



The Panchromatic Hubble Andromeda Treasury: Triangulum Extended Region (PHATTER). IV. Star Cluster Catalog

L. Clifton Johnson¹ , Tobin M. Wainer^{1,2} , Estephani E. TorresVillanueva^{2,3} , Anil C. Seth² , Benjamin F. Williams⁴ ,
Meredith J. Durbin⁴ , Julianne J. Dalcanton^{4,5} , Daniel R. Weisz⁶ , Eric F. Bell⁷ , Puragra Guhathakurta⁸ ,
Evan Skillman⁹ , and Adam Smercina⁴

PHATTER Collaboration

¹ Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA) and Department of Physics and Astronomy, Northwestern University, 1800 Sherman Avenue, Evanston, IL, 60201, USA; lcj@northwestern.edu

² Department of Physics and Astronomy, University of Utah, Salt Lake City, UT, 84112, USA

³ Department of Astronomy, University of Wisconsin-Madison, Madison, WI, 53706, USA

⁴ Department of Astronomy, University of Washington, Box 351580, Seattle, WA, 98195, USA

⁵ Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY, 10010, USA

⁶ Department of Astronomy, University of California, Berkeley, CA, 94720, USA

⁷ Department of Astronomy, University of Michigan, 323 West Hall, 1085 S. University Avenue, Ann Arbor, MI, 48105, USA

⁸ University of California—Santa Cruz, 1156 High Street, Santa Cruz, CA, 95064, USA

⁹ Minnesota Institute for Astrophysics, 116 Church Street SE, Minneapolis, MN, 55455, USA

Received 2022 January 21; revised 2022 August 15; accepted 2022 August 21; published 2022 October 14

Abstract

We construct a catalog of star clusters from Hubble Space Telescope images of the inner disk of the Triangulum Galaxy (M33) using image classifications collected by the Local Group Cluster Search, a citizen science project hosted on the Zooniverse platform. We identify 1214 star clusters within the Hubble Space Telescope imaging footprint of the Panchromatic Hubble Andromeda Treasury: Triangulum Extended Region (PHATTER) survey. Comparing this catalog to existing compilations in the literature, 68% of the clusters are newly identified. The final catalog includes multiband aperture photometry and fits for cluster properties via integrated light spectral energy distribution fitting. The cluster catalog's 50% completeness limit is $\sim 1500 M_{\odot}$ at an age of 100 Myr, as derived from comprehensive synthetic cluster tests.

Unified Astronomy Thesaurus concepts: Star clusters (1567); Triangulum Galaxy (1712); Catalogs (205)

Supporting material: machine-readable tables

1. Introduction

For decades, star clusters have been recognized as useful tracers of star formation. Rather than representing uniform tracers of star formation, we now understand that non-embedded, long-lived, gravitationally bound star clusters emerge from natal regions with relatively high gas surface density and star formation efficiency (Elmegreen & Efremov 1997; Kruijssen 2012; Grudić et al. 2021). Star clusters contain $\sim 5\%$ – 30% of the stellar mass formed (Johnson et al. 2016; Adamo et al. 2020) and are long-lasting remnants of peaks in hierarchically structured star-forming regions that survived the stellar feedback and gas removal dissolution that unbind most stellar groupings and natal structures.

We now have broad samples of clusters spanning a wide variety of galactic environments, where long-lived bound star clusters are tell-tale tracers of past episodes of intense, efficient star formation. Large imaging surveys using the Hubble Space Telescope (HST) have made significant progress in cataloging and characterizing star cluster populations in nearby (3–30 Mpc) galaxies (e.g., LEGUS, PHANGS-HST; Adamo et al. 2017; Lee et al. 2022). While the diversity of galactic environments included in these samples is very useful for purposes of galaxy-to-galaxy comparisons, individual star clusters at these distances

are only marginally resolved, limiting observational measurements to integrated properties.

In contrast, studies of neighboring galaxies in the Local Group provide a unique opportunity for detailed studies of external galaxies and their star clusters, yielding a rich picture of star formation observed at and below molecular cloud spatial scales—a level of detail not possible in more distant extragalactic targets. Due to their proximity and the spatial resolving power of HST, Local Group galaxies provide an unmatched opportunity to construct high-quality cluster catalogs and make detailed observations of these systems and their environments (e.g., Johnson et al. 2012). Local Group cluster catalogs reach low cluster mass completeness limits leading to increased sample sizes and diversity. Observations at these distances resolve individual cluster member stars leading to marked improvements in age-dating precision and usefulness to stellar evolution studies (Johnson et al. 2016; Girardi et al. 2020).

This paper studies the star cluster population of the Triangulum Galaxy (M33), whose intermediate galaxy mass and relatively active star formation provides a point of comparison to studies of the stellar cluster populations of the Andromeda galaxy (M31) and the Magellanic Clouds. Notably, M33 hosts a larger star formation rate (SFR) surface density (Σ_{SFR}) than the bulge-dominated, relatively quiescent M31 (Williams et al. 2021). Therefore, we expect observations of M33's young cluster population to unlock valuable new insights into star cluster formation and evolution. Triangulum's



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

relatively face-on orientation (inclination angle of 55° ; Koch et al. 2018) also presents an advantage over Andromeda in terms of line-of-sight dust attenuation and projection effects.

Studies of the star cluster population in M33 include results from both ground-based (e.g., Christian & Schommer 1982; San Roman et al. 2010) and space-based (e.g., Chandar et al. 1999, 2001; Park & Lee 2007; San Roman et al. 2009) observations, as summarized by Sarajedini & Mancone (2007). Previous work demonstrated HST’s utility for identifying star clusters, but the spatial coverage of M33’s star-forming disk was sparse and largely noncontiguous, preventing systematic studies of the cluster population. As a result, much of the past work to characterize M33’s star clusters made use of ground-based imaging and photometry (e.g., de Meulenaer et al. 2015; Fan & de Grijs 2014), especially from the Local Group Galaxy Survey (LGGS; Massey et al. 2006).

We note that an alternative catalog of M33 young star cluster candidates was published by Sharma et al. (2011) based on mid-infrared Spitzer $24\ \mu\text{m}$ source identification. This catalog should be sensitive to embedded clusters that are not detected by an optical search, and has been used for analysis of M33 clusters by a number of groups (e.g., González-Lópezlira et al. 2012; Pflamm-Altenburg et al. 2013; Corbelli et al. 2017). However, significant concerns about this catalog’s contamination by non-cluster objects and its suitability for star cluster studies have been well articulated by Sun et al. (2016). HST observations will have comparatively limited sensitivity to the earliest embedded stages of star cluster formation (1–3 Myr), but its high spatial resolution ($0''.1$ versus $\sim 6''$ for Spitzer $24\ \mu\text{m}$ images) remains the best avenue for the identification and analysis of star clusters at nearly every other age.

