundaries in the color–color diagrams based on AllWISE filters are needed to identify YSOs in future studies.

4. Five clusters containing 16 YSO candidates are identified in the IRAS 22147 region. Two clusters at 5.6 kpc are confirmed by the spatial extent of CO 1–0 and dust emission. Three clusters at 2.2 kpc are also co-spatial with available CO 1–0, but form a loosely connected population which might as well be regarded as a single, large cluster. The agglomerative hierarchical clustering algorithm used in the analysis successfully resolved clusters at 5.6 kpc, unlike the commonly used DBSCAN, which seems to be better suited to large datasets.

5. Physical parameters for 26 YSOs were obtained from the YSO SED model fitting (Robitaille 2017). The modeling confirms their status as YSOs and provides complementary measure to the classification obtained with color–color and machine-learning methods. However, the classification is strongly affected by the lack of high angular-resolution far-IR photometry.

Multi-wavelength datasets that include infrared wavelengths are necessary for the proper identification and classification of YSOs, in particular those in the deeply-embedded stages. High angular-resolution submillimeter spectroscopy of Class 0 and I sources are needed to investigate outflow activity and molecular abundances in the outer Galaxy. The latter is a necessary step toward the determination of the gas temperatures, densities, and UV radiation field strengths, as well as improving our understanding of the impact of metallicity on star formation.

References

Ali, B., Tobin, J. J., Fischer, W. J., et al. 2010, A&A, 518, L119
Anderson, L. D., Armentrout, W. P., Johnstone, B. M., et al. 2015, ApJS, 221, 26
Andre, P., Ward-Thompson, D., & Barsony, M. 1993, ApJ, 406, 122
Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Demleitner, M., & Andrae, R. 2021, AJ, 161, 147
Bania, T. M., Anderson, L. D., Balser, D. S., & Rood, R. T. 2010, ApJ, 718, L106
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Brunthaler, A., Reid, M. J., Menten, K. M., et al. 2011, Astron. Nachr., 332, 461
Carey, S., Ali, B., Berriman, B., et al. 2008, Spitzer Mapping of the Outer Galaxy (SMOG), Spitzer Proposal, 50398
Chen, Y., Girardi, L., Bressan, A., et al. 2014, MNRAS, 444, 2525
Chen, Y., Bressan, A., Girardi, L., et al. 2015, MNRAS, 452, 1068
Digel, S., Bally, J., & Thaddeus, P. 1990, ApJ, 357, L29
Dunham, M. M., Evans, Neal J., I, Bourke, T. L., et al. 2006, ApJ, 651, 945
Dunham, M. M., Crapsi, A., Evans, Neal J., I., et al. 2008, ApJS, 179, 249
Dunham, M. M., Allen, L. E., Evans, Neal J., I., et al. 2015, ApJS, 220, 11
Edris, K. A., Fuller, G. A., & Cohen, R. J. 2007, A&A, 465, 865
Elia, D., Schisano, E., Molinari, S., et al. 2010, A&A, 518, L97
Elia, D., Molinari, S., Fukui, Y., et al. 2013, ApJ, 772, 45
Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. 1996, in Proceedings of 2nd International Conference on Knowledge Discovery, 226
Evans, Neal J., I., Dunham, M. M., Jørgensen, J. K., et al. 2009, ApJS, 181, 321
Everitt, B., Landau, S., Leese, M., & Stahl, D. 2011, Cluster analysis, 5th edn. (Wiley)
Fang, M., Kim, J. S., Pascucci, I., & Apai, D. 2021, ApJ, 908, 49
Fazio, G. G., Hora, J. L., Allen, L. E., et al. 2004, ApJS, 154, 10
Fischer, W. J., Padgett, D. L., Stapelfeldt, K. L., & Sewiło, M. 2016, ApJ, 827, 96
Fontani, F., Cesaroni, R., & Furuya, R. S. 2010, A&A, 517, A56
Gaia Collaboration (Brown, A. G. A., et al.) 2021, A&A, 649, A1
Giannetti, A., Leurini, S., König, C., et al. 2017, A&A, 606, L12
Goicoechea, J. R., Cernicharo, J., Karska, A., et al. 2012, A&A, 548, A77
Green, D. A., & Joncas, G. 1994, A&AS, 104, 481
Griffin, M. J., Abergel, A., Abreu, A., et al. 2010, A&A, 518, L3
Gutermuth, R. A., Megeath, S. T., Myers, P. C., et al. 2009, ApJS, 184, 18
Heyer, M., & Dame, T. M. 2015, ARA&A, 53, 583
Heyer, M. H., Brunt, C., Snell, R. L., et al. 1998, ApJS, 115, 241

Heyer, M., Krawczyk, C., Duval, J., & Jackson, J. M. 2009, ApJ, 699, 1092
Jarrett, T. H., Cohen, M., Masci, F., et al. 2011, ApJ, 735, 112
Jiang, B. W., Deguchi, S., & Nakada, Y. 1996, AJ, 111, 231
Joncas, G., & Higgs, L. A. 1990, A&AS, 82, 113
Jones, O. C., Woods, P. M., Kemper, F., et al. 2017, MNRAS, 470, 3250
Kalenskij, S. V., Bachiller, R., Berulis, I. I., et al. 1992, AZh, 69, 1002
Karska, A., Kaufman, M. J., Kristensen, L. E., et al. 2018, ApJS, 235, 30
Kennicutt, R. C., & Evans, N. J. 2012, ARA&A, 50, 531
Kerton, C. R., & Brunt, C. M. 2003, A&A, 399, 1083
Koenig, X. P., & Leisawitz, D. T. 2014, ApJ, 791, 131
König, C., Urquhart, J. S., Wyrowski, F., Colombo, D., & Menten, K. M. 2021, A&A, 645, A113
Kothes, R., Fedotov, K., Foster, T. J., & Uyanıker, B. 2006, A&A, 457, 1081
Kristensen, L. E., & Dunham, M. M. 2018, A&A, 618, A158
Kristensen, L. E., van Dishoeck, E. F., Bergin, E. A., et al. 2012, A&A, 542, A8
Kryukova, E., Megeath, S. T., Gutermuth, R. A., et al. 2012, AJ, 144, 31
Kumar, M. S. N., Keto, E., & Clerkin, E. 2006, A&A, 449, 1033
Lundquist, M. J., Kobulnicky, H. A., Alexander, M. J., Kerton, C. R., & Arvidsson, K. 2014, ApJ, 784, 111
Mainzer, A., Bauer, J., Grav, T., et al. 2011, ApJ, 731, 53
Marton, G., Tóth, L. V., Paladini, R., et al. 2016, MNRAS, 458, 3479
Marton, G., Ábrahám, P., Szegedi-Elek, E., et al. 2019, MNRAS, 487, 2522
Meade, M. R., Whitney, B., Babler, B., et al. 2016, "SMOG IRAC Data Description" available at: https://irsa.ipac.caltech.edu/data/SPITZER/GLIMPSE/doc/smog_dataprod_v1.1.pdf
Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. 2020, Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien, r package version 1.7-4
Molinari, S., Brand, J., Cesaroni, R., & Palla, F. 1996, A&A, 308, 573
Molinari, S., Swinyard, B., Bally, J., et al. 2010, A&A, 518, A100
Molinari, S., Schisano, E., Faustini, F., et al. 2011, A&A, 530, A133
Palla, F., Brand, J., Cesaroni, R., Comoretto, G., & Felli, M. 1991, A&A, 246, 249
Pilbratt, G. L., Riedinger, J. R., Passvogel, T., et al. 2010, A&A, 518, L1
Poglitsch, A., Waelkens, C., Geis, N., et al. 2010, A&A, 518, A2
Pohl, M., Englmaier, P., & Bissantz, N. 2008, ApJ, 677, 283
R Core Team 2018, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing
Ragan, S., Henning, T., Krause, O., et al. 2012, A&A, 547, A49
Reid, M. J., Menten, K. M., Brunthaler, A., et al. 2014, ApJ, 783, 130
Rieke, G. H., Young, E. T., Engelbracht, C. W., et al. 2004, ApJS, 154, 25
Robitaille, T. P. 2017, A&A, 600, A11
Robitaille, T. P., Whitney, B. A., Indebetouw, R., & Wood, K. 2007, ApJS, 169, 328
Roman-Duval, J., Jackson, J. M., Heyer, M., Rathborne, J., & Simon, R. 2010, ApJ, 723, 492
Rosolowsky, E., & Leroy, A. 2006, PASP, 118, 590
Schölkopf, B., Williamson, R., Smola, A., Shawe-Taylor, J., & Platt, J. 1999, in Proceedings of the 12th International Conference on Neural Information Processing Systems, NIPS'99 (Cambridge, MA, USA: MIT Press), 582
Sewiło, M., Whitney, B. A., Yung, B. H. K., et al. 2019, ApJS, 240, 26
Shawe-Taylor, J. & Cristianini, N. 2004, Kernel Methods for Pattern Analysis (New York, NY, USA: Cambridge University Press)
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
Sodroski, T. J., Odegard, N., Arendt, R. G., et al. 1997, ApJ, 480, 173
Solarz, A., Bilicki, M., Gromadzki, M., et al. 2017, A&A, 606, A39
Stern, D., Eisenhardt, P., Gorjian, V., et al. 2005, ApJ, 631, 163
Tang, J., Bressan, A., Rosenfield, P., et al. 2014, MNRAS, 445, 4287
Taylor, A. R., Gibson, S. J., Peracaula, M., et al. 2003, AJ, 125, 3145
Urquhart, J. S., Busfield, A. L., Hoare, M. G., et al. 2008, A&A, 487, 253
Vapnik, V. N. 1995, The Nature of Statistical Learning Theory (New York, NY, USA: Springer-Verlag, Inc.)
Vioque, M., Oudmaijer, R. D., Schreiner, M., et al. 2020, A&A, 638, A21
Weingartner, J. C., & Draine, B. T. 2001, ApJ, 548, 296
Werner, M. W., Roellig, T. L., Low, F. J., et al. 2004, ApJS, 154, 1
Wilking, B. A., Mundy, L. G., Blackwell, J. H., & Howe, J. E. 1989, ApJ, 345, 257
Williams, J. P., de Geus, E. J., & Blitz, L. 1994, ApJ, 428, 693
Winston, E., Hora, J., Gutermuth, R., & Tolls, V. 2019, ApJ, 880, 9
Wouterloot, J. G. A., & Brand, J. 1989, A&AS, 80, 149
Wouterloot, J. G. A., Brand, J., & Fiegle, K. 1993, A&AS, 98, 589
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868

