Infrared surveys of Galactic star clusters

Valentin D. Ivanov,1 Maria Messineo,2 Qingfeng Zhu,3 Don Figer,2 J. Borissova,4 R. Kurtev4 and G. R. Ivanov5

1 ESO, Ave. Alonso de Cordova 3107, Casilla 19, Santiago 19001, Chile email: vivanov@eso.org
2 Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, 54 Lomb Memorial Drive, Rochester, NY 14623-5604, USA email: [messineo, figer]@cis.rit.edu
3 Key Laboratory for Research in Galaxies and Cosmology, The Chinese Academy of Sciences, University of Science and Technology of China, Hefei, Anhui, 230026, China email: zhuqf@ustc.edu.cn
4 Departamento de Física y Astronomía, Facultad de Ciencias, Universidad de Valparaíso, Ave. Gran Bretaña 1111, Playa Ancha, Casilla 53, Valparaíso, Chile email: [jura.borissova, radostin.kurtev]@uv.cl
5 Department of Astronomy, Sofia University, Bulgaria, 5 James Bourchier, 1164 Sofia, Bulgaria email: gr_ivanov@phys.uni-sofia.bg

Abstract. Many attempts have been made to carry out a complete observational census of Milky Way star clusters based on recent near- and mid-infrared surveys. However, more clusters are still being discovered, indicating that existing catalogs are incomplete. We attempt to estimate the total number of supermassive (SM; $M_{cl} \geq 10^4 M_\odot$) clusters in the Galaxy, and to improve the yield from the automated cluster searches. Assuming that the ‘local’ census of SM clusters is complete, and that their surface density accross the disk follows that of the stars, we predict that the Milky Way contains $\geq 81 \pm 21$ SM clusters. We apply a cluster-detection algorithm to the 2mass Point Source Catalog after a preliminary color and/or magnitude selection of the point sources to improves the surface-density cluster-to-field contrast. Our algorithm identified 94 new candidates, and re-identified 34 known clusters. During the visual inspection, we detected an additional 41 new candidates, and re-identified 32 known objects. Preliminary characterization suggests that the new list may contain red-supergiant, open and globular clusters.

Keywords. globular clusters: general, open clusters and associations: general, Galaxy: bulge, Galaxy: disk, Galaxy: structure, infrared: general

1. Introduction

Lada & Lada (2003) estimated that 70–90% of the stars formed in giant molecular clouds are born in embedded clusters. The optical census has yielded large numbers of clusters (i.e., Phelps & Janes 1993, 1994), but most objects with the highest scientific impact (i.e., the Arches cluster; Cotera et al. 1992, 1996; Figer 1995; Nagata et al. 1996) have been discovered in the infrared (IR). It is not surprising that once the TWO MICRON ALL SKY SURVEY (2mass; Skrutskie et al. 2006), the DEEP NEAR-INFRARED SOUTHERN SKY SURVEY (denis; Epchtein et al. 1997), the Spitzer Space Telescope’s GALACTIC LEGACY INFRARED MID-PLANE SURVEY EXTRAORDINAIRE (GLIMPSE; Benjamin et al. 2003) and, most recently, the UKIRT INFRARED DEEP SKY SURVEY (UKIDSS; Lawrence et al. 2007) became available, a number of teams attempted to search for as yet unknown obscured clusters.

There were two basic strategies to identify new cluster candidates. The first consists of targeted searches around known IR and radio sources, or in areas of interest (e.g.,
Figure 1. 2MASS three-color images of red-supergiant-containing clusters, RSGC 1 ([BDS2003] 122 or BDSB 122; left) and RSGC 2 (Stephenson 2; right). Blue, green and red correspond to $J$, $H$ and $K_s$, respectively. The fields cover 6.7×6.7 arcmin$^2$. North is up, East to the left.

Galactic bulge or known nearby star-forming regions; i.e., Dutra & Bica 2000; Kronberger et al. 2006). The second is unbiased: it is based on automated techniques that apply sets of objective rules to identify overdensities in flux or stellar surface-number density in the pixel data or in the point-source catalogs, respectively. The morphology of some known supermassive clusters clearly demonstrates that they can be extremely difficult to detect because of their small angular size and compact structure. Both methods were used for the first time to find new Milky Way clusters in near-IR (NIR) surveys by Reylé & Robin (2002), using the denis data. Ivanov et al. (2002) and Borissova et al. (2003) reported lists of cluster candidates identified as number-density peaks in the 2MASS Point Source Catalog (PSC; Cutri et al. 2003). These studies were soon followed by others (i.e., Froebrich et al. 2007a, 2007b, 2008; Mercer et al. 2005; Davies et al. 2007), who found additional cluster candidates.

