DRAFT VERSION DECEMBER 8, 2024
Typeset using LaTeX **twocolumn** style in AASTeX631

Local Group Cluster Search: Star Cluster Catalog for the Small Magellanic Cloud

HANZHANG FENG,[1] L. CLIFTON JOHNSON,[1] SMASH COLLABORATION, AND DELVE COLLABORATION

[1] *Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA) and Department of Physics and Astronomy, Northwestern University, 1800 Sherman Ave., Evanston, IL 60201, USA*

ABSTRACT

We construct a catalog of star clusters from Dark Energy Camera (DECam) images of the Small Magellanic Cloud (SMC) using image classifications collected by the Local Group Cluster Search, a citizen science project hosted on the Zooniverse platform. We identify 252 star clusters within the SMC survey footprint of the SMASH survey. Comparing this catalog to existing compilations in the literature, 41.4% of previous clusters are not confirmed, pointing to a significant population of false positive cluster detections in previous cluster compilations. The cluster catalog's 50% completeness limit is $\sim 331\ M_{\odot}$ at an age of 100 Myr, as derived from comprehensive synthetic cluster tests.

1. INTRODUCTION

Star cluster catalogs for the SMC date back to early work done with photographic plates (Kron 1956, Lindsay 1958) and have continually been added to and improved with successive improvements in observing technology and data quality. The catalogs by Bica and collaborators provide a state-of-the-art summary of past work in the Magellanic system. Specifically, Bica et al. (2008) reviews previous literature and catalogs of extended objects in the Magellanic system and presented a comprehensive catalog of 9305 objects. Bica et al. (2008) cross identified objects from different catalogs by their position, size and object class to eliminate duplicates. Additionally, the authors examined whether newly identified smaller objects were contained within larger objects, such as a cluster within an association, or a large cluster that is reported as two adjacent smaller clusters in another catalog. Among the 9305 objects examined by Bica et al. (2008), 636 are identified as cluster objects within SMC, of which 498 are classified as ordinary clusters. In 2020, Bica et al. (2020) presented an updated catalog of 2741 objects in SMC and Magellanic Bridge following similar cross-identification procedure. Among these objects, 809 are identified as cluster objects and 627 are identified as ordinary cluster.

Another recent set of catalogs focusing on clusters and associations in Magellanic system are Bitsakis et al. (2017) and Bitsakis et al. (2018). Similar to Bica et al. (2008) and Bica et al. (2020), Bitsakis et al. (2017) and Bitsakis et al. (2018) utilized multiband archival data. However, rather than relying on visual identification or adopting cluster identifications from previous studies, Bitsakis et al. (2017) developed a fully automated method using a star count algorithm to identify clusters. Using this algorithm, Bitsakis et al. (2017) found 4850 clusters in the LMC, Bitsakis et al. (2018) found 1319 clusters in the SMC. Bitsakis et al. (2018) compared their catalog with Bica et al. (2008) and found that it successfully recovered 211 out of the 515 clusters listed in the Bica et al. (2008) catalog for the central 18 deg^2 region of SMC. This implies that their catalog discovered 1,108 new clusters in the SMC.

The results of automated, algorithmic cluster identification have raised serious concerns about the nature of the catalogs that are produced. Piatti (2018) and Bica et al. (2020) both question the validity of the Bitsakis et al. (2018) catalog. Piatti (2018) conducted a cross-identification between the clusters in Bitsakis et al. (2018) and Bica et al. (2008) using their coordinates. Their procedure identified only 159 matching objects. Further cross-identification of Bitsakis et al. (2018) catalog with other major catalogs also yielded a low matching percentage ($\leq 20\%$). Piatti (2018) thus concluded that this catalog of SMC cluster is statistically incomplete. Additionally, Piatti (2018) analyzed the cluster frequency of Bitsakis et al. (2018) catalog compared to Bica et al. (2008) catalog and other more recent catalogs, concluding that the new detections in Bitsakis et al. (2018) are unlikely to be faint clusters that were previously undetectable. Bica et al. (2020) also conducted a cross-identification between the Bitsakis et al. (2018) catalog and their updated SMC cluster catalog. They concluded that most of the new detections in Bitsakis et al. (2018) should be classified as associations rather than clusters.

Corresponding author: L. Clifton Johnson
lcj@northwestern.edu

2. DATA

[As this paper is still in progress, I will only present sections that I (Hanzhang Feng) wrote in this draft.]