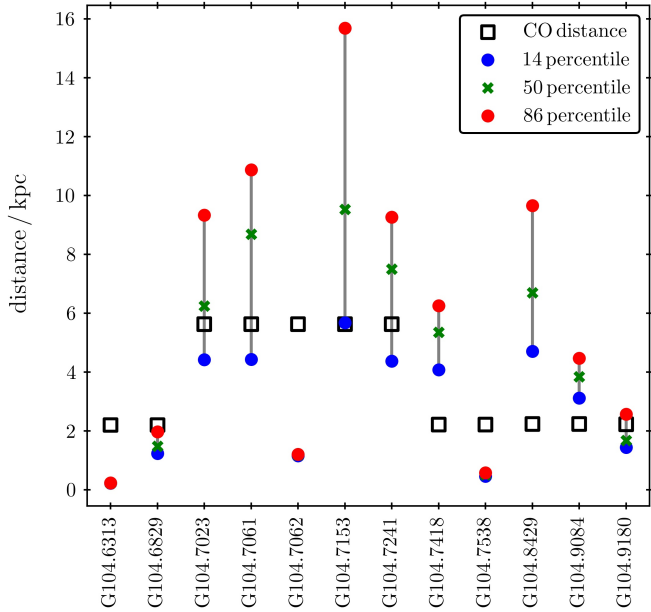# Appendix A: Kinematic distances versus the Gaia measurements

Fig. A.1. Comparison of the kinematic distances of YSOs in the IRAS 22147 region obtained using CO observations and the distances from Gaia (Bailer-Jones et al. 2021). Black empty squares show the kinematic distance obtained using CO emission (Section 3.1). Blue dots, green crosses, and red dots show the 14th, 50th (median), and the 86th percentile of photogeometric distance distribution. The full names of YSOs and their equatorial coordinates are provided in Table A.1.

We compared kinematic distances obtained using CO (Section 3.1) with the Gaia Early Data Release 3 catalog (Gaia Collaboration et al. 2021). We cross-correlated 49 YSOs in the IRAS 22147 region with the catalog by Bailer-Jones et al. (2021), which provides photogeometric distances for 1.35 billion Gaia targets. The distances are determined using a probabilistic approach involving information about the colors and apparent magnitudes of stars.

Twelve Gaia targets are found within ~1 arc sec separation from YSOs in the IRAS 22147 region (~25% of all YSOs). The remaining YSOs are located more than ~2.5 arcsec away from the Gaia targets, and are unlikely to be their true counterparts. Table A.1 shows the comparison of the distances obtained toward the matched sources using the CO emission (Section 3.1) and the photogeometric distances from Gaia (Bailer-Jones et al. 2021). Figure A.1 shows the same information in a graphical way, illustrating the uncertainties of the Gaia measurements. The uncertainties of the kinematic distances from CO depend on the accuracy of the rotational curves and proper motions of the sources, and may be probably as high as 50% of the quoted values.

Among the YSOs located at ~5.6 kpc, the kinematic distances of four out of five objects are consistent with the Gaia distances within uncertainties; i is G104.7062 tat appearsto be located at ~2.2 kpc. The discrepancies are more severe for objects with kinematic distances of ~2.2 kpc, with only two out of seven objects showing values consistent with Gaia. In case of G104.7418 and G104.8429, the Gaia distances suggest the position of the objects in the more distant cloud (at ~5.6 kpc).

On the contrary, G104.6313 and G104.7538 are located at even more nearby star-forming regions.

The discrepancies between kinematic distances and Gaia measurements point to the problem with the automatic selection of the best CO clump in case of CO lines at multiple velocities toward the same sources. In case of G104.7062, G104.6313, and G104.7538, the distances from Gaia are very precise and likely more reliable than the kinematic ones, due to uncertainties in using the rotational curve for the most nearby sources. Thus, we adopted the median distances from Gaia for the subsequent analysis of those objects.

Seven YSOs with Gaia distances passed our criteria for the SED modelling (Table 3). Three sources show sufficient agreement with the kinematic distances, which are adopted for the modeling (G104.6829, G104.7153, G104.9180). For the remaining four sources, we used Gaia distances to obtain new YSO SED models (G104.7062. G104.7538, G104.8429, and G104.9084).

Table A.2 shows the physical parameters of YSOs in the IRAS 22147 region obtained from the best-fit Robitaille (2017) SED models and the Gaia distances from Bailer-Jones et al. 2021 (Table A.1). An decrease of almost two orders of magnitude in source luminosity is seen for G104.7062 (from 1750 to 32 $L_\odot$) and G104.7538 (from 3 to 0.04 $L_\odot$) due to a factor of 4-5 smaller distances to these sources found by Gaia, as well as a slightly different sets of YSO models fitted (Table 2). Similarly, the luminosity of G104.8429 increases from 0.1 to 3.7 $L_\odot$ due to a larger distance to the source from Gaia by factor of 3. Finally, a similar source luminosity of G104.9084 (20.7 $L_\odot$) is obtained using the Gaia distance of 3.8 kpc and the kinematic distance of 2.2 kpc (27 $L_\odot$). The classification of the sources is not altered except G104.7062, ages of which are consistent with Class 0/I protostars. Figure A.2 shows the SEDs of YSOs and the best-fit Robitaille (2017) models (see also Appendix F).

Figure A.3 shows the impact of adopting the Gaia distances (Table A.1) on the clustering analysis. Among six sources with the significant differences between Gaia and CO distances, only three sources are cluster members when CO kinematic distances are assumed (G104.7062, G104.7418, and G104.9084; see also Section 3.5). Adopting the distance of 1.2 kpc for G104.7062 would result in the removal of the source from the cluster at 5.6 kpc; the ratio of Class II to Class I objects would decrease from 2.7 to 2.3, but the cluster center remains almost the same at l,b=(104.100°, 2.797°). Similarly, adopting a distance of 5.4 kpc for G104.7418 would remove it from the originally five-member cluster at 2.2 kpc; the ratio of Class II to Class I objects would decrease from 4.0 to 3.0 and the cluster center moves to (104.722°, 2.608°). Finally, G104.9084 is a member of a small, two-member cluster at 2.2 kpc, and adopting the new Gaia distance of 3.8 kpc would result in the cancellation of the cluster identification.

We note that the Gaia distances and their impact on the clustering does not seem to be consistent with the distribution of molecular gas in the IRAS 22147 region (Fig. 7). G104.7062 is located in the middle of the cluster at 5.6 kpc, where significant amount of molecular gas is present. The CO spectra at those positions do not contain any additional CO lines at velocities corresponding to cloud at smaller distances. A similar situation occurs at the southern cluster with G104.7418, which is located in the central part of the 2.2 kpc cluster, with no CO emission corresponding to larger distances. We conclude that the Gaia distances may not provide the improvement over the kinematic distances in case of star-forming regions.

Table A.1. Distances to YSOs based on the CO emission and the Gaia catalog (Bailer-Jones et al. 2021)

Nr	IRAC Designation^a 'SSTSMOGA'	RA (deg)	DEC (deg)	Separation^b (arc sec)	D_{kin}^{c} (pc)	D_{Gaia}^{d} (pc) 50th	14th	86th
1	**G104.6313+02.6211**	334.176540	59.876274	0.60	2197	224	218	231
2	G104.6829+02.6173	334.266213	59.901773	0.12	2197	1481	1227	1968
3	G104.7023+02.8142	334.079104	60.076155	0.11	5628	6245	4413	9334
4	G104.7061+02.7861	334.116800	60.054850	0.35	5628	8687	4425	10873
5	**G104.7062+02.7895**	334.113200	60.057750	0.15	5628	1179	1157	1200
6	G104.7153+02.7979	334.118870	60.069817	0.11	5628	9531	5675	15687
7	G104.7241+02.8075	334.122770	60.082638	0.09	5628	7498	4365	9265
8	**G104.7418+02.5949**	334.388550	59.915745	0.07	2215	5354	4070	6247
9	**G104.7538+02.8964**	334.072570	60.172980	0.17	2208	506	454	577
10	**G104.8429+02.6522**	334.493160	60.019398	1.05	2230	6691	4698	9651
11	**G104.9084+02.7625**	334.480220	60.147484	0.02	2230	3839	3109	4469
12	G104.9180+02.7798	334.477020	60.167213	0.18	2223	1673	1440	2567