The Panchromatic Hubble Andromeda Treasury: Triangulum Extended Region survey (PHATTER; Williams et al. 2021) of M33 delivers contiguous, multiband imaging of a majority of the galaxy’s star-forming disk, extending the same quality of data obtained by the Panchromatic Hubble Andromeda Treasury survey (PHAT; Dalcanton et al. 2012) of M31 to M33. Similarly, this work moves cluster studies in M33 into a new era using techniques and analysis that were employed to construct the PHAT cluster catalog (Johnson et al. 2015; hereafter J15) for M31.

Facing the absence of a robust algorithmic method for identifying clusters in Local Group galaxy images, we launched an online citizen science project, the Local Group Cluster Search (LGCS), to perform a visual search of the PHATTER data. We employ the crowdsourced methodology developed for PHAT and the Andromeda Project (J15) to construct a star cluster catalog. This approach improves on the subjectivity of expert-led searches conducted in the past for M33 (e.g., Christian & Schommer 1982; Chandar et al. 1999) using a “wisdom of the crowds” consensus classification technique, where an unbiased, repeatable result is obtained by averaging over tens of independent image classifications. In addition to producing a robust cluster catalog, we characterize catalog completeness using synthetic clusters inserted into the search images. Not only does this methodology produce useful results, but it facilitates meaningful engagement with project volunteers regarding astronomy and star cluster science.

In this paper, we present the survey-wide cluster catalog for PHATTER. We describe the LGCS project, its input data and preparation, and data collection results in Section 2. We analyze image classifications and outline the steps required to

produce a cluster catalog in Section 3. We present the final catalog in Section 4, followed by a characterization of the catalog’s completeness in Section 5. We derive integrated light ages and masses in Section 6 and place the new catalog in context with previous work in M33 and similar work in M31 in Section 7.

This catalog serves as the foundation for PHATTER survey cluster science. Future work includes the measurement of the cluster mass function (Wainer et al. 2022), measurement of the high-mass stellar initial mass function, calibration of stellar evolution models, and more. These studies will build upon and benefit from comparisons to the PHAT star cluster studies of M31, including measurements of the star cluster formation efficiency (Johnson et al. 2016), the cluster mass function (Johnson et al. 2017), and the high-mass stellar initial mass function (Weisz et al. 2015).

Throughout this work, we assume a distance to M33 of 859 kpc (distance modulus = 24.67; de Grijs & Bono 2014) where 1 arcsec is equivalent to ~ 4.2 pc.

2. Data

In this section, we describe the LGCS citizen science project and the underlying HST data that enable this study. We begin by describing the PHATTER imaging used for the project (Section 2.1) and the LGCS website interface (Section 2.2). Next, we discuss data collection and statistics regarding image classifications and project volunteers (Section 2.3). Finally, we discuss the creation of synthetic clusters used to characterize catalog completeness (Section 2.4).

2.1. PHATTER Images and Resolved Star Photometry

The HST images analyzed by the LGCS project were obtained as part of the PHATTER survey. Full details of the survey are presented in Williams et al. (2021), but here we highlight the features of this survey that are relevant to star cluster catalog work.

The PHATTER survey uses the same imaging strategy as the PHAT survey, where parallel observations are efficiently obtained with the Advanced Camera for Surveys (ACS) and Wide Field Camera 3 (WFC3). These observations are organized into three contiguous “bricks”, a 3×6 mosaic of WFC3 footprints formed from pairs of parallel ACS and WFC3 images that combine to create a rectangular region of fully overlapped spatial coverage in all observed passbands. This observing strategy yields images in six filters: F475W and F814W in the optical obtained with ACS/WFC; F275W and F336W in the near-UV obtained with WFC3/UVIS; and F110W and F160W in the near-IR (NIR) obtained with WFC3/IR. The PHATTER survey’s three bricks (54 individual fields of view) span the inner disk of M33, extending out to a galactocentric radius of ~ 4 kpc. We use three types of image products from the survey. First, drizzled single-pointing ACS images were used to create optical images with synthetic clusters inserted to minimize computational effort (see Section 2.4). Second, brick-wide mosaic images for each of the six filters were used for aperture photometry of the clusters, which provide the best overlapping spatial coverage and artifact removal (i.e., chip gaps and cosmic rays). Third, LGCS search images (see Section 2.2) were extracted from survey-wide optical mosaic images. All images have an image scale of $0''.05\ \text{pixel}^{-1}$ and are astrometrically aligned to Gaia DR2 with 3 mas (7 mas) residuals for ACS/WFC and WFC3/

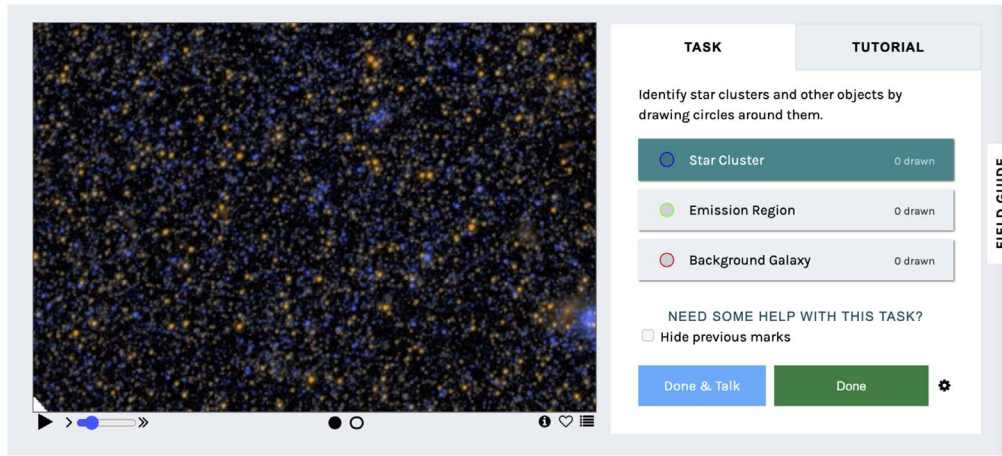


Figure 1. Screenshot of the LGCS annotation interface.

UVIS (WFC3/IR), and are combined and distortion corrected using AstroDrizzle from the DrizzlePac package (STSCI Development Team 2012; Hack et al. 2013; Avila et al. 2015).

In addition to the images, we also use PHATTER resolved star photometry catalogs presented by Williams et al. (2021). This point-spread function photometry was measured simultaneously in all six filters using DOLPHOT,¹⁰ an updated version of the HSTphot photometry package (Dolphin 2000).

We use the PHATTER photometry catalogs to quantify and map stellar density across the survey footprint. Specifically, we define and use two quantities: N_{MS} , the number of upper main-sequence stars selected using a color-magnitude cut of $F475W < 24$ and $F475W - F814W < 1$; and N_{RGB} , the number of bright RGB stars defined using a polygon region in the optical color-magnitude diagram (CMD) that mimics a NIR-based selection used in M31 by J15, where $F475W - F814W > 1.5$ and $F814W$ brighter than ~ 22.5 .