3. CATALOG CONSTRUCTION

4. SMC STAR CLUSTER CATALOG

5. CATALOG COMPLETENESS

To analyze the completeness of our cluster sample, we synthesized 2000 images of clusters with varying ages, masses, and magnitudes. These synthetic images were mixed with real SMASH cluster images for users to identify. The classification results of these synthetic clusters are shown in the left panels of Figure 1. Here, we applied the same detection threshold used for our catalog, $f_{cluster} = 0.561$. In the right panels of Figure 1, the detection results of synthetic cluster samples are divided into 2D bins by age, mass, and G-band magnitude. The fraction of detected clusters in each 2D bin are shown by color shading.

For each age bin, a completeness curve is derived by fitting the data into the logistic function:

$$C = (1 + e^{-k(x - M_{50})})^{-1}, \quad (1)$$

as applied in Johnson et al. (2022). In this equation, k is the slope of the logistic function, M_{50} is the 50% completeness limit, and x is the physical input parameter — either the G-band magnitude or $\log(M/M_{\odot})$.

Figure 2 provides a rough representation of the logistic function fit results. In this figure, we divide the age range into three broad bins: young, middle-aged, and old clusters. We then fit the completeness as a function of mass and magnitude separately for each age group. Within each sub-figure, mass and magnitude are further divided into several bins. The calculated completeness for each bin, along with their errors, is shown by the blue lines in Figure 2. The light red lines show the logistic curve fit results, while the red dots indicate the M_{50} values as derived from the logistic function fit.

In Figure 1 we fitted the same logistic function for each age bin of width log(Age/yr) = 0.3. For each age bin, mass is divided into bins of width $\log(\mathrm{Mass}/M_{\odot}) = 0.18$; G-band magnitude is divided into bins of width 0.81. We record the M_{50} value for each age bin, and fit these values into

$$M_{50}(\tau) = a \times e^{b(\tau - \tau_{\min})} + c, \quad (2)$$

as in Johnson et al. (2022), to find the age dependence of M_{50}. In this equation, $\tau \equiv \log(\mathrm{Age/yr})$, and $\tau_{\min}$ represents the median τ for the youngest bin. Here, $\tau_{\min} = 6.85$. We fit for constants a, b, and c by minimizing χ^2. The best-fit values for mass are $(a, b, c) = 0.01, 1.60, 2.34$, with reduced $\chi^2 = 3.03$. However, for G-band magnitude, the magnitude-age correlation seems linear. We thus fit linearly and get $M_{50,mag}(\tau) = 0.524\tau + 12.59$, with reduced $\chi^2 = 5.09$. The resulting 50% completeness curves are plotted as dashed black lines in Figure 1.

Table 1. 50% completeness analysis results for different N_{MS} groups

N_{MS}	$M_{50,mass}(\tau = 8)$	a	b	c	reduced χ^2
Total (0-1711)	2.41	0.01	1.6	2.34	3.03
0-27	2.17	0	2.56	2.16	99
27-188	2.36	0	2.8	2.35	5.38
288-462	2.53	0.32	0.48	1.98	5.71
462-1711	2.68	0.09	0.88	2.43	10.45
Cluster catalog	2.52				

NOTE—XXX

5.1. *Additional Completeness Analysis*

In addition to age, mass, and magnitude, another factor influencing catalog completeness is the local stellar density around clusters. In regions of higher stellar density, clusters may be more difficult to identify due to reduced contrast with the surrounding environment. In this paper, stellar density is characterized by the number of main-sequence stars within each cutout image, N_{MS}. For synthetic clusters, N_{MS} ranges from 0 to 1711. For real SMASH cluster images, N_{MS} ranges from 2 to 1524.

We divided the synthetic cluster samples into four groups based on N_{MS} and repeated the procedure described above. For each group, we derived the parameters [a,b,c] along with their corresponding reduced χ^2. The results are recorded in Table 1. As a reference, we extracted the 50% completeness limit mass at log(age/yr) = 8 for each group, denoted as $M_{50,mass}(\tau = 10^8)$.

A plot of $M_{50,mass}(\tau = 8)$ against log of median N_{MS} of each group is shown in Figure 3. The red line in Figure 3 represents the linear fit of this data, given by $M_{50,mass}(\tau = 8) = 0.12 \cdot log(N_{\mathrm{MS}}) + 1.85$, where N_{MS} is the number of main-sequence stars. Using this fit, we estimated the $M_{50,mass}(\tau = 8)$ for the 228 clusters in our catalog for which star count data are available. The mean value of their $M_{50,mass}(\tau = 8)$ is 2.52.