Our team pursued a rigorous campaign to determine the nature of the candidates and develop relevant diagnostic techniques (Borissova et al. 2005, 2006, 2008; Beletsky et al. 2009; Ivanov & Borissova 2002; Ivanov et al. 2002, 2005a, 2005b; Kurtev et al. 2007, 2008). Our NIR imaging of candidates selected through visual inspection suggests ~40–50% contamination by multiple stars and extended sources. The visual inspection also shows that automated searches fail in the areas of highly variable extinction, i.e., they produce many spurious detections around dark dusty clouds or in areas of lower extinction, as also pointed by Froebrich et al. (2007a). The incompleteness of NIR-based cluster searches was also demonstrated in the mid-IR Spitzer Space Telescope-based work of Mercer et al. (2005). Most disturbingly, some of their mid-IR cluster candidates are clearly visible on 2MASS images but they were missed by both automated searches and visual inspection.

These considerations prompted us (i) to put in prospective the results from the current searches by estimating, at least approximately, the total number of detectable clusters in the Milky Way and (ii) to further improve our search algorithm. The visual inspection of three-color images suggest that most clusters stand out really well because cluster colors differ from those of foreground and background stars, because of either intrinsically different stellar populations or the dust reddening suffered by the clusters at greater distances than the foreground stars (Figure 1). The cluster-number estimate is based on the fundamental assumptions that the clusters follow the spatial distribution of the field stars and that we know the local density of the most massive clusters ($M_{\text{cl}} \geq 10^4 M_\odot$), which we used as normalization.

Here we describe additional color and magnitude criteria to the ‘normal’ overdensity search, slightly modified with respect the description in Ivanov et al. (2002). We optimized
our criteria for clusters containing red supergiants (RSGs), because many of the known supermassive Milky Way clusters host this type of star (i.e., Davies et al. 2007). The criteria can easily be modified to favor clusters containing other stellar types. We provide a list of new cluster candidates, and some re-identified clusters and cluster candidates. For all objects, we carry out a preliminary analysis and characterization based on the available 2MASS and GLIMPSE data.

2. Total number of Milky Way clusters

We attempt here to obtain a zeroth-order semi-empirical estimate of the total number of clusters in the Milky Way from simple scaling relations, using the following assumptions:

1. The surface density of clusters across the Galactic X–Y plane follows the stellar surface density, described by an exponential law (de Vaucouleurs 1959) with the same scale length \( R_{\text{eff}} \) as for the stars.

2. All ‘local’ supermassive clusters with masses \( M_{\text{cl}} \geq 10^4 \, M_\odot \) in the solar vicinity are known, so we can scale the exponential law to match the measured surface density in the high-mass bin of the mass distribution.

3. The cluster mass distribution is described by a power law with an index of \( -2 \) (Piskunov et al. 2008) and cutoffs at \( 10^2 \) and \( 10^5 \, M_\odot \), so we can extrapolate the number of clusters to lower-mass bins.

The total number of clusters in the Milky Way is obtained by integrating over both the spatial extent of the Galaxy and the cluster mass range.

The main argument supporting the first assumption comes from the estimate of Lada & Lada (2003) that most stars formed in giant molecular clouds are born in embedded clusters. A word of caution is needed as regards the third assumption, however. Larsen (2006) argued for a truncated cluster initial mass function, which can be parametrized well using a Schechter function (Schechter 1976) as demonstrated by Gieles (2009). The tentative current upper-mass limit is \( \sim 10^5 \, M_\odot \) for our Galaxy, well above the mass bin that we use for normalization, so it is likely to have a negligible effect on our estimates.

We adopt an exponential law with \( R_{\text{eff}} = 2.5 \, \text{kpc} \), starting at \( R_{\text{in}} = 2 \, \text{kpc} \) and with an outer cutoff at \( R_{\text{max}} = 20 \, \text{kpc} \) (Robin et al. 2003; Picaud & Robin 2004; Brand & Wouterloot 2007). The innermost disk profile is flat, with a constant surface density at a level equal to that of the exponential law at \( R_{\text{in}} \) (Freudenreich 1998; Lépine & Leroy 2000; López-Corredoira et al. 2001). The profile is shown in Figure 2 (left, top).

