Obscured star clusters in the Inner Milky Way.
How many massive young clusters are still awaiting detection?

Akash Gupta[1,2*], Valentin D. Ivanov[2], Thomas Preibisch[1], and Dante Minniti[3,4]

[1] Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians-Universität München, Scheinerstr. 1, 81679 München, Germany
e-mail: agupta@ph1.uni-koeln.de, preibisch@usm.uni-muenchen.de
[2] European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching bei München, Germany
e-mail: vivanov@eso.org
[3] Instituto de Astrofísica, Depto. de Ciencias Físicas, Facultad de Ciencias Exactas, Universidad Andrés Bello, Av. Fernandez Concha 700, Las Condes, Santiago, Chile
[4] Vatican Observatory, Vatican City State, V-00120, Italy

Reporter：Zeng Guangya

outline

- 1. Introduction
- 2. Cluster search
- 3. Cluster detection completeness
- 4. Discussion and conclusions

Introduction

- Infrared (IR) surveys allow us to find star clusters close to the Galactic center, like the massive Arches cluster (Nagata et al. 1993, 1995) and to address the question how close to the center of the Milky Way the clusters can survive (Minniti et al. 2021a). However, the near-IR cluster census is also subjected to incompleteness, as shown by Ivanov et al. (2005), because the extinction in the inner Milky Way can be considerable even at these wavelengths (Kurtev et al. 2008).
- In this work, they pursue two goals, with corresponding improvements. First, they want to address, albeit for now in a limited way, the long-neglected question of how complete the existing cluster catalogs are.
- Their other aim is to complete the Milky Way cluster census further while taking advantage of the new and improved search algorithms, finding undiscovered and highly obscured clusters. To underline, their search is optimized for the detection of distant and highly reddened clusters – the type that are likely to suffer the worst incompleteness.
- Their two goals are intertwined because the completeness analysis requires having at hand a reliable cluster detection tool.

Cluster search

Search method

- They have used the GLIMPSE catalog from Spitzer Science (2009) which combines GLIMPSE-I v2.0, GLIMPSE-II v2.0, and GIMPSE-3D.
- Ordering Points To Identify the Clustering Structure Clustering algorithm (OPTICS; Ankerst et al. 1999) is a density-based clustering algorithm, a further development of Density-Based Spatial Clustering of Applications with Noise (DBSCAN; Ester et al. 1996). The method allows for a hierarchy of clusters and given the nature of the star formation which tends to occur in structures of different sizes, from giant star-forming regions to compact star clusters they consider it important to preserve this hierarchy.

- It proved unpractical to run the cluster search over the entire GLIMPSE footprint, because of the memory and speed requirements to handle the entire catalog. Therefore, they fragmented it into $1^\circ \times 1^\circ$ tiles for easier data handling.
- Next, to ensure good-quality data they considered only GLIMPSE sources with photometric errors of <0.2 mag in both [3.6] and [4.5] bands. They also set an upper color limit of $[3.6]-[4.5]=4$ mag, because their experiments indicated that the surface density of these sources is typically too low to meet the minimum number of cluster members that they require.
- Last, they applied color criteria as a proxy for the reddening and for the distance to individual stars, assuming the reddening and the distance are proportional, to zero order. Running the search on each color bin separately minimizes the field star contamination and improves the cluster-to-field contrast.

- In the end, they adopted a single color bin of [0.6, 4] as a compromise between two goals: to ensure that they identify the reddest and most obscured candidates and to exclude the vast blue foreground stellar population.
- They investigated how the OPTICS parameters eps, min_samples affect the candidate yield. A larger eps and a smaller min_samples increase the number of identified candidates; a smaller eps and a large min_samples tend to detect more compact candidates, missing sparser clusters.
- They adopt the (12, 6') parameter combination as a compromise - to be sensitive to compact clusters but also to make their detection more complete.