Notes. $^{(a)}$Sources with the differences in the distances calculated using CO and Gaia exceeding the Gaia uncertainties are in bold. $^{(b)}$A separation between the YSO and the closest Gaia target from the Bailer-Jones et al. (2021) catalog. $^{(c)}$A kinematic distance from the CO emission (Section 3.1). $^{(d)}$The three quantiles of a distance distribution from Bailer-Jones et al. (2021): the median (the 50th percentile), the 14th, and 86th percentiles.
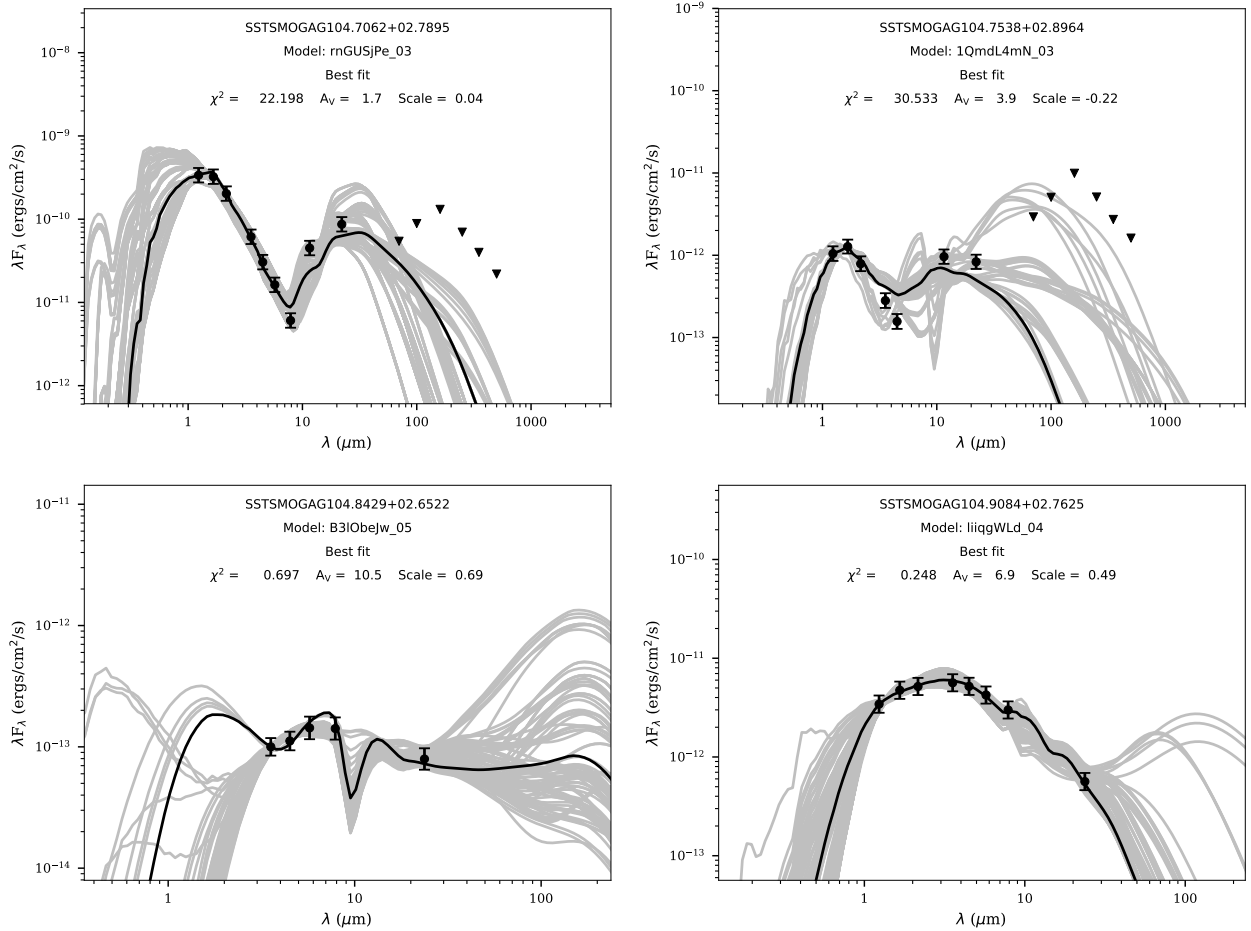
Fig. A.2. SEDs of candidate YSOs with distances adopted from the Gaia DR3 catalog (see Table A.2), see also Fig. 9.

Appendix B: Photometry of YSO candidates in the IRAS 22147 region

Tables C.1 and C.2 show the multi-wavelength photometry for 35 YSOs in the IRAS 22147 region identified using the color–color diagrams and the SVM learning algorithm, for

which the distances have been assigned (see Section 3). The table includes nine sources without 12 μm photometry for which the SED modeling was not performed.

Table C.3 shows the photometry of the remaining 14 YSOs identified in the IRAS22147 region, for which the distances are

Table A.2. Physical Parameters for a Subset of YSO Candidates in the IRAS 22147 region with distances from the Gaia DR3 catalog

Nr	IRAC Designation 'SSTSMOGA'	Model Set[a]	Dist.[b] (kpc)	$R_\star$ ($R_\odot$)	$T_\star$ ($T_\odot$)	$L_\star$ ($L_\odot$)	$M_\star$ ($M_\odot$)	Age (Myr)
1	G104.7062+02.7895	sp–h-i	1.2	9.3	4500	31.8	1.3	0.3
2	G104.7538+02.8964	sp–h-i	0.5	0.5	3440	0.04	–	–
3	G104.8429+02.6522	sp–smi	6.7	1.7	6220	3.7	–	–
4	G104.9084+02.7625	sp–s-i	3.8	2.9	7210	20.7	–	–

[a]See Tables 2 and 3 for the meaning of characters in the model set names, and the remarks to the physical quantities, respectively.
[b]Distances taken from Bailer-Jones et al. (2021).
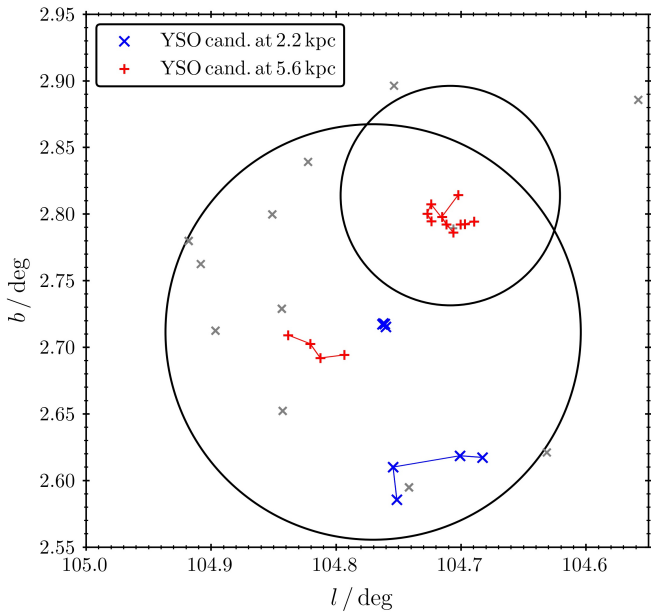
Fig. A.3. Minimum spanning tree for YSO candidates in the IRAS 22147 region adopting Gaia distances for G104.7062, G104.6313, and G104.7538 (shown as gray x's), see also Fig. 6.

not available. The SEDs of these sources could not have been modeled with the Robitaille (2017) YSO models (Section 3.7).

Appendix C: Comparison of the YSO classification with the literature results

Table C.4 shows a source-by-source comparison of the YSO classification for 24 sources that were identified here and in Winston et al. (2019) (see Section 3.4). For completeness, we include also the classification based on color–color diagrams, where we did not require the distance determination (column 'CC (no dist.)'). This method is not optimal, as we could not properly isolate AGN candidates, but allows us to double-check the classification obtained using the machine learning techniques (column 'ML').

A comparison of the columns labeled "ML" and " MIPS" shows a very good agreement between our work and Winston et al. (2019), with 14 Class I and 10 Class II sources identified here, and 11 Class I and 13 Class II sources cataloged in Winston et al. (2019). Six sources have different classification: four are classified as Class I here and as Class II in Winston et al. (2019), and two are classified here as Class II (ML) and as Class I in Winston et al. (2019). One additional source,

G104.8383+02.7091, is classified as Class II in our catalog and in Winston et al. (2019) using IRAC and MIPS bands, but appears as Class I when only IRAC bands are used in both works; we consider it as a Class II source in this paper. Overall, 75% of the YSO candidates in both works share the same classification.

Appendix D: Bolometric luminosities of YSOs

We used the photometry of YSOs in the IRAS 22147 region to calculate bolometric luminositeis, L_{bol}, using trapezoidal summation (Table C.1 and C.2). For sources with available *Spitzer*/MIPS photometry at $24\,\mu m$, we omit the WISE 12 and $22\,\mu m$ points due to the larger beam and possible source confusion. In the calculations, we used upper limits to constrain the SED shape in the far-IR range (see Appendix F) and, thus, the calculated luminosities shall be considered as upper limits of the actual L_{bol}.

Figure D.1 shows the distribution of the resulting L_{bol} of YSOs in the IRAS22147 region. The objects are divided by their kinematic distances into two groups to illustrate the range of bolometric luminosities, which can be probed using the available infrared photometry. YSOs located at ~5.6 kpc are relatively luminous objects with L_{bol} in the range of ~$10 - 1000$ L$_\odot$, corresponding to low- and intermediate-mass YSOs. Among more nearby sources, less luminous sources are also detected in the infrared, with L_{bol} in the range of ~$0.1 - 10$ L$_\odot$. Note that the single source with L_{bol} below 1 L$_\odot$ at 5.6 kpc is G104.8206; unfortunately, its distance cannot be verified with Gaia (Appendix A).
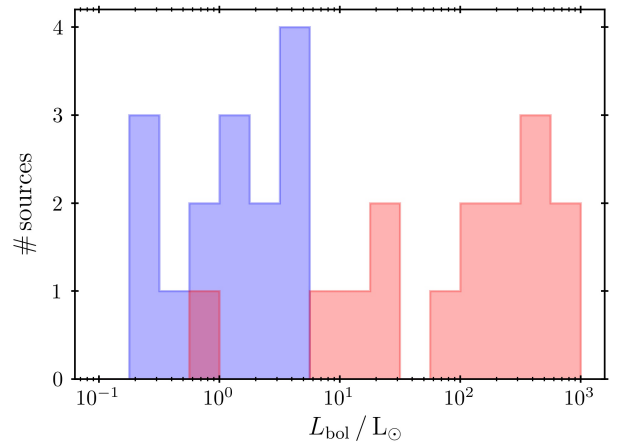
Fig. D.1. Histogram of bolometric luminosities for YSOs located at ~2.2 kpc (in blue) and at ~5.6 kpc (in red), adopting kinematic distances.

Table C.1. Near- and mid-infrared photometry of YSO candidates in the IRAS 22147+5948 field. The columns represent the 2MASS JHK_s, *Spitzer* IRAC, and AllWISE 12 μm bands. Units are mJy and the upper limits are 3σ.