2.2. LGCS Interface

The Local Group Cluster Search (LGCS)¹¹ is a citizen science project built and hosted on the Zooniverse¹² platform. The project is a direct follow-on of the Andromeda Project (J15), but was built using the Zooniverse’s Project Builder¹³ platform tools rather than being built as a custom project-specific website. The Project Builder platform allowed the research team to build and configure the project without the effort or assistance of the Zooniverse web development team, though LGCS has fewer custom features than the Andromeda Project (e.g., no interactive walk-through of interface tools during the project tutorial). The main capabilities provided by the Zooniverse platform, however, remain the same: an interactive user interface that enables image annotation, web hosting for the project page and image data, subject image selection and queuing, feedback confirming the correct identification of synthetic clusters in the images (see 2.4), and the storage of classification responses.

The scope of the LGCS project extends beyond the search of PHATTER imaging presented here; the project hosts a visual

cluster search of SMASH (Nidever et al. 2017) imaging of the Large and Small Magellanic Clouds, and will expand to additional data sets in the future. This study focuses solely on the results of the PHATTER M33 search, while results from other LGCS searches will be published separately in future work.

Volunteers who participate in the LGCS search of the PHATTER data use a simple annotation interface to mark objects of interest, as shown in Figure 1. Specifically, participants are asked to mark star clusters, background galaxies, and nebulous emission regions using a circular tool with adjustable size. Users click the center of an object in the image and drag outward to set a circular marker’s radial size. Two images are shown to participants: a color image constructed from the F475W and F814W bands, and an inverted grayscale F475W image. The two-band composite image provides important color information, while the inverted singleband image provides high contrast to improve the detection of faint clusters. Individual $\sim 36'' \times 25'' (\sim 150 \times 100 \text{ pc})$ subimages are extracted from survey-wide drizzled mosaic images. These subimages spatially overlap by 100 pixels ($5''$) to minimize edge effects on the search results.

In addition to the image annotation interface, the LGCS project also hosts helpful resources for volunteers. Participants are presented a tutorial for instruction about the task during their first visit to the classification page, which includes a short demonstration video and text instructions. Volunteers can also access a number of “About” pages that summarize the project’s research goals and background information, a field guide that provides detailed examples of target objects, and a forum called “Talk” to facilitate interactions between with the research team and project participants.

2.3. Data Collection and Classification Statistics

The LGCS project collected image classifications from 2019 January 8 to 2019 February 28. During this time, LGCS volunteers submitted 269,645 classifications, where one classification denotes a participant’s response specifying the location and size of any objects they identify in an image. Each search image was classified by at least 60 unique volunteers.

A total of 2757 users participated in the LGCS project, 1517 of whom participated as registered users using a Zooniverse account. Contributions from individuals who were not logged in are identified and grouped via an anonymized hash of the

¹⁰ <http://americano.dolphinsim.com/dolphot/>

¹¹ <https://www.clusterssearch.org/>

¹² <https://www.zooniverse.org/>

¹³ <https://www.zooniverse.org/lab>

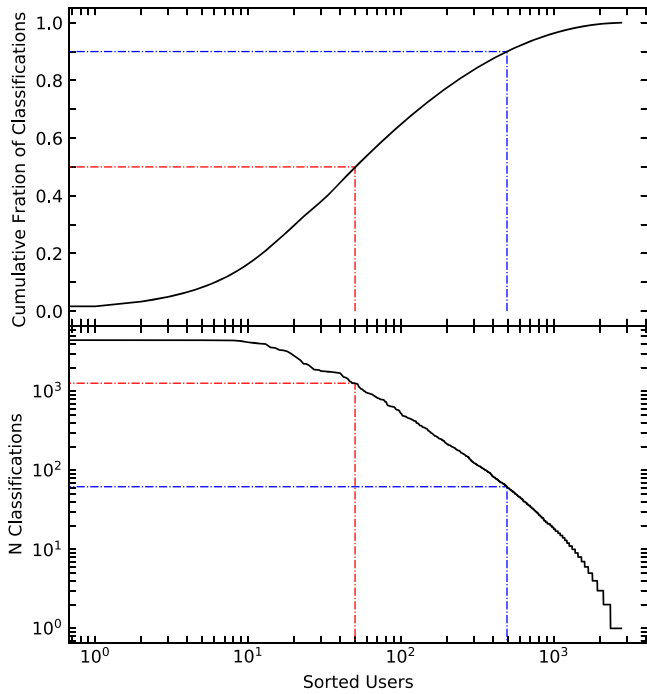


Figure 2. Classification statistics of individual LGCS users, sorted in decreasing order of the total submitted classifications. Top: the cumulative fraction of the LGCS’s $\sim 269,000$ classifications. Bottom: the number of classifications submitted by the N th ranked volunteer. The red dashed–dotted lines illustrate that 50% of all the classifications were performed by the 50 most active volunteers, who each contributed at least 1267 classifications. The blue dashed–dotted lines illustrate that 90% of all classifications were performed by 463 volunteers who each contributed at least 62 classifications.

participant’s IP address. We note that while unregistered users represent 45% of the volunteers by number, these users only contributed 8% of the total classifications. Beyond this difference between registered and unregistered users, the distribution of effort across the pool of participants varies significantly, as shown in Figure 2. The median number of classifications per person is nine (19 for registered users), but half of all the classification effort was provided by the 50 most active volunteers—each of whom contributed at least 1267 classifications. This behavior is consistent with trends seen for the Andromeda Project (J15) and other Zooniverse projects (Spiers et al. 2019), where a relatively small group of active volunteers contribute a bulk of the total effort.

2.4. Synthetic Cluster Generation

A key part of our analysis is incorporating synthetic clusters with known ages, masses, and radii into the visual search. Synthetic clusters were inserted into the same LGCS images used for the cluster search, but were analyzed separately. The synthetic clusters were created following a procedure similar to the one used by J15. In short, a synthetic cluster’s individual member stars were drawn with masses following a Kroupa (2001) initial mass function (IMF) and stellar properties for a specified age using PARSEC 1.2S + COLIBRI PR16 isochrones (Bressan et al. 2012; Marigo et al. 2017). Spatial positions were drawn from a King (1962) profile with a specified effective radius, R_{eff} , and a concentration ($R_{\text{tidal}}/R_{\text{core}}$) of 30.

The sample of synthetic clusters were created with the following properties:

1. Ages were drawn randomly from a grid of $\log(\text{Age}/\text{yr})$ values ranging from 6.6 to 10.1 incremented every 0.05 dex.
2. Masses were drawn randomly from a continuous uniform distribution of $\log(\text{Mass}/M_{\odot})$ values ranging from 2.0 to 5.0.
3. A fixed Solar metallicity ($Z = 0.0152$) was assumed for ages younger than 5 Gyr. For older ages, the metallicity was randomly drawn from a set of five discrete values ($Z = [0.0152, 0.005, 0.0015, 0.0005, 0.00015]$) so that the sample of older synthetic clusters would span a metallicity range resembling that of Galactic and extragalactic globular clusters.
4. Extinctions were drawn from an exponential A_V distribution which ranges from the foreground Milky Way extinction value of 0.11 to 3.0 mag, following the expression $P(A_V) \propto e^{-(A_V/1.34)}$; this is the same distribution used for M31 synthetic clusters by J15.
5. Effective radii (R_{eff}) were drawn from the distribution of measured values obtained for M31 clusters by J15, but biased to larger R_{eff} values to ensure sufficient number statistics of diffuse clusters in the high- R_{eff} tail for completeness determination purposes.