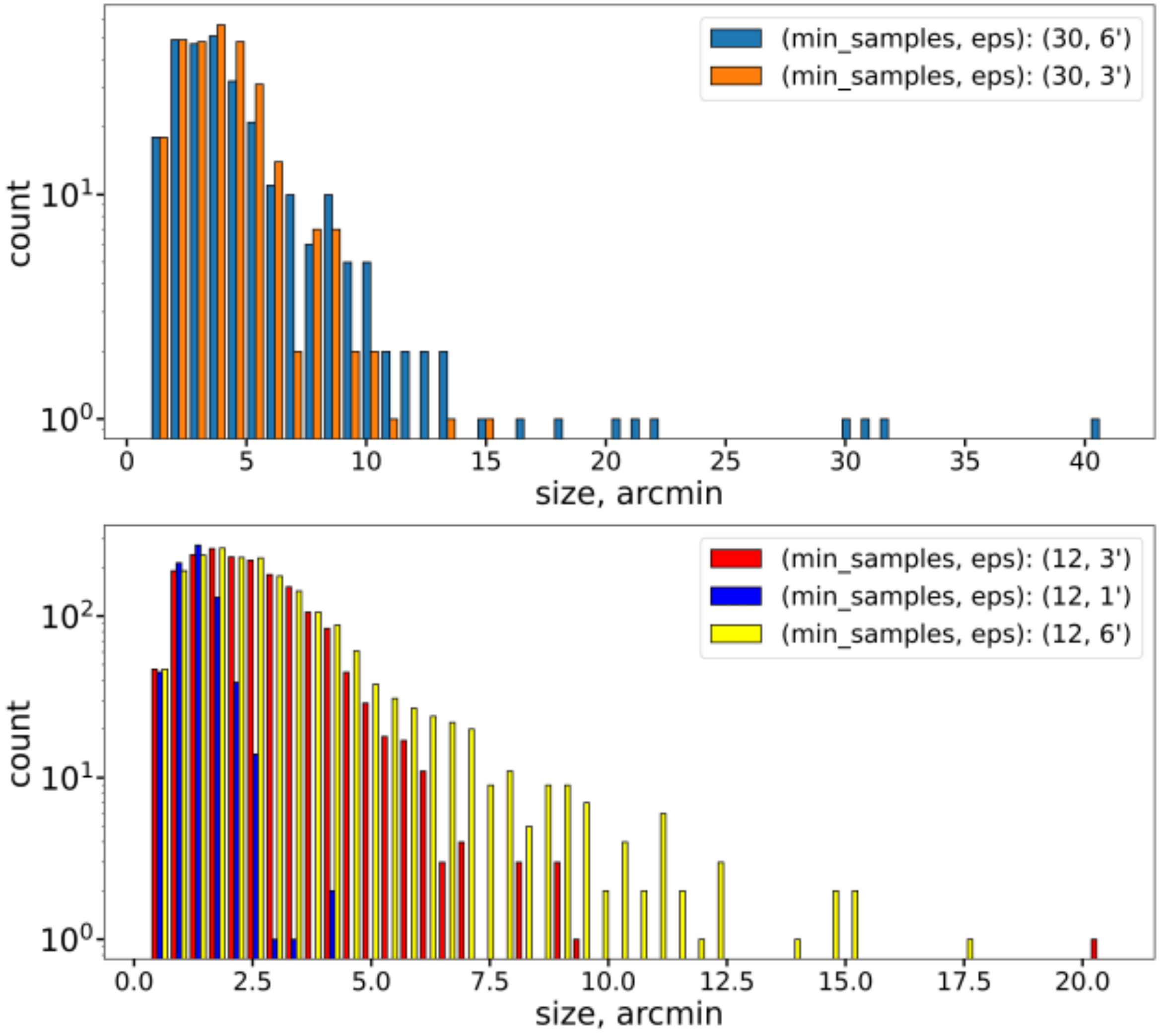