Nr	IRAC Designation SSTSMOGA	2MASS J	2MASS H	2MASS K_s	IRAC 3.6 μm	IRAC 4.5 μm	IRAC 5.8 μm	IRAC 8.0 μm	WISE 12 μm
1	G104.5582+02.8857	–	–	–	1.29 ± 0.26	1.51 ± 0.30	1.61 ± 0.32	1.69 ± 0.34	–
2	G104.6313+02.6211	1.04 ± 0.21	0.93 ± 0.19	0.90 ± 0.18	0.50 ± 0.10	0.36 ± 0.07	0.26 ± 0.05	–	–
3	G104.6697+02.9357	–	0.37 ± 0.07	0.61 ± 0.12	–	1.52 ± 0.30	–	2.21 ± 0.44	1.62 ± 0.32
4	G104.6829+02.6173	0.80 ± 0.16	1.59 ± 0.32	1.80 ± 0.36	1.60 ± 0.32	1.50 ± 0.30	1.32 ± 0.26	1.65 ± 0.33	–
5	G104.6895+02.7945	–	1.26 ± 0.25	8.68 ± 1.74	63.09 ± 12.62	128.00 ± 25.60	230.30 ± 46.06	300.60 ± 60.12	484.29 ± 96.86
6	G104.6968+02.7925	–	–	–	0.79 ± 0.16	1.59 ± 0.32	2.63 ± 0.53	5.60 ± 1.12	57.43 ± 11.49
7	G104.7003+02.7923	–	–	–	2.01 ± 0.40	3.73 ± 0.75	5.33 ± 1.07	0.25 ± 0.05	–
8	G104.7009+02.6186	–	–	–	0.14 ± 0.03	0.17 ± 0.03	0.19 ± 0.04	2.19 ± 0.44	–
9	G104.7023+02.8142	1.46 ± 0.29	2.00 ± 0.40	2.11 ± 0.42	2.38 ± 0.48	2.62 ± 0.52	2.36 ± 0.47		< 0.46
10	G104.7061+02.7861	1.31 ± 0.26	2.47 ± 0.49	2.90 ± 0.58	2.56 ± 0.51	2.35 ± 0.47	–		–
11	G104.7062+02.7895	141.80 ± 28.36	183.10 ± 36.62	149.00 ± 29.80	74.19 ± 14.84	46.57 ± 9.31	31.70 ± 6.34	16.24 ± 3.25	176.97 ± 35.39
12	G104.7118+02.7921	1.28 ± 0.26	14.91 ± 2.98	42.53 ± 8.51	56.40 ± 11.28	39.40 ± 7.88	–	18.44 ± 3.69	202.63 ± 40.53
13	G104.7153+02.7979	1.71 ± 0.34	2.56 ± 0.51	3.10 ± 0.62	1.94 ± 0.39	1.42 ± 0.28	–		87.72 ± 17.54
14	G104.7237+02.7946	–	0.68 ± 0.14	1.15 ± 0.23	1.41 ± 0.28	1.67 ± 0.33	1.57 ± 0.32	1.42 ± 0.28	–
15	G104.7241+02.8075	2.36 ± 0.47	2.92 ± 0.58	–	2.27 ± 0.45	1.88 ± 0.38	1.29 ± 0.26		17.01 ± 3.40
16	G104.7270+02.8003	1.51 ± 0.30	5.59 ± 1.12	18.63 ± 3.73	89.42 ± 17.88	142.80 ± 28.56	196.20 ± 39.24	348.60 ± 69.72	900.95 ± 180.19
17	G104.7418+02.5949	3.96 ± 0.79	6.47 ± 1.29	5.78 ± 1.16	3.12 ± 0.62	3.27 ± 0.66	1.30 ± 0.26		< 0.24
18	G104.7513+02.5857	–	–	–	0.23 ± 0.05	0.22 ± 0.04	0.16 ± 0.03	0.20 ± 0.04	< 0.34
19	G104.7538+02.8964	0.44 ± 0.09	0.72 ± 0.14	0.58 ± 0.12	0.34 ± 0.07	0.24 ± 0.05	–		3.78 ± 0.76
20	G104.7544+02.6101	–	–	–	2.10 ± 0.42	3.19 ± 0.64	4.09 ± 0.82	4.12 ± 0.83	4.00 ± 0.80
21	G104.7601+02.7154	–	–	–	0.42 ± 0.08	0.49 ± 0.10	0.62 ± 0.12	0.79 ± 0.16	–
22	G104.7611+02.7176	–	–	–	0.37 ± 0.07	0.45 ± 0.09	0.45 ± 0.09	0.47 ± 0.09	–
23	G104.7622+02.7179	–	–	–	0.54 ± 0.11	0.74 ± 0.15	0.93 ± 0.19	0.88 ± 0.18	3.29 ± 0.66
24	G104.7631+02.7173	–	–	–	0.49 ± 0.10	0.78 ± 0.16	1.00 ± 0.20	1.01 ± 0.20	–
25	G104.7934+02.6942	–	0.51 ± 0.10	0.82 ± 0.16	1.39 ± 0.28	1.96 ± 0.39	2.54 ± 0.51	3.25 ± 0.65	3.37 ± 0.67
26	G104.8126+02.6920	–	–	–	2.74 ± 0.55	6.24 ± 1.25	11.99 ± 2.40	18.67 ± 3.73	22.97 ± 4.59
27	G104.8206+02.7026	–	–	–	0.27 ± 0.05	0.44 ± 0.09	0.63 ± 0.13	0.97 ± 0.19	–
28	G104.8225+02.8393	–	–	–	0.66 ± 0.13	1.10 ± 0.22	1.79 ± 0.36	2.43 ± 0.49	3.53 ± 0.71
29	G104.8383+02.7091	–	0.37 ± 0.07	1.71 ± 0.34	4.13 ± 0.83	5.74 ± 1.15	8.73 ± 1.75	10.28 ± 2.06	11.31 ± 2.26
30	G104.8429+02.6522	–	–	–	0.12 ± 0.02	0.17 ± 0.03	0.28 ± 0.06	0.38 ± 0.08	–
31	G104.8436+02.7289	–	–	–	0.11 ± 0.02	0.17 ± 0.03	0.25 ± 0.05	0.32 ± 0.06	–
32	G104.8510+02.7998	–	–	–	2.21 ± 0.44	2.77 ± 0.55	3.52 ± 0.70	4.00 ± 0.80	4.58 ± 0.92
33	G104.8965+02.7124	–	–	–	0.54 ± 0.11	0.50 ± 0.10	0.39 ± 0.08	0.30 ± 0.06	2.98 ± 0.60
34	G104.9084+02.7625	1.44 ± 0.29	2.68 ± 0.54	3.80 ± 0.76	6.81 ± 1.36	7.93 ± 1.59	8.26 ± 1.65	8.01 ± 1.60	6.62 ± 1.32
35	G104.9180+02.7798	0.27 ± 0.05	0.59 ± 0.12	–	0.19 ± 0.04	0.12 ± 0.03	–		1.03 ± 0.21

Table C.2. Photometry of YSO candidates in the IRAS 22147+5948 field from mid-infrared to submillimeter wavelengths. The columns represent the AllWISE 22 μm, *Spitzer* MIPS 24 μm and *Herschel* PACS and SPIRE bands. Units are mJy and the upper limits are 3σ.

Nr	IRAC Designation SSTSMOGA	WISE 22 μm	MIPS 24 μm	PACS 70 μm	PACS 100 μm	PACS 160 μm	SPIRE 250 μm	SPIRE 350 μm	SPIRE 500 μm
1	G104.5582+02.8857	–	3.46 ± 0.69	–	–	–	–	–	–
2	G104.6313+02.6211	–	–	–	–	–	< 26.25	< 19.13	< 20.39
3	G104.6697+02.9357	4.52 ± 1.20	5.51 ± 1.10	–	–	–	< 543.11	< 439.78	< 354.30
4	G104.6829+02.6173	–	3.07 ± 0.61	–	–	–	< 14.11	< 14.21	< 14.89
5	G104.6895+02.7945	1441.33 ± 288.27	951.30 ± 190.26	3812.09 ± 762.42	4602.46 ± 920.49	6556.28 ± 1311.26	4978.77 ± 995.75	2566.71 ± 513.34	2656.18 ± 531.24
6	G104.6968+02.7925	172.81 ± 34.56	–	< 638.26	< 1629.24	< 4786.60	< 4657.96	< 4425.38	< 3035.44
7	G104.7003+02.7923	–	–	970.26 ± 194.05	2180.21 ± 436.04	5872.59 ± 1174.52	5667.15 ± 1133.43	4845.84 ± 969.17	3304.47 ± 660.89
8	G104.7009+02.6186	–	0.76 ± 0.15	–	–	–	< 17.52	< 11.93	< 16.96
9	G104.7023+02.8142	< 2.34	–	< 137.91	< 264.69	< 698.90	< 546.92	< 416.00	< 383.71
10	G104.7061+02.7861	–	–	< 1205.16	< 2483.55	< 5822.72	< 5420.34	< 3622.00	< 3088.32
11	G104.7062+02.7895	652.17 ± 130.43	–	< 1266.84	< 2953.93	< 6997.07	< 5817.88	< 4658.27	< 3651.07
12	G104.7118+02.7921	837.84 ± 167.57	–	< 840.59	< 2140.81	< 5910.42	< 4916.79	< 4984.93	< 3561.18
13	G104.7153+02.7979	331.71 ± 66.34	–	< 897.40	< 1881.94	< 3962.98	< 2977.20	< 2390.63	< 1994.98
14	G104.7237+02.7946	–	–	< 298.37	< 835.45	< 2920.31	< 4428.71	< 3691.58	< 2952.51
15	G104.7241+02.8075	124.73 ± 24.95	–	< 215.06	< 562.18	< 1561.44	< 1580.04	< 1315.64	< 1074.96
16	G104.7270+02.8003	2400.83 ± 480.17	–	5572.69 ± 1114.54	7435.41 ± 1487.08	10568.03 ± 2113.61	5931.67 ± 1186.33	3791.80 ± 758.36	2551.67 ± 510.33
17	G104.7418+02.5949	< 1.45	–	–	–	–	–	–	–
18	G104.7513+02.5857	< 1.74	–	–	–	–	–	–	–
19	G104.7538+02.8964	6.25 ± 1.25	–	< 68.21	< 168.90	< 529.21	< 425.21	< 317.81	< 269.32
20	G104.7544+02.6101	16.78 ± 3.36	14.68 ± 2.94	–	–	–	–	< 13.41	< 17.90
21	G104.7601+02.7154	–	–	< 38.26	< 122.08	< 507.21	< 800.02	< 666.55	< 579.15
22	G104.7611+02.7176	–	–	< 37.69	< 110.58	< 706.18	< 872.82	< 940.50	< 634.56
23	G104.7622+02.7179	21.98 ± 4.40	–	< 70.86	< 194.06	< 781.75	< 1034.92	< 1113.86	< 634.56
24	G104.7631+02.7173	–	–	< 60.01	< 198.57	< 771.71	< 1034.92	< 1113.86	< 634.56
25	G104.7934+02.6942	13.98 ± 2.80	13.45 ± 2.69	–	–	–	< 351.64	< 540.90	< 398.58
26	G104.8126+02.6920	54.85 ± 10.97	36.49 ± 7.30	–	–	–	–	–	–
27	G104.8206+02.7026	–	2.85 ± 0.57	–	–	–	–	–	–
28	G104.8225+02.8393	9.22 ± 1.84	6.81 ± 1.36	–	–	–	< 340.48	< 334.73	–
29	G104.8383+02.7091	27.90 ± 5.58	11.41 ± 2.28	–	–	–	–	–	–
30	G104.8429+02.6522	–	0.64 ± 0.13	–	–	–	–	–	–
31	G104.8436+02.7289	–	1.10 ± 0.22	–	–	–	–	–	–
32	G104.8510+02.7998	20.34 ± 4.07	12.58 ± 2.52	–	–	–	–	–	–
33	G104.8965+02.7124	6.85 ± 1.37	0.65 ± 0.13	–	–	–	–	–	–
34	G104.9084+02.7625	6.20 ± 1.24	4.54 ± 0.91	–	–	–	–	–	–
35	G104.9180+02.7798	4.27 ± 0.86	–	–	–	–	–	–	–