6. DISCUSSION

6.1. *Comparison to Existing SMC Cluster Catalogs*

6.1.1. *Primary Catalog Comparisons: Bica et al. (2020)*

We compared our catalog of SMC star clusters with that of Bica et al. (2020) based on their coordination, using a search radius of 30 arcsec. Bica et al. (2020) recorded 627 clusters within SMC and Magellanic Bridge, of which our survey area covers 558 of them. Among these 558 star cluster objects, 477 were identified by at least one LGCS user. Applying the

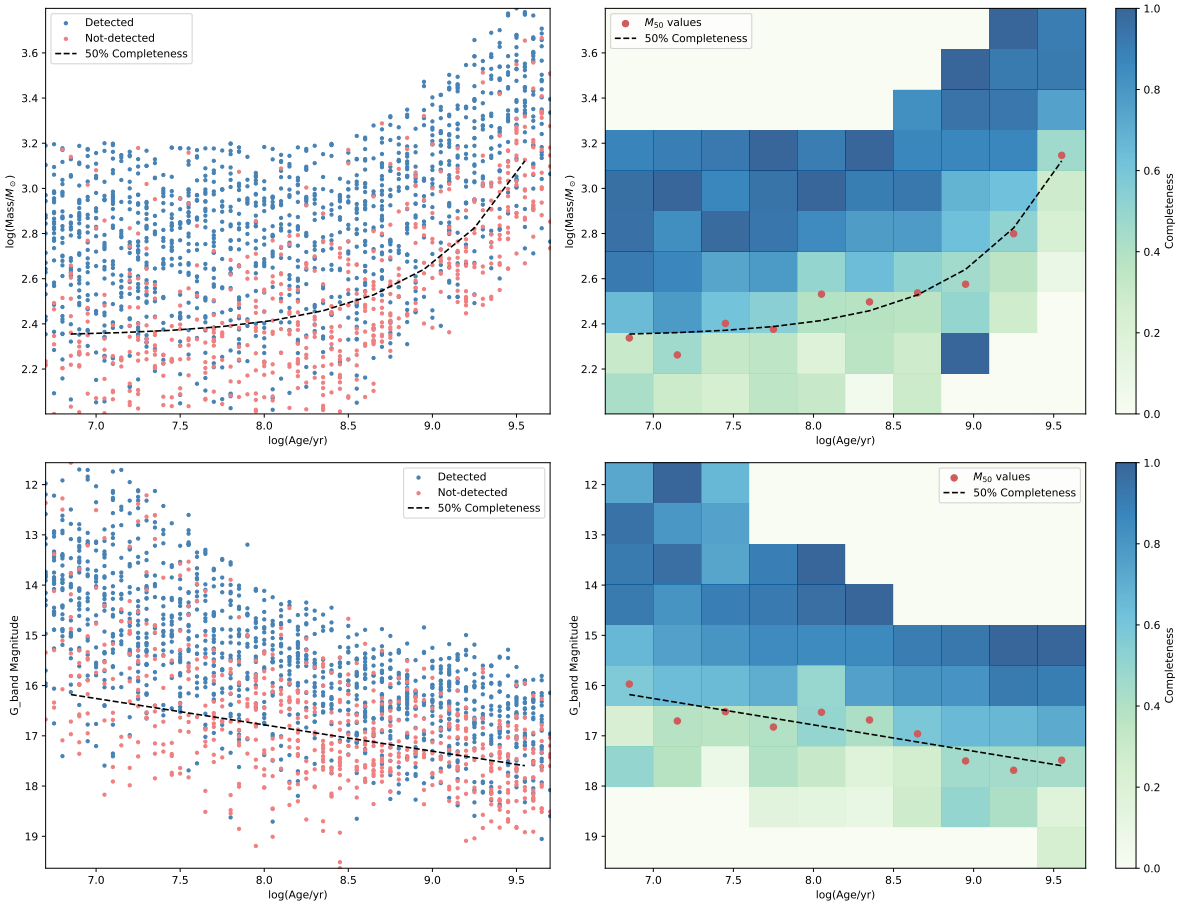

Figure 1. Completeness results based on the analysis of synthetic clusters. The top-left panel shows the identification results for all synthetic clusters, plotted by mass and age. The bottom-left panel plots identification results by G-band magnitude and age. In both panels, blue dots represent detected clusters ($f_{cluster} \geq 0.561$), while red dots represent non-detections ($f_{cluster} < 0.561$). The right panels present these identification results in 2D binned histogram. The fraction of detection of each bin is indicated by the shading. Orange dots in right panels represent the fitted M_{50} values derived from logistic function fit. The dashed black lines present the 50% completeness curves.

criterion $f_{cluster} \geq 0.561$, our catalog recovered 231 clusters from Bica et al. (2020).