For the purpose of local normalization of the supermassive clusters, we considered the objects from table 4 in Messineo et al. (2009), with some updates discussed in detail in Ivanov et al. (2010). Up to which distance from the Sun is the supermassive cluster sample likely complete? The bottom right-hand panel of Figure 2 shows the \( K \)-band extinction, \( A_K \), as a function of the distance, \( D \), from the Sun. Some clusters, even at \( D \sim 7–8 \, \text{kpc} \), suffer from low extinction: \( A_K \sim 0.5–1.5 \, \text{mag} \), i.e., they are located behind ‘holes’ in the Galactic dust. None of the known clusters suffer from extinction much higher than \( A_K \sim 3 \, \text{mag} \), regardless of distance, suggesting that extinction, and not distance, is the main obstacle that prevents us from detecting clusters in the Galactic plane. Indeed, the apparent magnitude of the stars in a cluster near the Galactic Center would be only about 0.6 mag fainter than that for a cluster at \( D \sim 6 \, \text{kpc} \), if it was not for the effects of extinction. Bearing in mind the limited statistics, we conclude that the sample is likely to become incomplete for \( A_K \geq 3 \, \text{mag} \), which for at least some clusters occurs at a distance \( D \sim 7–8 \, \text{kpc} \), as can be seen in Figure 2. Indeed, the cumulative distribution (right, middle panel) deviates from the expected distribution for constant cluster surface...
Figure 2. (left) Exponential profile of the cluster surface density, normalized at the peak value (top), fractional and number distribution of the supermassive clusters as a function of Galactocentric distance (middle and bottom, respectively). See Section 2 for details. (right) Distribution of the supermassive clusters as a function of their distance to the Sun (top) and the corresponding cumulative histogram (middle). The K-band absorption as function of distance is also shown (bottom). The vertical long-dashed lines mark the maximum distance to which we consider the sample complete. The centers of the bins that fall below this limit are marked with crosses for clarity. The short-dashed line shows the predicted cumulative distribution for homogeneously distributed supermassive clusters with constant surface density, $n = 0.062 \pm 0.016 \, \text{kpc}^{-2}$. The dashed lines are the 1σ Poissonian uncertainties, based on the cluster numbers.

density (short-dashed line) for greater distances. Based on these considerations, we will make a somewhat arbitrary assumption that the sample is complete to $D \leq 6 \, \text{kpc}$, an upper limit consistent with the data, keeping in mind that the known clusters may reside in low-extinction windows throughout the disk. Seven clusters are located within the nearest 6 kpc. Averaging within this volume, we obtain a cluster surface density $n = 0.062 \pm 0.016 \, \text{kpc}^{-2}$. The error is random and is dominated by the small-number statistics. This calculation neglects the local density gradient.

Scaling the modified exponential radial profile to match the ‘local’ surface density, and integrating over the entire Galaxy, we estimate that it contains $\sim 81 \pm 21$ clusters with masses in the range $10^4 - 10^5 \, \text{M}_\odot$. We remind the reader that this number is a lower limit. What fraction of these clusters falls within the boundaries of our cluster search ($0 \leq l \leq 360 \, \text{deg}$, $-1.5 \leq b \leq 1.5 \, \text{deg}$)? Adopting a vertical cluster distribution scale height of $H_d = 55 \, \text{pc}$ (Phelps & Janes 1994) and a displacement of the Sun from the Galactic disk along the $Z$ axis of 15 pc (Drimmel & Spergel 2001), we estimate that $\sim 79 \pm 20$ of the $\sim 81 \pm 21$ supermassive clusters in the Milky Way probably fall into the area covered by our search, of which $\sim 39 \pm 10$ are closer than the Galactic Center (i.e., observable), approximately three times exceeding the number of known objects.

3. Search algorithm with color/magnitude criteria

We imposed color/magnitude criteria on the stars in the 2MASS PSC to isolate the cluster members on unique sequences without counterparts on the field color–magnitude diagrams (CMDs). We create a two-step search algorithm, (i) pre-selection of stars from the 2MASS PSC that meet certain color and/or magnitude criteria and (ii) detection of statistically significant peaks in the density map constructed from this pre-selection.