After creating a parent population of artificial clusters, we chose a subset to insert into the LGCS images. This selection was based on cluster magnitude and age, and it was designed to produce a sample of synthetic clusters that spans the full range of detectability, from easily detected to undetectable. Specifically, we adopt the following magnitude limits: $18.5 < m_{\text{F475W}} < 22$ for $6.6 < \log(\text{Age}/\text{yr}) < 8.0$; $19.5 < m_{\text{F475W}} < 22.5$ for $8.0 < \log(\text{Age}/\text{yr}) < 9.0$; and $20 < m_{\text{F475W}} < 22.5$ for $9.0 < \log(\text{Age}/\text{yr}) < 10.0$.

We inserted the magnitude-selected sample of synthetic clusters into the F475W and F814W images using DOLPHOT. One synthetic cluster was added per LGCS search image, positioned pseudorandomly within the image, avoiding positions within 120 pixels of the edge. Because DOLPHOT places the synthetic clusters into each individual frame before drizzling, we use single-field images as opposed to multi-image mosaics for insertion to minimize computational complexity. The insertion locations were chosen to avoid chip edges and gaps to ensure that the synthetic images were essentially identical to the original search images.

We created two batches of synthetic clusters, each with 848 objects for a total of 1696 synthetic clusters. The first batch was randomly assigned to LGCS images spanning the entire PHATTER survey footprint, resulting in a diverse set of cluster-image pairs across the full range of galactic environments. The second batch was assigned spatial locations in a targeted manner, such that young clusters (< 100 Myr) were placed in regions of the footprint with a high density of bright, blue stars, as defined by their high N_{MS} values ($N_{\text{MS}} > 1200$ stars per search image). The remaining older synthetic clusters were distributed across the remaining fields with lower N_{MS} values. The targeted placement of this second batch ensures that sufficient numbers of young synthetic clusters fall within young star-forming regions, safeguarding our ability to derive catalog completeness for the key population of young star clusters.

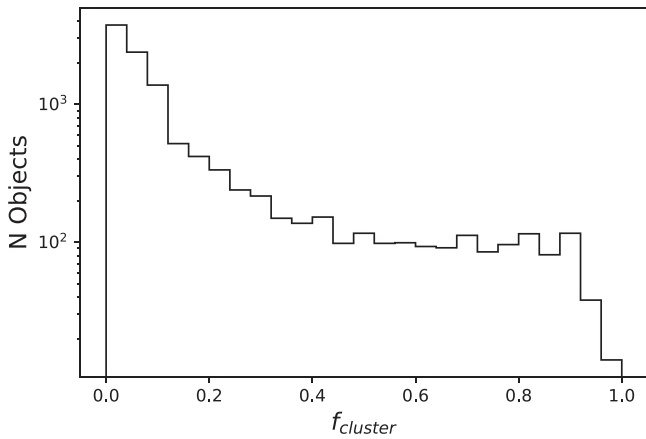


Figure 3. Distribution of f_{cluster} values for our full sample of cluster candidates.

3. Catalog Construction

The process of converting LGCS image classifications into a star cluster catalog involves combining 60 independent classifications from each image into a consensus result regarding the presence, location, and size of candidate clusters (and other objects). This target number of classifications per image is selected as a balance of keeping the statistical errors of the classification results low and maintaining a reasonable total runtime for the project.

The first step in this process is to compile and combine the candidate identifications made by LGCS participants. We merge identifications following the procedure described in detail in Appendix A of J15: we aggregate markings for each individual search image by clustering marker centers and merging overlapping candidates, then we combine the per-image lists of identifications into a survey-wide data product by running a spatial match to merge duplicate candidates in regions of overlapping image coverage.

We use the fraction of classifications where the object is detected as the principal indicator of significance. We compute four fractional metrics to characterize each candidate:

1. f_{view} is the fraction of total classifications where a candidate is identified as any class of object.
2. f_{cluster} is the fraction of total classifications where a candidate is identified as a star cluster.
3. f_{galaxy} is the fraction of total classifications where a candidate is identified as a background galaxy.
4. f_{emission} is the fraction of total classifications where a candidate is identified as an emission region.

These quantities are related by:

$$f_{\text{view}} = f_{\text{cluster}} + f_{\text{galaxy}} + f_{\text{emission}}. \quad (1)$$

We note that the definitions of these metrics differ slightly from those used by J15, such that all four metrics are normalized by the total number of available image classifications.

The aggregation process produced a set of 10,926 unique identifications. This total number includes many low-significance objects, with only 4780 candidates having $f_{\text{view}} \geq 0.1$. For this paper, we focus primarily on the cluster candidates; please see Appendix A for a discussion of the background galaxy and emission region results.

We present a histogram of f_{cluster} values in Figure 3. Our visual inspection of cluster candidates confirmed that as f_{cluster}

decreases, the quality of the cluster candidates is lower. We find that for $f_{\text{cluster}} > 0.6$, there are very few contaminants; among 841 candidates, only one has $f_{\text{galaxy}} > 0.1$, and that object is eliminated by the subsequent weighted cuts (see Section 3.1).

3.1. User Weighting

While the f_{cluster} metric assumes that each volunteer is equally skilled at identifying star clusters, multiple citizen science projects (e.g., Willett et al. 2013; Jayasinghe et al. 2019; Eisner et al. 2021) have found that weighting volunteer responses according to their task performance can increase sample quality and decrease catalog contamination as a function of completeness. To maximize the usefulness of LGCS volunteer contributions, we follow the methodology of J15 and weight classifications based on the volunteers' performances in identifying star clusters. In this section, we demonstrate that employing user weighting significantly improves the resulting cluster catalog.

We calculate user weights based on the agreement between the classifications of the user and the entire set of LGCS participants. We use two separate weights: one for detections, objects the user identified; one for non-detections, objects the user did not identify. A user's detections are weighted according to the average f_{cluster} of all cluster identifications made by the individual, such that classifications made by those who tend to identify good candidates with higher f_{cluster} are assigned greater weights than those who identify worse candidates with lower f_{cluster} . A user's cluster non-detections are weighted according to the fraction of high-quality clusters ($f_{\text{cluster}} > 0.6$) that the user sees and detects, such that classifications by those who rarely miss good clusters carry greater weight than those who are more selective and identify fewer clusters.