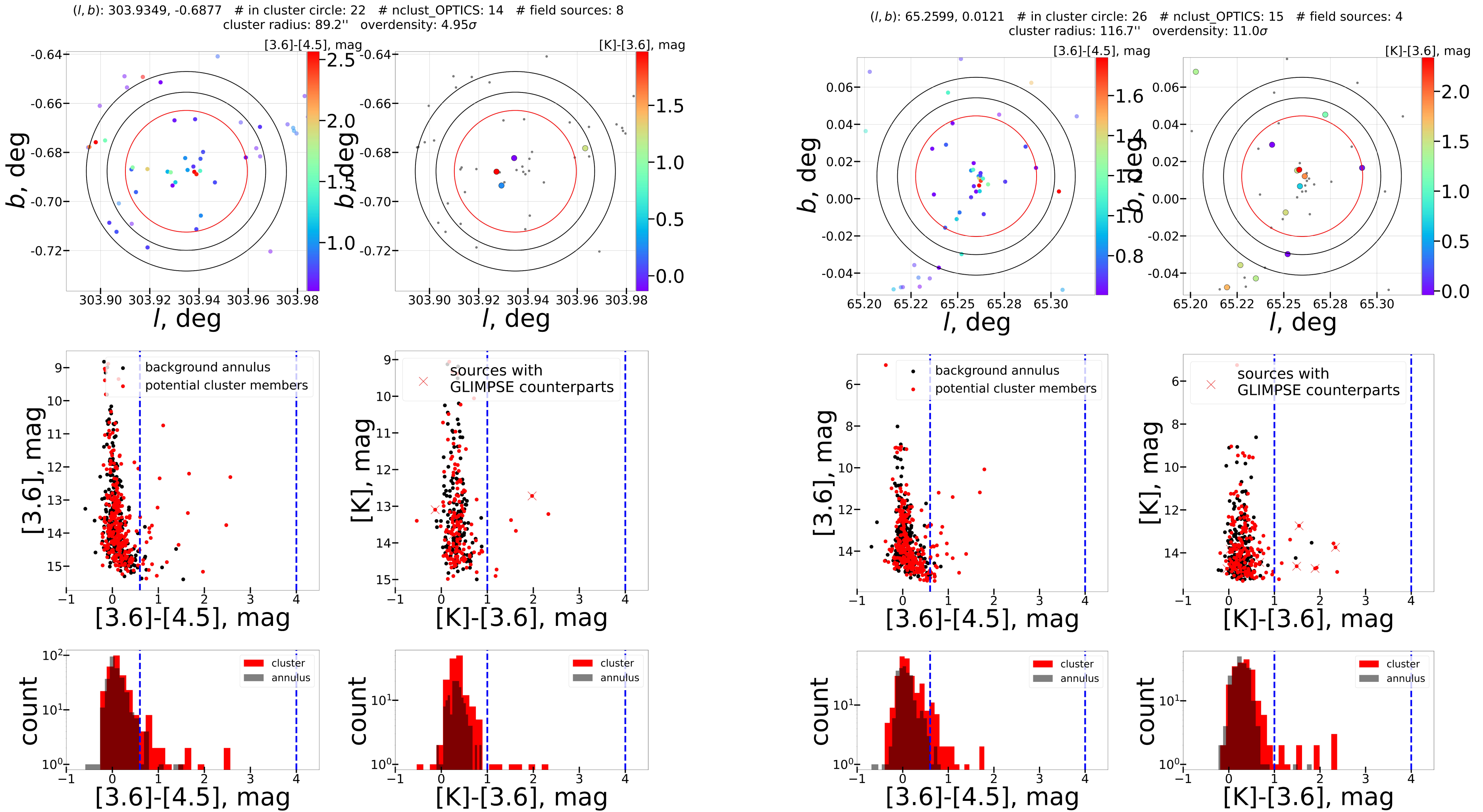