The RSG stars in known RSG clusters (Figure 3) occupy relatively compact loci at $1.2 \leq J - K_s \leq 2.2 \, \text{mag}$ for RSGC1 ([BDS2003] 122 or BDSB 122) and $0.2 \leq J - K_s \leq 1.7 \, \text{mag}$ for RSGC2 (Stephenson 2). These limits were chosen to separate the
Figure 3. 2MASS color–magnitude diagrams of red-supergiant-containing clusters, RSGC1 ([BDS2003] 122 or BDSB 122; left) and RSGC 2 (Stephenson 2; right) and their surroundings. Stars within $R_{cl} \leq 2$ arcmin are shown in the top panels and stars within equal-area annuli, $7.0 \leq R \leq 7.28$ arcmin, are shown in the bottom panels. The double-hatched areas show the combination of the color and magnitude limits designed to select only RSG stars, and the single-hatched areas represent the adopted magnitude limits. The results from number counts are also given. They indicate that the total number of stars inside the ‘cluster’ is smaller than the total number of stars in the ‘field’ annulus: 347 versus 414 for RSGC 1 and 376 versus 426 for RSGC 2. This is caused by crowding and the presence of bright stars in the cluster area. It is probably the reason why unconstrained number counts used in Ivanov et al. (2002) and Froebrich et al. (2007a) failed to detect these clusters.

apparent regions of the most significant overdensities with respect to the field, but they can be limited further. The optimal color limits for each individual cluster depend on the distance, reddening and cluster age. To alleviate these complications, we adopted only an RSG magnitude limit of $K_s \leq 9.0$ mag. For the RSGs in a 12 Myr cluster with absolute magnitudes of $M_K \sim -10$ to $-12$ mag (Davies et al. 2008), this corresponds to a maximum distance modulus of $(m - M)_K \sim 12$–21 mag, or a distance limit of $\sim 16$–40 kpc for $A_K = 3$ mag, or $\sim 4$–10 kpc for $A_K = 6$ mag. Therefore, unless the extinction significantly exceeds that for the Arches, Quintuplet or the Galactic Center cluster (i.e., more than one magnitude in $A_K$), the adopted magnitude limit should not exclude RSG-dominated clusters on the near side of the Milky Way’s center.

The second step in the search is to build two-dimensional surface-density maps to detect the peaks. This is similar to the technique described by Ivanov et al. (2002), except—to make the search independent from a fixed grid—we calculated the surface density within a radius of $R_{\text{max}} = 1.5$ arcmin around each pre-selected star, using only the other pre-selected stars. The surface densities suffer from the problem of not having been measured at equidistant points, which is inconvenient for further analysis, so we still introduce a grid, but it is finer than the $R_{\text{max}}$: the bin size is $0.5 \times 0.5$ arcmin$^2$, effectively preserving the angular ‘resolution’ determined by $R_{\text{max}}$. Next, we remove the background by subtracting from the grid a median-smoothed surface-density distribution, where the median-filter size was $21 \times 21$ bins ($10.5 \times 10.5$ arcmin$^2$). The final step was to identify clusters as groups of neighboring bins, where neighbors are defined as bins separated by less than $R_{\text{neighbor}} = 1.5$ arcmin. We found it advantageous to smooth the two-dimensional histogram with a $5 \times 5$ boxcar function before running the cluster-identification routine. This improved the detection of irregularly shaped candidates and objects that are larger than $R_{\text{neighbor}}$. 
Once the cluster candidates were identified, we visually inspected the 2MASS-generated CMDs and three-color images of objects covering at least four grid bins, with an overdensity of $\geq 10\sigma$ above the background. We also inspected the GLIMPSE survey, where available, for a limited number of objects that were considered interesting because of their apparent morphology or their CMDs. A study of more GLIMPSE and UKIDSS images of our candidates is forthcoming. The visual inspection helped to discover serendipitously a number of new clusters that were missed by the algorithm. These were usually compact, and we attribute the detection problem to the fundamental limitation of the 2MASS PCS (or any other catalog) that cannot resolve crowded regions.

For the purpose of this study, we limited ourselves to within $\pm 1.5$ deg from the Galactic equator, an area of 1080 deg$^2$. This region includes the most obscured parts of the Galaxy, where the cluster census is likely to be least complete. Barring extinction effects, it covers the highest apparent stellar-density areas in the Milky Way, and therefore the location of most active cluster formation and disruption.

The list of identified objects, together with preliminary characterization from public data, will appear in Ivanov et al. (2010). All identified objects are concentrated within $\sim 60$ deg from the Galactic Center (i.e., roughly within the solar circle). There are more objects within the $0 \leq l \leq 60$ deg range than in $300 \leq l \leq 360$ deg (133 versus 68), for both automatically (91 versus 37) and visually (41 versus 31) detected cluster candidates, contrary to the samples of Froebrich et al. (2007b), Mercer et al. (2005) and the combined Bica and Dutra catalogs. This result may reflect a spatial variation in the global star-formation rate in the Galaxy, e.g., due to the Galactic bar. Our search may possibly be selecting more distant clusters. Incidentally, the three known RSG clusters are located in the overdense quadrant.