We note that volunteer classifications of synthetic cluster images are omitted from user metric calculations. This omission ensures that user classification metrics are based only on real data, and that catalog completeness is not biased due to user weighting.

We calculate two user metrics for all volunteers to quantify the detection and non-detection behaviors described above: $\bar{f}_{\text{cluster}}$, the average f_{cluster} of all clusters a user identifies, and $f_{\text{consensus}}$, the fraction of high-quality ($f_{\text{cluster}} > 0.6$) clusters a user saw that they identified. Figure 4 shows these user metrics for all volunteers who contributed more than 20 classifications. Many users lie toward the upper right corner of the plot, representing users who excel in both user metrics. In contrast, users in the top left are conservative classifiers who identify clusters with high f_{cluster} , but miss a significant fraction of commonly identified clusters. Those in the bottom right are liberal classifiers who include all good clusters in their identifications, but at the expense of including lower-quality clusters as well.

We examine trends in user behavior by grouping users into bins according to their total classification count, with the average user metrics of each group plotted as colored stars in Figure 4. Volunteers with higher classification counts (redder points) tend to have higher user metrics scores, which may indicate that volunteers become more skilled on average as they classify an increasing number of images.

Because an individual's user metrics become noisy at small numbers of classifications, we replace the metrics of users with

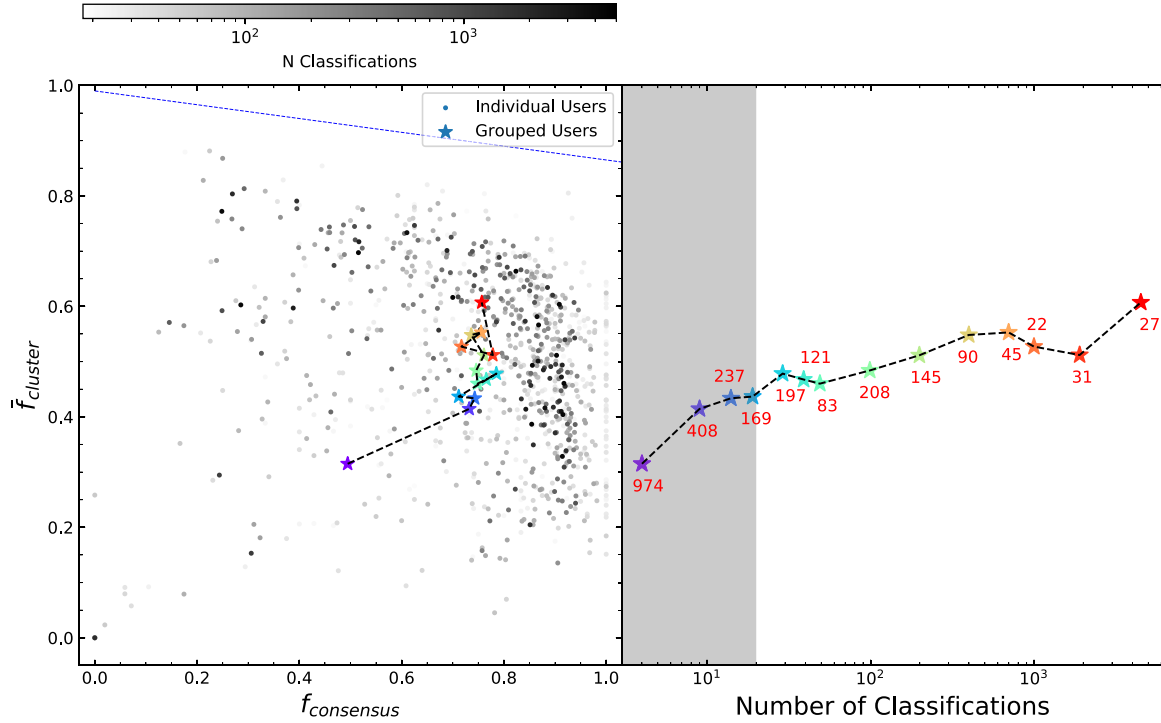


Figure 4. Left: $f_{\text{consensus}}$ and $\bar{f}_{\text{cluster}}$ user metric scores used for weighting. Circular points represent metrics for individuals with >20 classifications; point shading denotes the volunteer classification count, where darker colors highlight individuals with more classifications. Star markers represent the average metrics for a group of users, binned in groups according to classification count. The blue dashed line represents the maximum $\bar{f}_{\text{cluster}}$ a volunteer can achieve based on the fraction of high-quality consensus clusters identified. Right: star markers correspond to the same binned user groups presented in the left panel, but now showing the average number of classifications for each group on the x-axis. The red number next to each star is the number of users represented in each respective bin. For volunteers with <20 classifications (gray shaded region) users were weighted according to aggregate user metric values calculated for the binned group rather than using individual values.

≤ 20 classifications with the aggregate values of $\bar{f}_{\text{cluster}}$ and $f_{\text{consensus}}$ obtained for their binned groups. The use of aggregate metrics and weights for users with low classification counts has little impact on the catalog results due to the small percentage of total classifications contributed by these individuals (see Figure 2).

To convert a volunteer’s user metric results into a classification weight, we adopt a generalized logistic function:

$$W(x) = B \times \left(A + \frac{1}{1 + e^{-m_{\text{logistic}}(x - b_{\text{logistic}})}} \right), \quad (2)$$

where x represents $\bar{f}_{\text{cluster}}$ for detection weights and $f_{\text{consensus}}$ for non-detection weights, while m_{logistic} and b_{logistic} are the slope and position of maximum growth of the logistic curve, respectively. The coefficients A and B are normalization constants set such that W varies between 0 and 1 over the x interval $[0, 1]$. We seek to identify values of m_{logistic} and b_{logistic} for detection and non-detection weights that maximize cluster catalog completeness and minimize contamination.

To fit for an optimal weighting scheme, we first compile a set of “expert” ratings to use as a reference when computing completeness and contamination metrics for a given set of weighting parameters. A group of four coauthors visually inspected clusters and scored them on a scale of 1–3: 1 is a definite cluster, 2 is a possible cluster, and 3 is a non-cluster. The four coauthors ranked all marginal cluster candidates ($0.35 < f_{\text{cluster}} < 0.5$) where we expect the greatest variety in quality. Additionally, one coauthor ranked a broader range of candidates ($f_{\text{cluster}} \gtrsim 0.25$) to confirm that objects with $f_{\text{cluster}} < 0.35$ were low-quality identifications. The average of these ranks, S_{expert} , is then used to categorize clusters and

contaminants. Clusters with $S_{\text{expert}} < 1.5$ were declared good clusters, and $S_{\text{expert}} > 2.5$ were considered contaminants.

Using the expert ratings, we construct a completeness versus contamination curve for our unweighted sample by varying the f_{cluster} threshold from 0 to 1, as shown in Figure 5. We define the minimum distance from the curve to the lower left corner of this plot (i.e., an optimal sample with 100% completeness and no contaminants) as d_{optimal} , and use this metric to evaluate, rank, and optimize the adjustable weighting parameters.