Table C.3. Multi-wavelength photometry of YSO candidates in the IRAS 22147+5948 field with unknown distances. The columns represent the 2MASS JHK_s, *Spitzer* IRAC, AllWISE 12 and 22 μm and *Spitzer* MIPS 24 μm. Units are in mJy.

Nr	IRAC Designation SSTSMOGA	2MASS			IRAC				WISE		MIPS
		J	H	K_s	$3.6\,\mu m$	$4.5\,\mu m$	$5.8\,\mu m$	$8.0\,\mu m$	$12\,\mu m$	$22\,\mu m$	$24\,\mu m$
1	G104.5742+02.6247	0.36 ± 0.06	0.54 ± 0.07	0.82 ± 0.09	0.77 ± 0.03	0.80 ± 0.03	0.75 ± 0.05	0.72 ± 0.04	2.13 ± 0.15	5.92 ± 0.84	0.90 ± 0.26
2	G104.5824+02.6207	–	–	–	0.46 ± 0.03	0.77 ± 0.03	1.07 ± 0.05	1.52 ± 0.07	3.96 ± 0.15	15.69 ± 1.14	3.83 ± 0.37
3	G104.5939+02.8413	0.57 ± 0.05	0.55 ± 0.07	0.53 ± 0.09	0.24 ± 0.03	0.16 ± 0.01	–	–	1.50 ± 0.10	4.56 ± 0.86	–
4	G104.5956+02.8382	–	–	–	0.14 ± 0.03	0.11 ± 0.01	–	–	1.55 ± 0.12	4.38 ± 0.80	–
5	G104.5984+02.7886	–	–	–	0.10 ± 0.01	0.09 ± 0.01	–	–	0.95 ± 0.13	3.76 ± 0.81	–
6	G104.6468+02.8399	–	–	–	0.28 ± 0.01	0.20 ± 0.01	–	–	2.50 ± 0.21	9.81 ± 1.13	–
7	G104.6796+02.8554	–	–	–	10.47 ± 0.46	31.65 ± 0.79	63.56 ± 1.32	78.03 ± 1.71	285.06 ± 3.94	1054.61 ± 22.34	–
8	G104.6993+02.7714	3.93 ± 0.09	6.65 ± 0.07	6.19 ± 0.09	3.57 ± 0.10	2.29 ± 0.06	1.63 ± 0.09	–	22.66 ± 0.48	237.19 ± 7.43	–
9	G104.7034+02.7674	3.19 ± 0.10	4.06 ± 0.07	3.15 ± 0.09	1.62 ± 0.05	1.04 ± 0.03	0.65 ± 0.08	–	25.76 ± 0.44	92.02 ± 4.32	–
10	G104.7076+02.7809	1.50 ± 0.05	1.80 ± 0.09	1.41 ± 0.09	0.73 ± 0.04	0.47 ± 0.03	–	–	87.69 ± 1.94	658.70 ± 16.99	–
11	G104.7301+02.7741	–	–	–	0.23 ± 0.03	0.32 ± 0.02	–	–	11.40 ± 0.38	42.77 ± 2.21	–
12	G104.8880+02.6685	–	–	–	0.33 ± 0.01	0.33 ± 0.01	0.38 ± 0.03	1.78 ± 0.06	1.54 ± 0.12	8.22 ± 0.95	8.31 ± 0.23
13	G104.9019+02.7100	–	–	–	0.33 ± 0.02	0.43 ± 0.01	0.61 ± 0.04	1.15 ± 0.05	1.97 ± 0.24	5.51 ± 0.83	4.22 ± 0.45
14	G104.9468+02.8214	–	–	–	0.19 ± 0.01	0.12 ± 0.01	–	–	0.56 ± 0.09	2.28 ± 0.67	–

Table C.4. YSO classification obtained in this work and in Winston et al. (2019)

Nr	IRAC Designation[a] 'SSTSMOGA'	Class (this work)[b]			Class (Winston et al. 2019)[c]	
		ML	CC	CC (no dist.)	IRAC	MIPS
1	G104.5582+02.8857	II[(t)d]	II	II	II	II
2	G104.5742+02.6247	II	–	II	II	II
3	G104.5824+02.6207	I	–	I	I	I
4	G104.6796+02.8554	I	–	I	I	I
5	G104.6829+02.6173	II[(t)]	II	II	II	II
6	G104.6895+02.7945	I[(t)]	I	I	I	I
7	**G104.6968+02.7925**	II	–	–	I	I
8	G104.7003+02.7923	I[(t)]	I	I	I	I
9	G104.7023+02.8142	II[(t)]	II	II	II	II
10	G104.7237+02.7946	II[(t)]	II	II	II	II
11	G104.7270+02.8003	I[(t)]	I	I	I	I
12	G104.7513+02.5857	II[(t)]	II	II	II	II
13	**G104.7544+02.6101**	I[(t)]	I	I	II	II
14	G104.7601+02.7154	II[(t)]	II	II	II	II
15	G104.7611+02.7176	II[(t)]	II	II	II	II
16	**G104.7622+02.7179**	I[(t)]	I	I	II	II
17	**G104.7631+02.7173**	I[(t)]	I	I	II	II
18	G104.7934+02.6942	I[(t)]	I	I	I	I
19	G104.8126+02.6920	I[(t)]	I	I	I	I
20	G104.8206+02.7026	I[(t)]	I	I	I	I
21	G104.8225+02.8393	I[(t)]	I	I	I	I
22	**G104.8383+02.7091**	II[(t)]	II	I	I	II
23	**G104.8510+02.7998**	I[(t)]	I	I	II	II
24	**G104.9019+02.7100**	II	–	I	I	I

Notes. [(a)]For sources with different classifications obtained here and in Winston et al. (2019), the names are written in bold. [(b)]ML, CC, and CC (no dist.) refer to the classifications obtained using machine learning (Section 3.3), color–color diagrams when the kinematic distance was also available (Section 3.2), and color–color diagrams created using the criteria which did not require distance (see text). [(c)]IRAC and MIPS refer to the classification obtained in Winston et al. (2019) using only IRAC bands, and IRAC and MIPS bands, respectively. [(d)]For sources which belonged to the training or test samples in the ML calculations, the class is adopted from the CC method and marked with [(t)].

Appendix E: Multi-wavelength images of the IRAS 22147 region

We present multi-wavelength images of the IRAS 22147 region: 2MASS J, H, and K_s; *Spitzer*/SMOG IRAC 3.6, 4.5, 5.8, and 8.0 μm; *Spitzer*/SMOG MIPS 24 μm; AllWISE 12 and 22 μm; *Herschel* PACS 70, 100, and 160 μm; and *Herschel* SPIRE 250, 350, and 500 μm and VLA 1.4 GHz. All the images show the same field of view. The wavelengths and angular resolutions of the surveys are indicated in the images.