Fig. 1: Distribution of candidates cluster sizes for different search parameters.

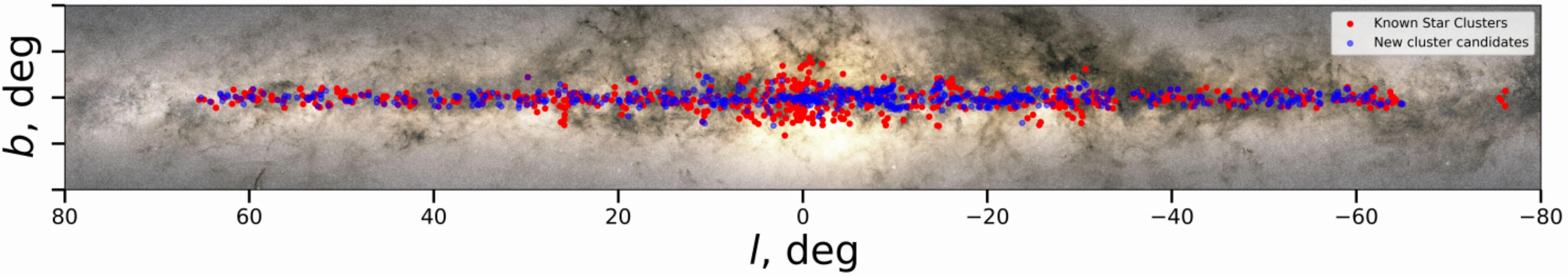
Screening of cluster candidates

- Their search with the adopted parameters yielded 10907 candidates. The experience of previous searches has shown that many, if not most of them are not real clusters.
- In the absence of spectroscopic observations, the sole means to verify the nature of the candidates is an inspection of the CMDs and 3-color images from the available MIR and NIR surveys.
- The main criteria for the true cluster nature of a candidate having a statistically significant excess of stars near the center, that these stars are more reddened than their surrounding counterparts, that they cluster in the CMDs in a locus that resembles a reddened main sequence or red giant branch and show circular symmetry.

- Figure 2 shows a CMD inspection image of a known embedded cluster (left) and a candidate from their catalog (right). The candidate has higher overdensity, supporting the clustered nature of this object. The red dots mark the stars within the cluster region and black dots are the stars in the comparison field annulus (both regions have the same areas and are marked with red and black circles, respectively, on the right panels).



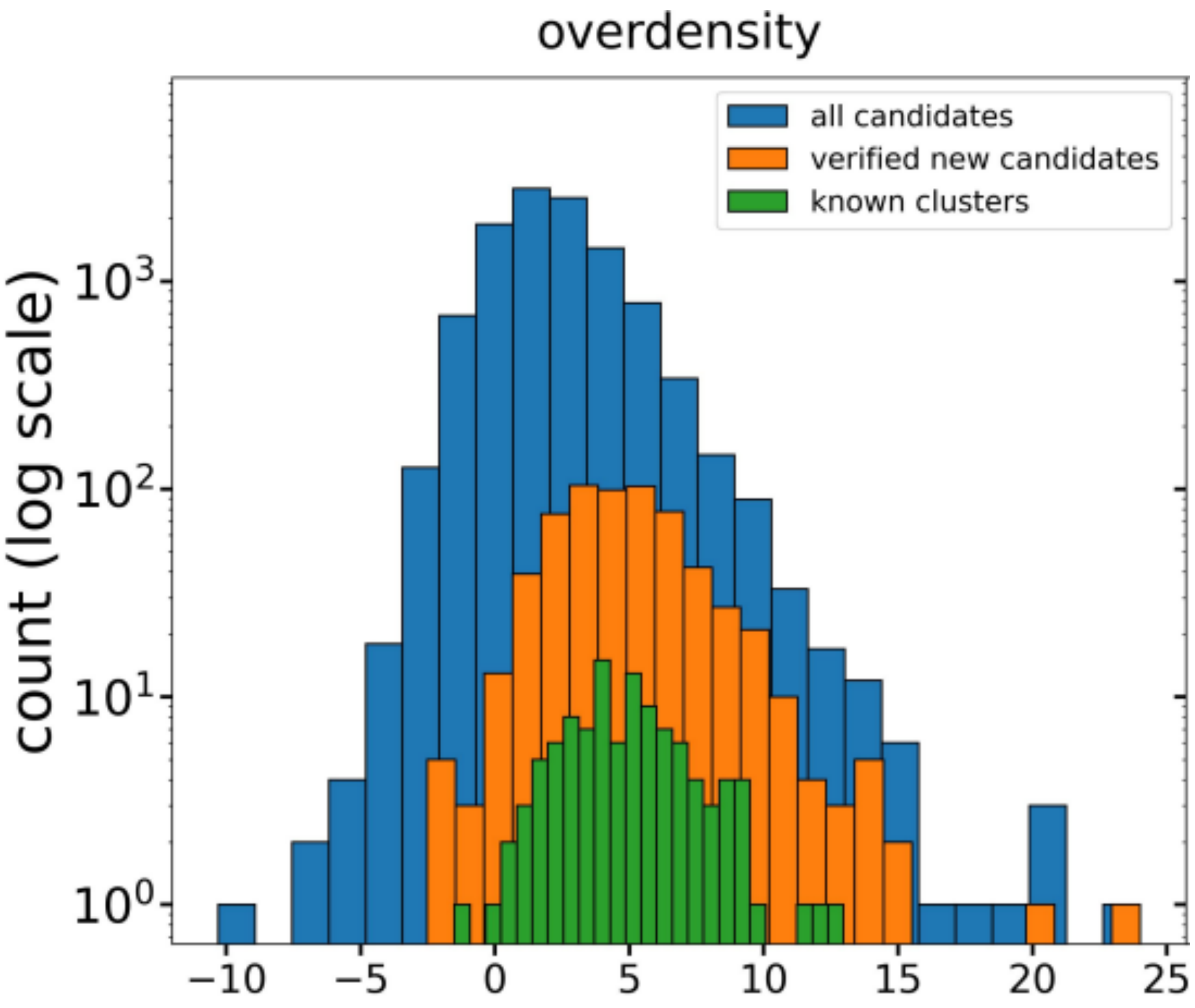
- The inspection also included the 3-color 2MASS (JHKS bands), WISE (W1, W2, W4 bands), and GLIMPSE ([3.6], [4.5], and [8.0] bands) images of candidates that passed the CMD check. This step is important for excluding candidates located next to dense dust clouds that generate a necklace-like chain of clusters.
- Summarizing, these two steps of screening reduced the sample size to 659 candidates. Their location on the Milky Way map is shown in Fig. 4.