We conduct an iterative, grid-based search for an optimal set of detection and non-detection weighting function parameters that minimize the d_{optimal} metric for the completeness versus contamination curve. For each grid point, we calculate detection and non-detection user weights from $\bar{f}_{\text{cluster}}$ and $f_{\text{consensus}}$ user metrics, respectively, using logistic functions with specified values of m_{logistic} and b_{logistic} parameter values to make the metric-to-weight transformation. We then compute user-weighted f_{cluster} values, $f_{\text{cluster},W}$, for each cluster candidate, construct a completeness versus contamination curve, and calculate the associated d_{optimal} .

We identify the set of logistic function parameters that minimize d_{optimal} , and thus produce an optimal cluster catalog that maximizes completeness and minimizes contamination. We find that the following weighting parameters produce the best weighted catalog: detection weight parameters of $(m_{\text{logistic}}, b_{\text{logistic}}) = (30.0, 0.45)$; non-detection weight parameters of $(m_{\text{logistic}}, b_{\text{logistic}}) = (30.0, 1.1)$. The optimally weighted catalog has significantly lower contamination as a function of completeness than the unweighted catalog, as shown in Figure 5, demonstrating that the application of user weights improved the quality of the cluster catalog we produced.

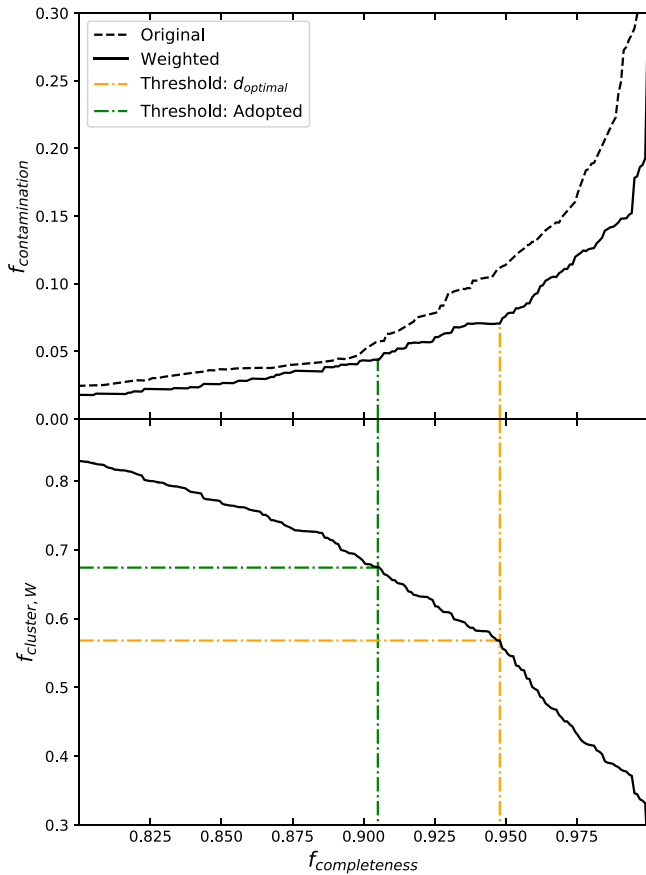


Figure 5. Top: catalog completeness vs. contamination for the original, unweighted sample (dashed black line), and the optimal user weighting system (solid black line). The orange dashed line represents the d_{optimal} threshold, where d_{optimal} is the distance from this point to the bottom right corner of the plot. The green line shows the adopted threshold chosen to reduce catalog contamination. Bottom: catalog completeness vs. $f_{\text{cluster},W}$, which provides reference to where each threshold is drawn.

3.1.1. Catalog Threshold Selection

With the user weighting parameters fixed, we move on to selecting a catalog threshold. J15 choose a $f_{\text{cluster},W}$ cutoff that corresponds to the point on the completeness versus contamination curve where d_{optimal} is minimized. The minimum d_{optimal} point corresponds to a $f_{\text{cluster},W}$ threshold of 0.568, 94.8% completeness, and 7.1% contamination, indicated by the orange lines in Figure 5.

We note two key differences between the J15 and LGCS completeness versus contamination curves: $f_{\text{contamination}}$ values for LGCS are smaller by approximately a factor of 2, and the original unweighted curve (and the weighted curve to a lesser degree) shows a distinct change in slope behavior at $f_{\text{completeness}} \sim 0.9$ in Figure 5. In addition, a qualitative evaluation of the d_{optimal} -based threshold concluded that the resulting cluster sample includes a higher number of contaminants than desired, leading us to reevaluate our choice of $f_{\text{cluster},W}$ threshold.

Based on the poor assessment of the initial $f_{\text{cluster},W}$ threshold, we seek an alternative, more conservative catalog limit. We target a greater $f_{\text{cluster},W}$ value that corresponds to a point on the weighted completeness versus contamination curve near the transition in slope at $f_{\text{completeness}} \sim 0.9$. We find that by applying a factor of 2 scaling to the $f_{\text{contamination}}$ component of the d_{optimal} distance calculation, motivated by the $\sim 2\times$ scaling difference between the LGCS and J15 curves, we

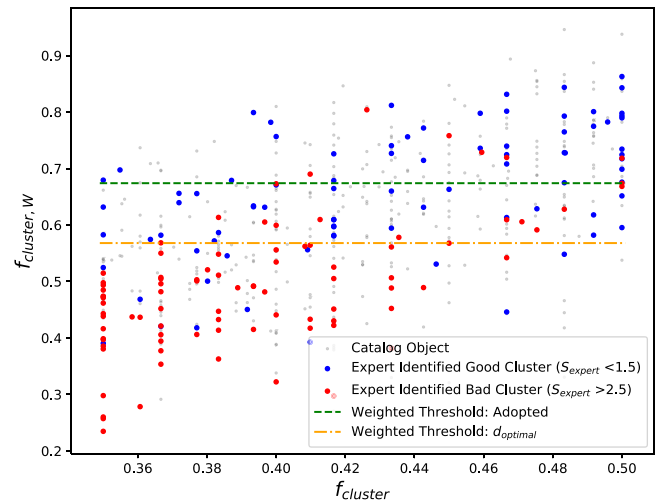


Figure 6. Weighted and unweighted f_{cluster} values are plotted for the subset of expert-ranked clusters. The blue dots are expert-identified good clusters, and the red dots are expert-identified contaminants. The dashed green line represents the $f_{\text{cluster},W}$ catalog threshold of 0.674, while the orange dashed line represents the $f_{\text{cluster},W}$ catalog threshold of 0.568. We prefer the higher adopted $f_{\text{cluster},W}$ threshold as it rejects a larger number of expert-identified bad cluster candidates, as seen by the larger number of red points that fall below the green horizontal line.

identify a viable threshold that meets all the above criteria. The resulting $f_{\text{cluster},W}$ catalog threshold is 0.674, which corresponds to 90.5% completeness and 4.4% contamination, indicated by the green lines in Figure 5.