Appendix F: SED fits of YSO candidates in the IRAS 22147 region

Figure F.1 shows the SEDs of YSOs in the IRAS 22147 region with the best-fit Robitaille (2017) models. The resulting physical parameters are provided in Table 3.
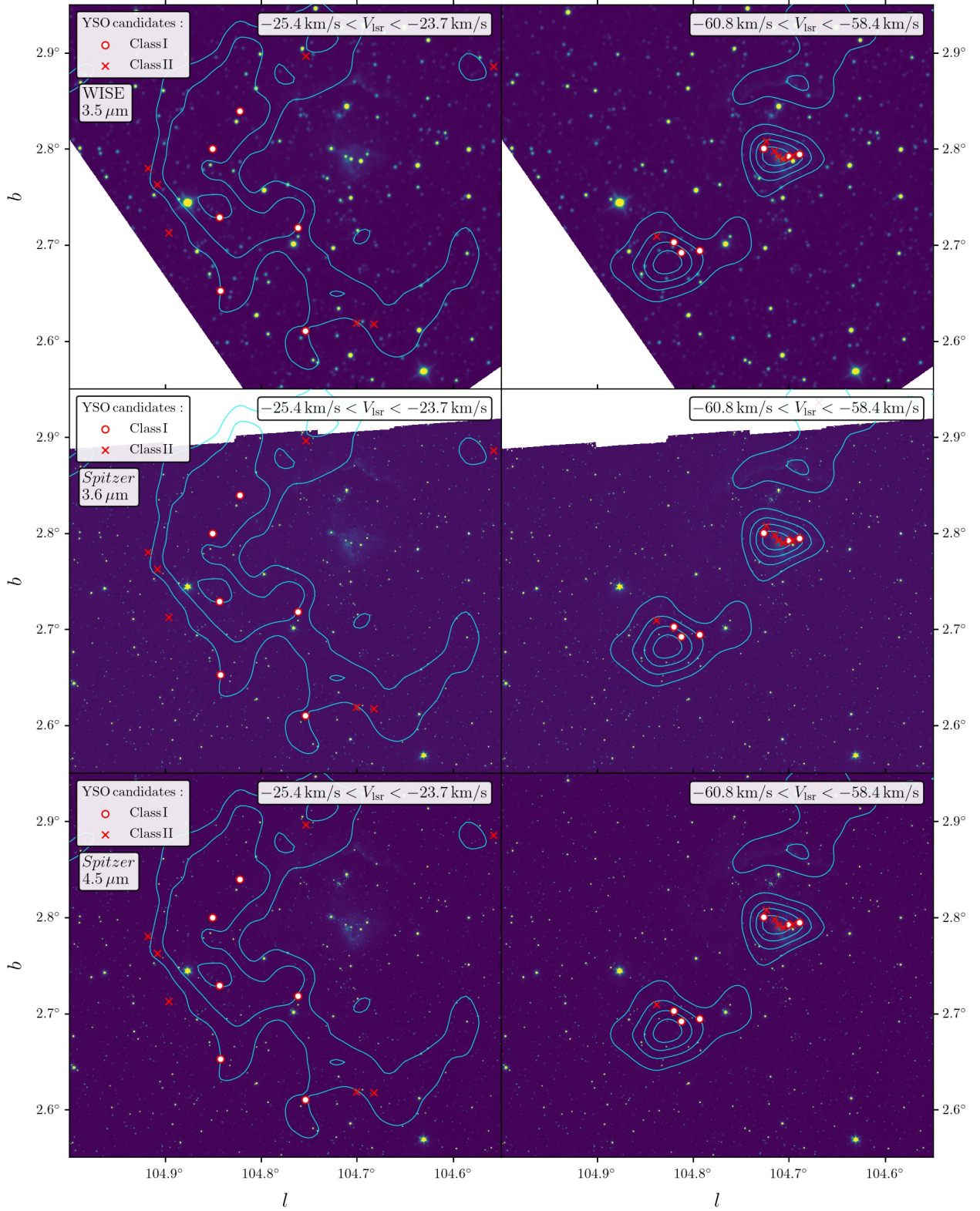
Fig. E.1. Available multi-band imaging of the IRAS 22147 region, as noted in the top left corner of each row. The left and right columns in each row correspond to the two main local standard of rest (LSR) velocity ranges found in the ^{12}CO $J = 1 - 0$ cubes, as specified in the top right corner of each stamp. We overlaid the YSO candidates in red and the ^{12}CO $J = 1 - 0$ contours in cyan color.
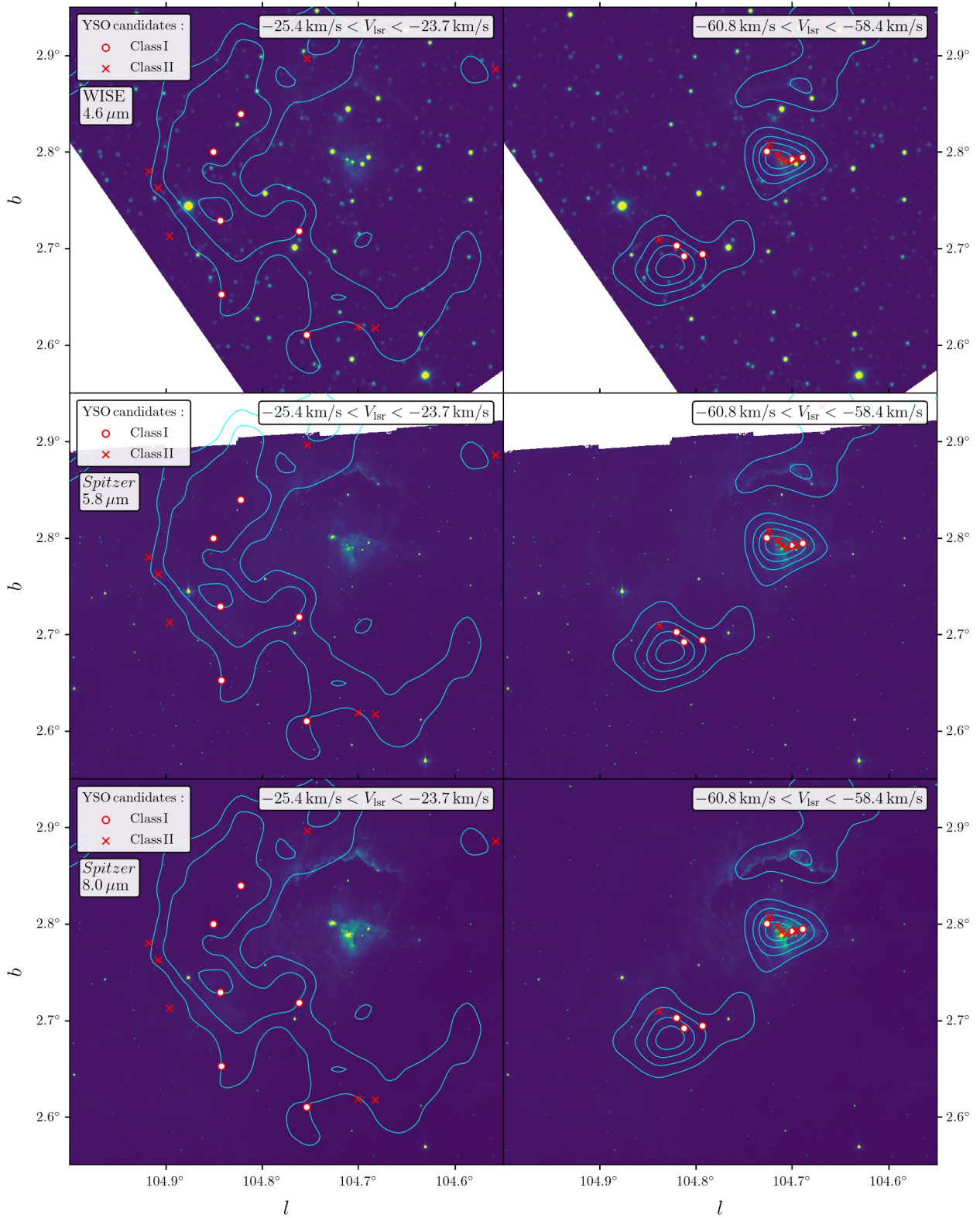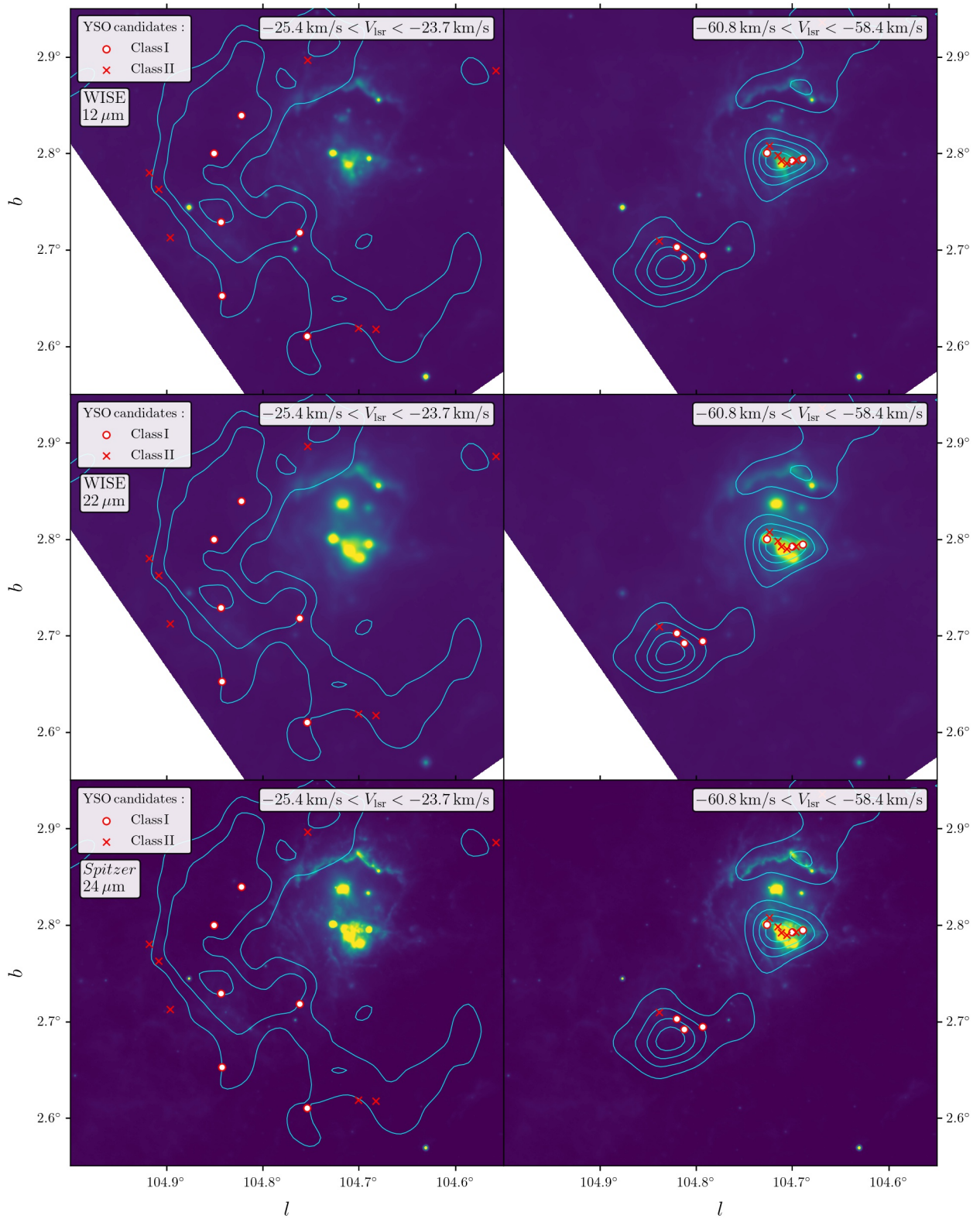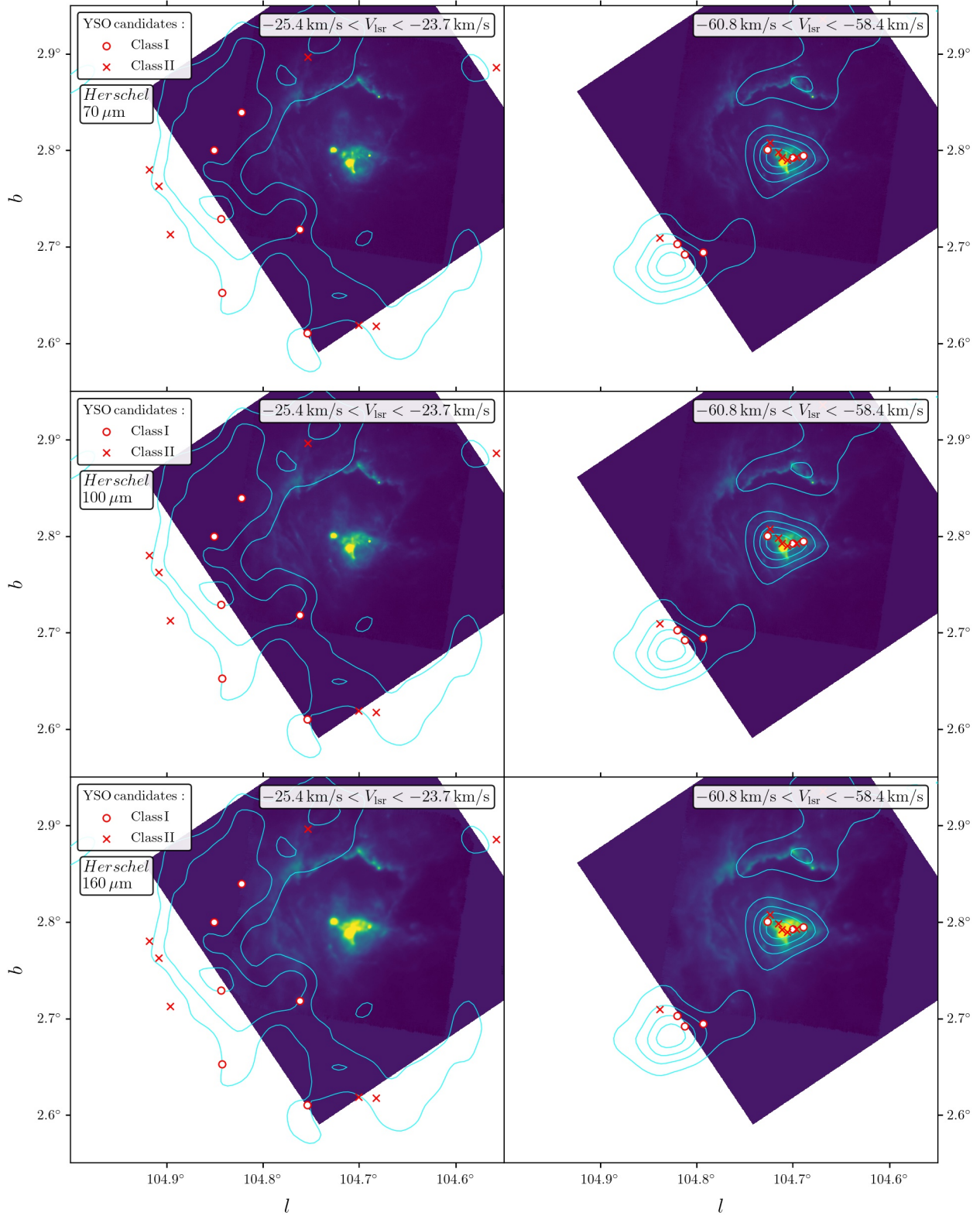
Fig. E.1. (continued).