Figure 4 shows the distribution of accepted and rejected star clusters from different sources cited by Bica et al. (2020). In this figure, the total length of each bar represents the total number of star clusters reported by this author that fall within our survey area (note that most clusters are identified by multiple authors, so the sum of clusters from each source exceeds 558). The light blue segments indicate clusters within the survey area that were not identified by any LGCS user. The dark blue segments represent clusters that were identified and met the criterion $f_{cluster} \geq 0.561$, while the red segment represents clusters that were identified but have $f_{cluster} < 0.561$, and were thus rejected from our catalog.

From Figure 4, it is shows that most of the rejected clusters were from OGLE II (Pietrzynski & Udalski 1999), H86 (Hodge 1986), and BS (Bica & Schmitt 1995). This outcome is expected, as these observations were conducted during the late photographic plate era or early CCD era. Before that time, observational resolution was too low to identify anything other than the most prominent clusters, which explains the full acceptance rate of the oldest catalogs Kron (1956), Lindsay (1958). Decades later, improved telescopes allowed for the detection of fainter and smaller clusters. However, the limited resolution still made it challenging to distinguish real clusters from false detections caused by random stellar overdensities. This issue improved with advancements in instrumentation, reflected in the low rejection rate of RZ (Rafelski & Zaritsky 2005). Note that although OGLE IV (Sitek et al. 2017) is one of the newest surveys, its coverage is limited to the noisy outer regions of the SMC, as central areas were already covered by OGLE II. This explains the high rejection rate of its cluster catalog.

7. SUMMARY

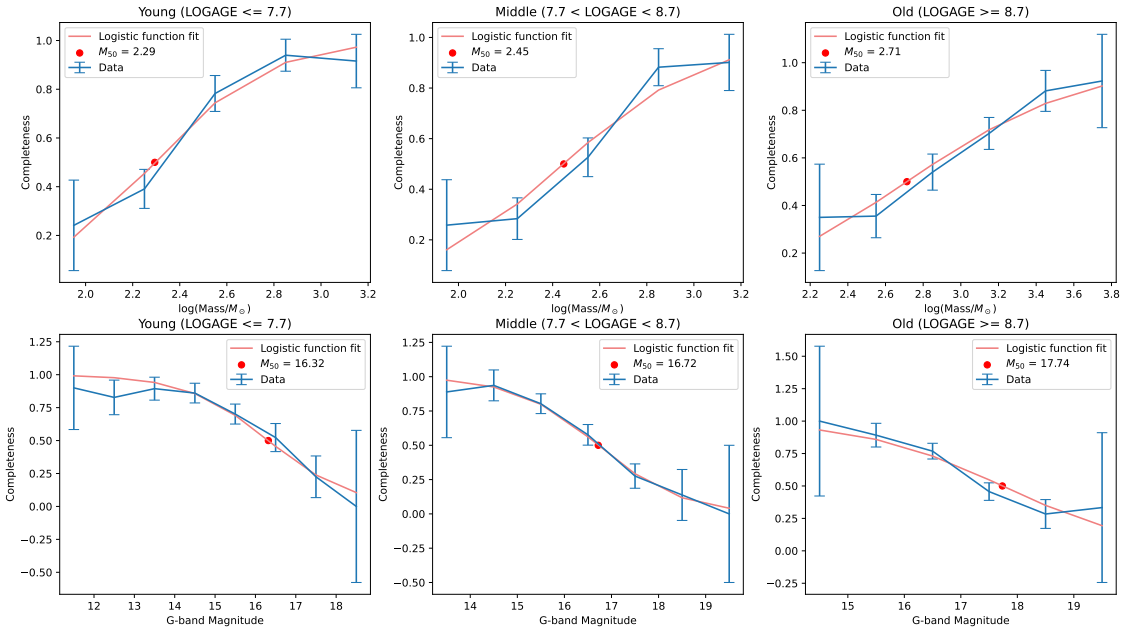

Figure 2. Logistic function fit example. Each column represents an age bin. In top panels, completeness of each age bin was fitted as a function of mass. In bottom panels, completeness of each age bin was fitted as a function of G-band magnitude. The blue lines present the calculated completeness of each mass bin or magnitude bin, along with their errors. The light red lines represent the logistic curve fit results. The red dots indicate the M_{50} values derived from the logistic function fit.