We demonstrate the impacts of our weighting system in Figure 6 for the expert-classified subsample of cluster candidates with $0.35 \leq f_{\text{cluster}} \leq 0.5$ where weighting and threshold selection has the greatest impact. Expert-identified good clusters ($S_{\text{expert}} < 1.5$) are plotted in blue, and expert-identified bad clusters ($S_{\text{expert}} > 2.5$) are shown in red. This plot shows how the weighted f_{cluster} system is more effective at separating good candidates from bad candidates than the unweighted system, due to the improved separation of blue and red points by horizontal lines of constant $f_{\text{cluster},W}$ over vertical lines of constant f_{cluster} . We can also see that the number of bad candidates that are rejected by the higher, adopted $f_{\text{cluster},W}$ threshold is larger than the number rejected by the original, unscaled d_{optimal} -based $f_{\text{cluster},W}$ threshold, justifying our choice of the more conservative catalog threshold.

4. PHATTER Star Cluster Catalog

We apply the catalog construction techniques and user weighting described in Section 3 to derive a final cluster catalog. We present a sample of 1214 clusters that were selected using an $f_{\text{cluster},W}$ threshold of 0.674, which reflects a conservative selection of clusters that minimizes contamination by non-cluster candidates. Table 1 reports cluster position, radius, and classification metrics (including weighted $f_{\text{cluster},W}$) for each of our cataloged clusters. We show the spatial distribution of the clusters in Figure 7.

We also present information for an additional 3566 candidate identifications with $f_{\text{cluster},W} < 0.674$ and $f_{\text{view}} \geq 0.1$ in Appendix A, allowing catalog users to make alternative choices of catalog $f_{\text{cluster},W}$ thresholds based on the specific needs and requirements of a particular science use case.

We present an example cluster, PHATTER 22, in Figure 8 to illustrate the data available for the PHATTER cluster sample.

Table 1
PHATTER Cluster Catalog

ID	R.A. (J2000)	DECL. (J2000)	R_{ap} (")	R_{eff} (")	f_{view}	f_{cluster}	f_{galaxy}	f_{emission}	$f_{\text{cluster,W}}$	Flags	m_{apcor}	m_{275}	σ_{275}	m_{336}	σ_{336}	m_{475}	σ_{475}	m_{814}	σ_{814}	m_{110}	σ_{110}	m_{160}	σ_{160}
1	23.553754	30.479462	1.90	0.23	1.0000	0.9667	0.0000	0.0333	1.0000	...	−0.00	...	...	...	...	18.39	0.02	17.70	0.04	...	...	...	...
2	23.611876	30.696172	1.66	0.48	1.0000	1.0000	0.0000	0.0000	1.0000	...	−0.11	>21.01	...	>21.10	...	19.63	0.06	18.73	0.08	...	...	...	...
3	23.606064	30.699199	1.54	0.36	1.0000	1.0000	0.0000	0.0000	1.0000	...	−0.05	20.09	0.18	20.42	0.01	20.21	0.10	19.41	0.25	...	...	...	...
4	23.434333	30.514082	1.90	0.49	1.0000	0.9833	0.0000	0.0167	0.9999	...	−0.07	>21.12	...	>22.39	...	19.84	0.11	18.25	0.05	17.41	0.05	16.61	0.05
5	23.583846	30.659212	1.86	0.37	0.9836	0.9016	0.0000	0.0820	0.9772	...	−0.02	19.29	0.10	19.00	0.06	18.76	0.02	18.00	0.06	17.64	0.12	17.21	0.16

Note. Note that the R_{ap} parameter gives the median of the user-clicked radii, which we use as the aperture for the photometry measurements (see Section 4.1).
(This table is available in its entirety in machine-readable form.)

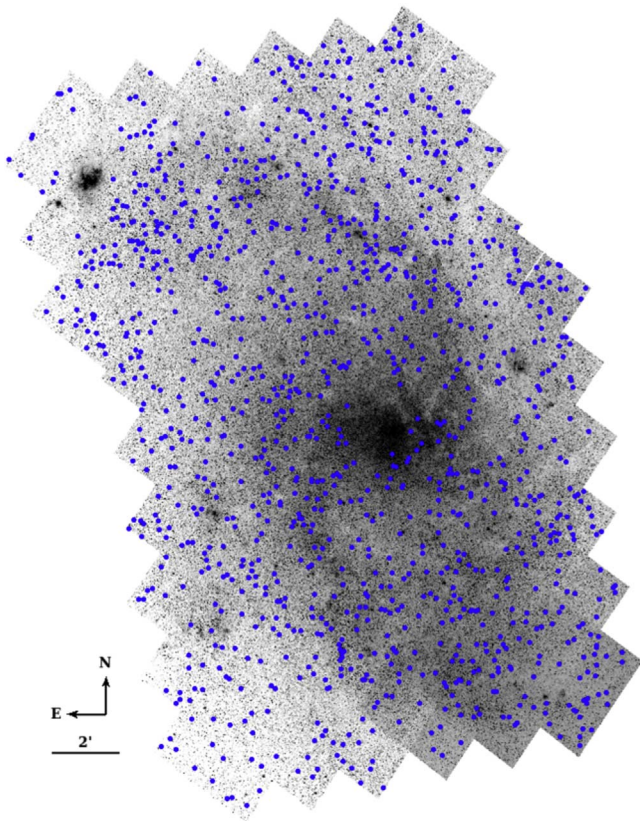


Figure 7. Spatial distribution of PHATTER star clusters (blue) overlaid on a F475W survey-wide mosaic image.

Optical cutout images display individual cluster identifications made by LGCS volunteers as well as the final cluster aperture for the cluster, and the images show that star clusters appear as collections of individually resolved members stars in the PHATTER imagery. The survey’s panchromatic images produce six-band spectral energy distributions (SEDs) from integrated light photometry (see Section 4.1) and cluster CMDs from resolved star photometry catalogs, both of which are used to fit for cluster properties (see Section 6 and Wainer et al. 2022).

4.1. Cluster Photometry

Aperture photometry for candidate clusters and other objects are measured using the same techniques and code employed by J15, which we summarize here. Photometric apertures for an object are centered at the mean position and extend to the median radius (R_{ap}) drawn by LGCS participants. The sky background is measured in 10 annuli, each with an area equal to that of the photometric aperture, extending from $1.2\text{--}3.4 R_{\text{ap}}$. These sky annuli facilitate robust determinations of the sky flux levels and associated uncertainties, which are important given that background variations often dominate the photometry error budget. Photometry is reported in the VEGAMAG system for the native HST bandpasses and calibrated using zeropoints obtained from relevant instrument websites for ACS¹⁴ (Bohlin 2016), WFC3/UVIS¹⁵ (2017 values; Deustua et al. 2017) and WFC3/IR¹⁶ (2012 values).