Fig. E.1. (continued).
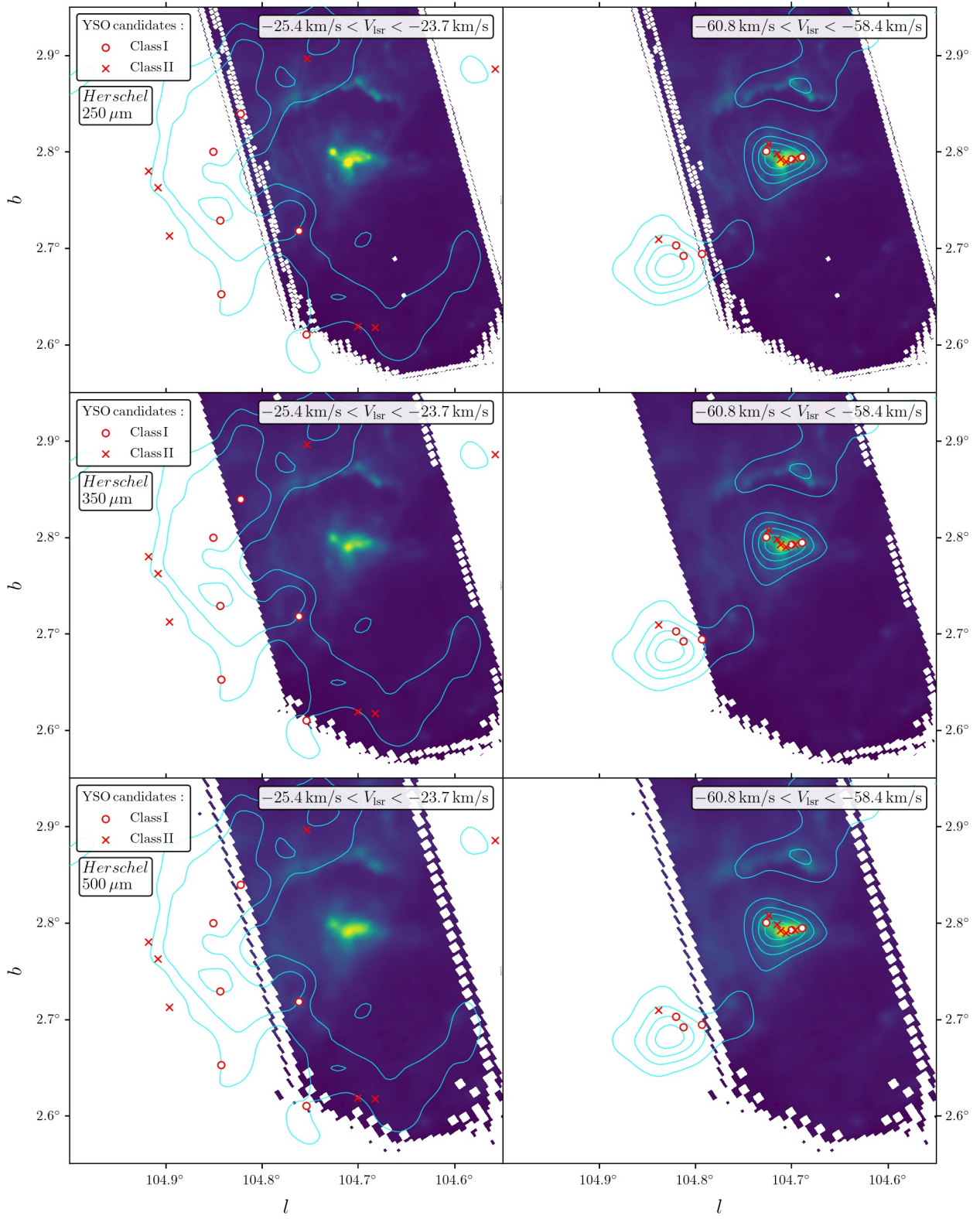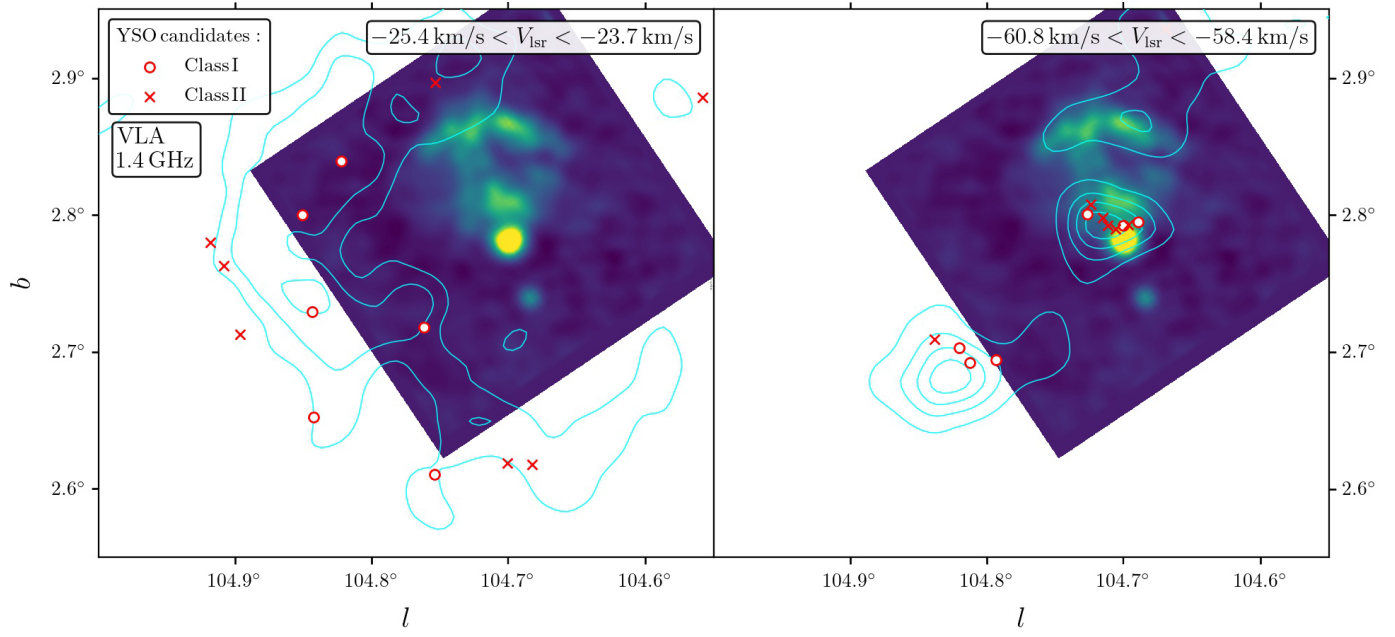
Fig. E.1. (continued).
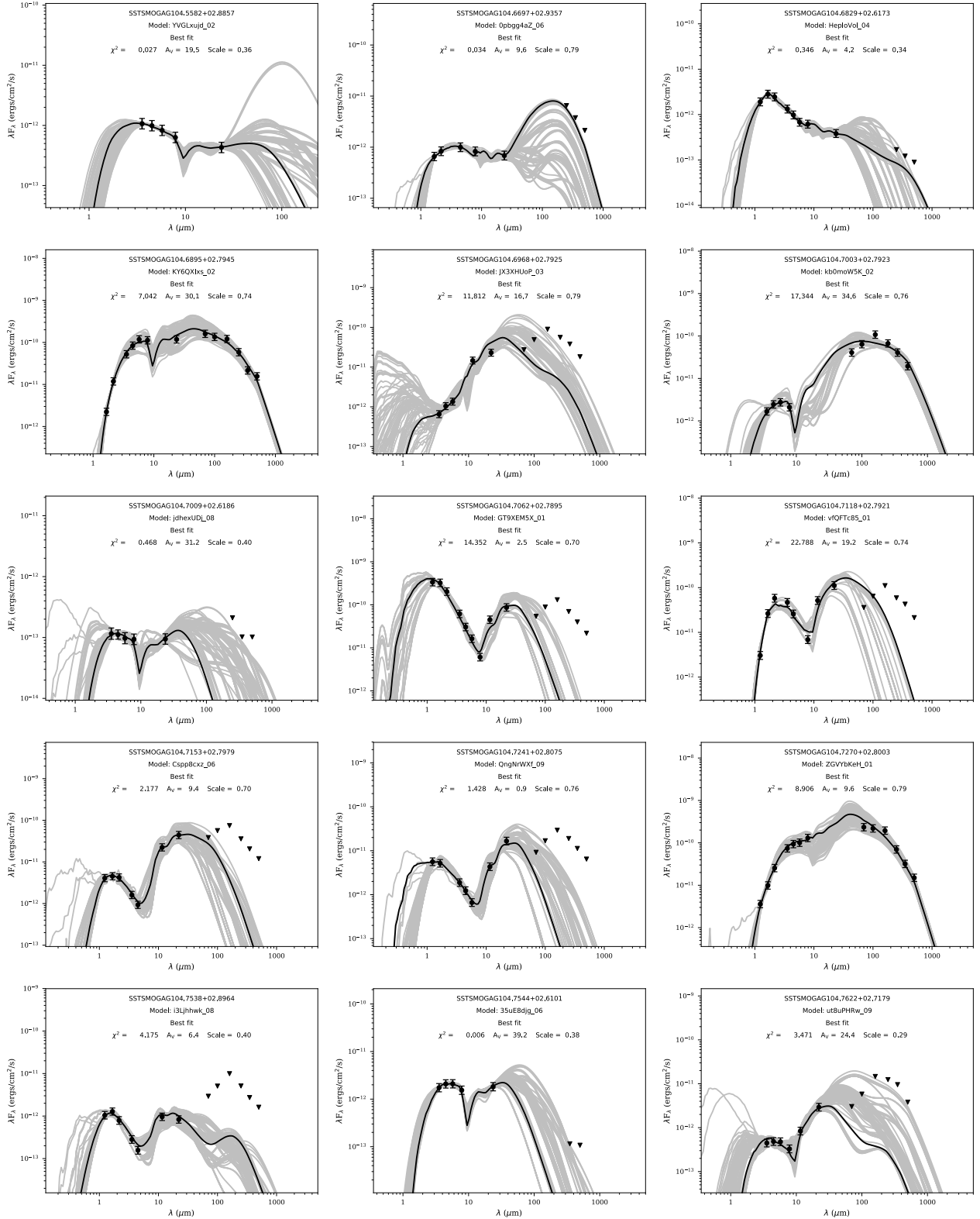
Fig. E.1. (continued).