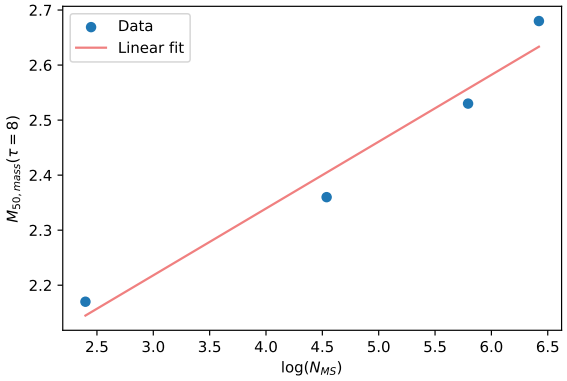

Figure 3. A plot of 50% complete mass at log(age/yr) = 8 against the number of main sequence stars in each synthetic cluster. The blue dots show the fitted values of $M_{50,mass}(\tau = 8)$ as derived from Equation 2 against the median value of N_{MS} for each synthetic cluster group. The red line shows the linear fit of these data points.

REFERENCES

Bica, E., Bonatto, C., Dutra, C. M., & Santos, J. F. C. 2008, MNRAS, 389, 678

Bica, E., Westera, P., Kerber, L. d. O., et al. 2020, AJ, 159, 82

Bica, E. L. D., & Schmitt, H. R. 1995, ApJS, 101, 41

Bitsakis, T., Bonfini, P., González-Lópezlira, R. A., et al. 2017, ApJ, 845, 56

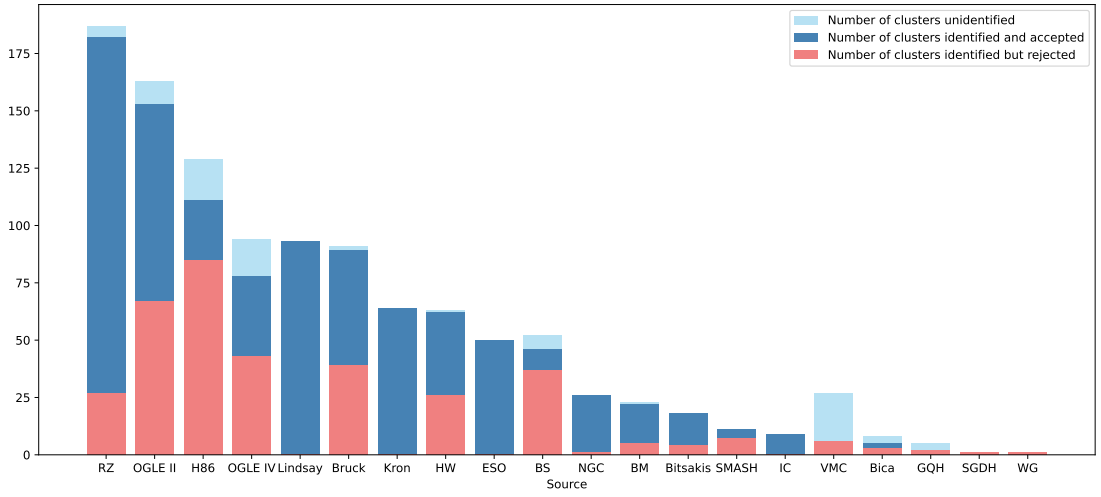

Figure 4. The ratio of undetected, accepted, and rejected clusters identified by authors cited in the Bica et al. (2020) catalog. The total bar length represents the number of clusters reported by each author within our survey area. The light blue segment indicates clusters that were not identified, the blue segment represents clusters identified and accepted by our catalog ($f_{cluster} \geq 0.561$), and the red segment represents clusters identified but rejected ($f_{cluster} < 0.561$). Note that most clusters were identified by multiple sources, thus the sum of numbers of clusters exceeds 558, the number of Bica et al. (2020) clusters within our survey area.

Bitsakis, T., González-Lópezlira, R. A., Bonfini, P., et al. 2018, ApJ, 853, 104

Hodge, P. 1986, PASP, 98, 1113

Johnson, L. C., Wainer, T. M., Torresvillanueva, E. E., et al. 2022, ApJ, 938, 81

Kron, G. E. 1956, PASP, 68, 125

Lindsay, E. M. 1958, MNRAS, 118, 172

Piatti, A. E. 2018, MNRAS, 478, 784

Pietrzynski, G., & Udalski, A. 1999, AcA, 49, 433

Rafelski, M., & Zaritsky, D. 2005, AJ, 129, 2701

Sitek, M., Szymański, M. K., Udalski, A., et al. 2017, AcA, 67, 363