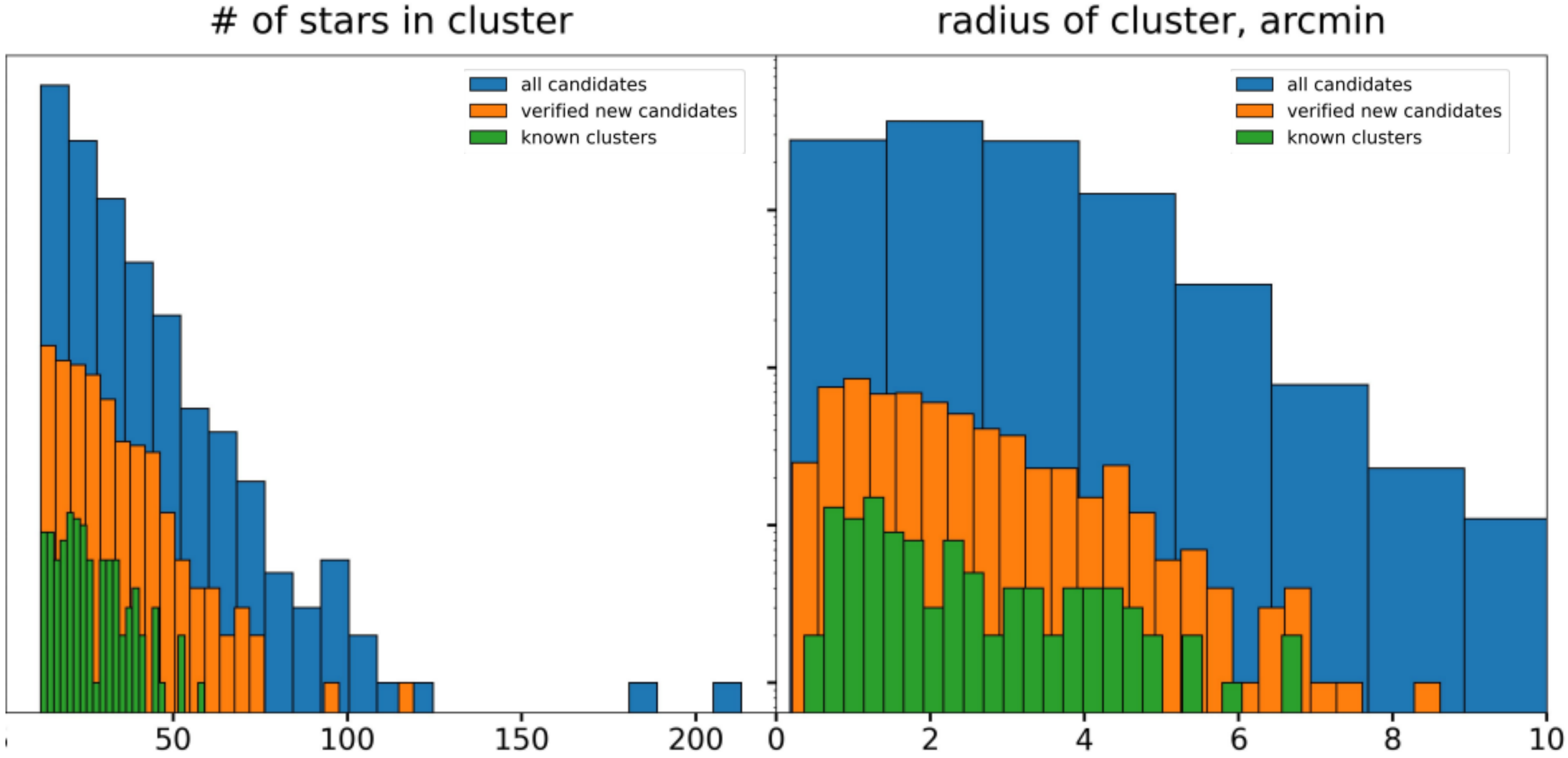
Properties of the sample of cluster candidates