Fig. E.1. (continued).

Fig. F.1. SEDs of candidate YSOs with well-fit Robitaille (2017) YSO models. The best-fit model is indicated with the black solid line; gray lines show the YSO models with χ^2 between $\chi^2_{\rm best}$ and $\chi^2_{\rm best}+F$, where F is a threshold parameter which we set to 3 (Sewiło et al. 2019). Filled circles and triangles are valid flux values and flux upper limits, respectively. The values of a reduced χ^2 and interstellar visual extinction for the best-fit model are indicated in the plots.
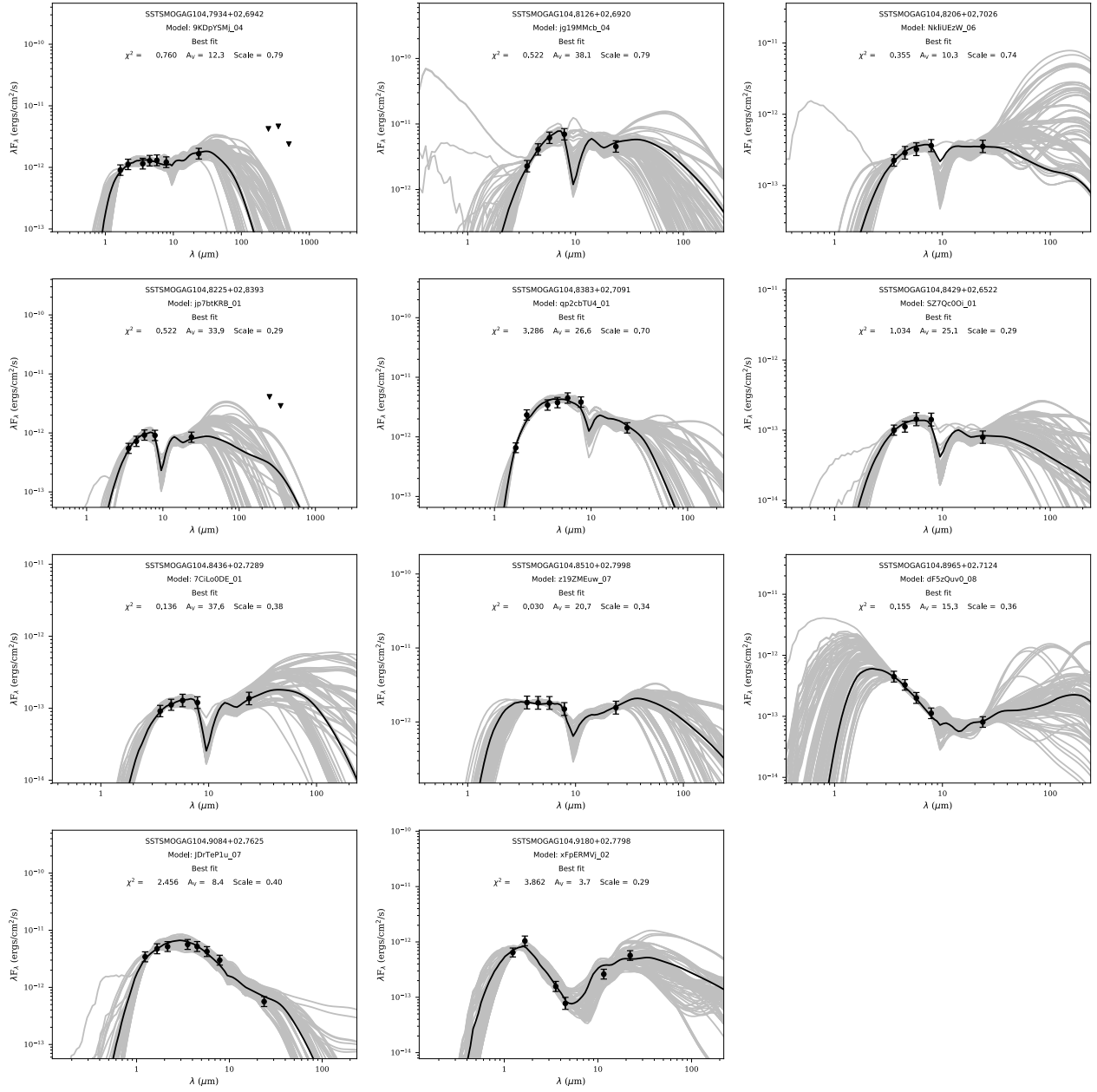
Fig. F.1. (continued).