The automatically discovered cluster candidates are, on average, larger than those detected visually, with mean radii of $102 \pm 45$ versus $70 \pm 40$ arcsec and median radii of 100 versus 60 arcsec, respectively. The new automatically detected candidates have, on average, $\sim 10$ arcsec larger radii than the re-identified known clusters, which suggests that our algorithm finds objects that are more likely to have been missed because they are spread over large areas and have lower contrast with respect to the foreground and background populations. On the other hand, the new, visually detected candidates have on average $\sim 20$ arcsec smaller radii than the re-identified previously known objects. Most previous visual searches were pointed at known radio or mid-IR sources, and typically discovered young clusters with nebulosities. Our visually detected candidates are free from this bias, and we find many clusters without diffuse emission which are likely evolved to beyond the O-star stage.

We attempted to assess the nature of the candidates, even of the known but poorly studied clusters that we re-identified, using three sources:

1. Morphology of the 2MASS CMDs constructed from stars within the cluster radius and a comparison-field CMD based on a circular annulus around the object;
2. Morphology on $\sim 900 \times 900$ arcsec$^2$ three-color 2MASS images;
3. Literature data of sources within 4 arcmin from the candidate.

This is not always straightforward because of significant field contamination and possible random alignments of objects at different distances. Nevertheless, this assessment is useful as a first approximation and as basis for selecting objects for follow-up studies.

We found a total of 25 RSG cluster candidates, of which 21 among the automatically identified candidates (15 new and 6 re-identified previously known clusters). However, among the latter group only four are truly known to contain RSGs, the remaining two are known but poorly studied objects, and we suspect that they contain RSG stars based on their 2MASS CMD features. Finally, we have four more candidates among the new visually
identified objects. We have 36 candidates for old open clusters or globulars, of which 20 were automatically (17 new and 3 previously known) and 16 were visually identified (9 new, 7 previously known). Most appear relatively poor, but this may be a combined effect due to distance, dust extinction and the shallow survey depth. There are 12 massive cluster candidates in our sample, 9 of which were automatically identified (2 new and 7 previously known objects) and 3 were visually identified (1 new and 2 previously known objects). Finally, we provide a list of 15 candidate cluster pairs with separations $\delta \theta \leq 5$ arcmin. All but three of the latter show indications of physical connections.

4. Summary and conclusions

1. The Milky Way is likely to contain $\gtrsim 81 \pm 21$ clusters with masses of $10^4 - 10^5 M_\odot$, estimated from the assumption that the cluster surface density in the disk is described by an exponential law with the same scale length as for the stars, normalized at the ‘local’ density of such clusters. The error is random and reflects only the small number of the clusters used for the normalization. If half of these clusters are located on the far side of the Galactic Center and remain inaccessible because of high extinction, about 30 supermassive clusters remain to be discovered on the near side of our Galaxy.

2. We developed a new search algorithm that uses color/magnitude selection of stars, optimized to pre-select a certain type of clusters, but with higher contrast than a general, unconstrained algorithm. We applied it to the 2MASS PSC, searching mostly (but not exclusively) for potential RSG clusters. We discovered serendipitously a number of additional candidates, usually unresolved.

3. The existing Galactic cluster census is incomplete. We identified 135 new cluster candidates (94 automatically and 41 visually) in a $\pm 1.5$ deg strip along the Galactic plane. We also re-identified 65 known clusters (or cluster candidates, since the cluster nature of many of them has not yet been confirmed; 34 automatically and 32 visually), including most of the well-known supermassive/RSG clusters, giving us confidence in the power of the new selection method.

4. A preliminary characterization is reported for 25 RSG cluster candidates, 36 globular or old open cluster candidates, and 12 massive cluster candidates. These classifications are based only on 2MASS, and (if available) literature information.

5. The number of massive cluster candidates is lower than predicted. This is expected because compact, supermassive clusters like the Arches cluster cannot be resolved by 2MASS. Some additional supermassive clusters similar to RSGC 1, 2 and 3 may hide among our RSG cluster candidates. However, we expect that many of our candidates, upon closer scrutiny, will turn out to be spurious detections, i.e. holes, in the dust.

6. Last but not least, we emphasize that the objects in our lists are candidates until their cluster nature is confirmed based on deeper imaging or spectroscopy.

References