- A SIMBAD search indicated that 106 of the 659 candidates were known: 12 are open clusters, 1 is a globular (2MASS GC01; Hurt et al. 2000) and the rest are extremely young embedded star clusters residing in star-forming regions.

- The verified candidates and the bonafide clusters tend to present somewhat higher overdensities than the average for the initial selection. Most of the highest overdensities with $\sigma \geq 20$ tend to exhibit cluster-like CMDs and/or morphologies and pass through the screening. The rejected high-overdensity candidates are located at the edges of dark clouds.



- The histogram of the number of member stars for the screened sample spans a similar range as the histogram of known clusters and they both have similar shapes. The few outliers again are objects towards the Galactic Center where the crowding is very high. Almost all candidates with more than 75 stars from the initial sample were rejected.
- Finally, the range of measured radii spans 0.5–8.5' . This includes somewhat larger objects than most known clusters. Most of the largest candidates in the initial sample are rejected.



Cluster detection completeness

Generation of artificial clusters

- The most robust way of estimating the detection rate is to carry out a controlled experiment by adding a sample of simulated star clusters. Simulating the entire Milky Way cluster population and measuring the detection rates for different classes of clusters is difficult and here they only consider the case of the most massive clusters, analogous to Westerlund 2.
- First, they removed the field contamination from the sources in the region of the prototype cluster Westerlund 2. They adopted a cluster radius 1' , then they defined an annulus (with the same area as the cluster region and centered on the cluster) where they sample the field population for statistical decontamination of the cluster.

Generation of artificial clusters

- The next step was to shift the "pure" Westerlund 2 population to a grid of predefined positions, extinctions, and reddenings where the artificial clusters would be located in the innermost region of the Milky Way - the region that is the most difficult to cluster search due to crowding and extinction. First, they corrected for the distance modulus and the extinction of Westerlund 2 itself. Then, the data must undergo three modifications: adding to the apparent magnitudes the respective distance modulus, adding to the color the reddening according to the extinction law of Rieke & Lebofsky (1985), and accounting for the decreased angular separation between sources, because of the increased distance.

Recovery of the artificial clusters

- The same algorithm that was used for the cluster search was applied to the catalogs with the artificial clusters. The fraction of recovered clusters varies between 70 % and 95 %. Nearby clusters are easier to identify than more distant ones. However, the higher extinction seems to help to find clusters - possibly because it sets the cluster in color space further apart from the contaminating foreground population that shows bluer colors than the candidate cluster member stars. Spatially, the innermost region at $-2 \leq l \leq 2$ deg stands out with a somewhat lower recovery rate, probably because of the worse crowding near the Galactic Center. In Galactic latitude, there seems to be no drop in the recovery rate at the position of the Milky Way plane within the range of latitude b that is covered by their simulation.

Discussion and conclusions

- They applied a new cluster finding algorithm - OPTICS - on the GLIMPSE survey point source catalog to identify obscured star clusters located in the inner Milky Way and report nearly 500 new objects; they also recovered about 140 previously identified ones.
- The classification and characterization of the new candidates remains outside the scope of this work but the properties of recovered known clusters do hint at the possibility that most of the new candidates would also be embedded and maybe a few would be highly obscured open or globular clusters.
- Here they addressed the important but often neglected question of how successful their algorithm is in finding clusters with a simulation, adding semi-artificial clusters to the GLIMPSE point source catalog, and running the same search algorithm trying to recover them. The achieved recovery fraction is high – in the range 70–95 %, suggesting that the near side of the Milky Way may harbor ~1-3 additional supermassive star clusters. In other words, no large population of hidden supermassive clusters resides inside the Milky Way.

Thanks