We note that these adopted zeropoints are consistent to within 1% (2%) to alternative “2020 values” for ACS and WFC3/UVIS (WFC3/IR).

Six-band integrated photometry for the final cluster sample is presented in Table 1, and equivalent photometry for ancillary cluster candidates is presented in Appendix A. Measured magnitudes are reported for detections with a signal-to-noise ratio (S/N) ≥ 3 and 3σ upper limits are reported for non-detections. Blank entries denote cases of incomplete image coverage in that photometric passband.

The radial light profile of each cluster is measured from the F475W image, and the half-light radius, R_{eff} , is derived through interpolation of the radial profile. Aperture corrections are computed assuming a King (1962) light profile with a fixed concentration ($c = R_{\text{tidal}}/R_{\text{core}} = 7$), scaled to match the measured R_{eff} . When applied to the aperture magnitudes,¹⁷ these corrections yield an estimate of the total cluster light that accounts for the flux that falls outside the photometric aperture, R_{ap} . The median correction is -0.04 mag and the 25th-to-75th interquartile range spans from -0.09 to -0.01 . We report the F475W R_{eff} measurements and aperture corrections in Table 1.

We flag objects with large R_{eff} ($\geq 0''.8$ or $\sim 3\text{pc}$) that are also bright ($m_{\text{F475W}} < 19.0$) and blue ($m_{\text{F336W}} - m_{\text{F475W}} < -0.5$) as possible associations. Eight such objects are identified in Table 1.

We plot a UV–optical color–color diagram for the PHATTER cluster sample in Figure 9. As expected, many of the clusters are quite blue ($\text{F336W} - \text{F475W} < 0.5$), indicating young ages. When compared to the color–color distribution of PHAT clusters from M31 (J15), we see the difference between the younger and bluer cluster population from M33 that hosts ongoing star formation versus the quiescent cluster population from M31 that hosts relatively larger numbers of globular clusters and other older clusters ($>1\text{--}3$ Gyr).

5. Catalog Completeness

We determine the completeness of our cluster sample by analyzing synthetic clusters inserted into the LGCS images. We analyze the classification metrics and catalog inclusion status of the synthetic cluster sample as a function of cluster luminosity, age, and mass. We also examine the impact of cluster size and environment on cluster detection and catalog completeness.

We process classification results for the synthetic clusters through the same catalog creation pipeline used for the real LGCS cluster candidates, applying the same classification weighting and $f_{\text{cluster,W}}$ detection threshold. We present the synthetic cluster sample and detection results as an ancillary table in Appendix A. The synthetic cluster results are shown in the left panels of Figure 10, where black points indicate detected clusters and red points indicate non-detections. The middle-left panel shows a 2D binned representation of the results. At higher luminosities and masses, a larger fraction of clusters are detected, and at the highest masses, nearly all synthetic clusters are detected.

The behavior of completeness with age is somewhat more complicated. Clusters at a fixed mass are more frequently detected at young ages due to their brighter luminosities. At fixed luminosity, clusters are more frequently detected at older ages due to their broader distribution of light across member stars, in contrast to young clusters that typically have a small number of very bright massive stars.

¹⁴ <https://acszeropoints.stsci.edu/>

¹⁵ <https://www.stsci.edu/hst/instrumentation/wfc3/data-analysis/photometric-calibration/uv-vis-photometric-calibration>

¹⁶ <https://www.stsci.edu/hst/instrumentation/wfc3/data-analysis/photometric-calibration/ir-photometric-calibration>

¹⁷ $m_{\text{total}} = m_{\text{ap}} + m_{\text{apcor}}$

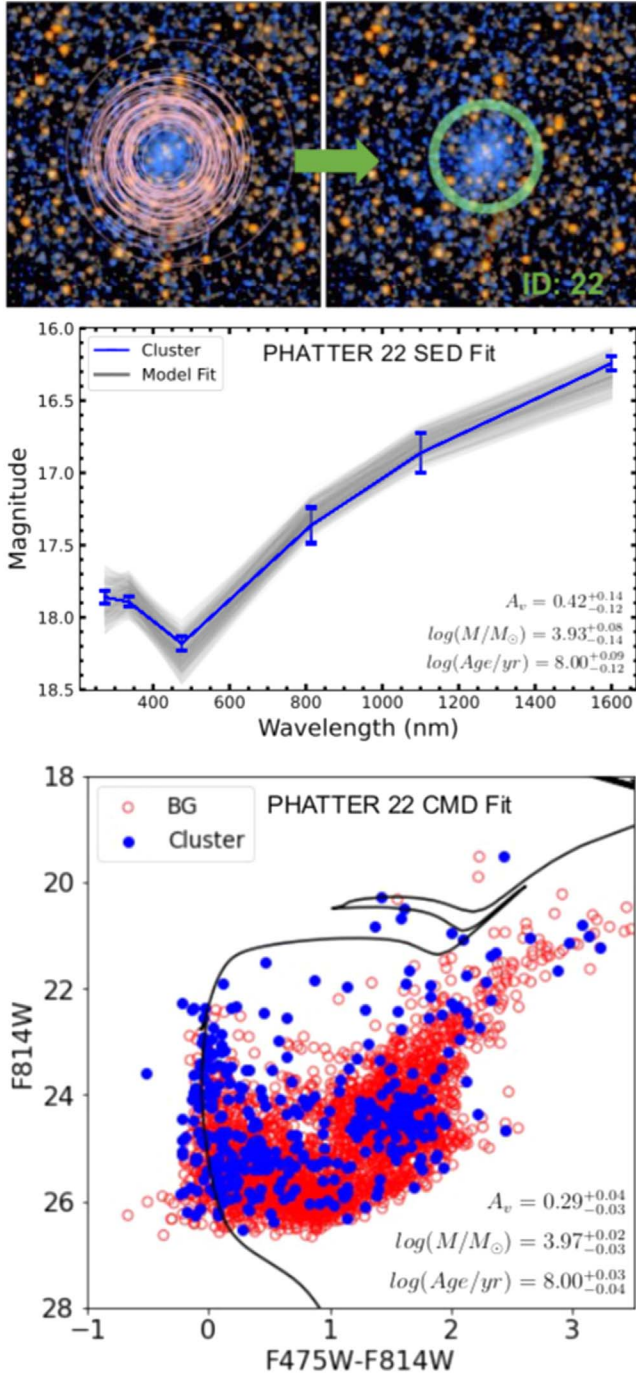


Figure 8. An example cluster: PHATTER 22. Top: two $12'' \times 12''$ F475W + F814W color images, where the pink circles represent user cluster identifications, and the green circle represents the final cluster aperture (R_{ap} of $2''.04$) derived from the median radius of the individual user apertures. Middle: the PHATTER 22 SED (blue) created from six-band integrated photometry. The results of SLUG SED fitting for PHATTER 22 are printed in the lower right (50th percentile value with the 16th–84th percentile confidence interval), and gray lines show the 100 best-fit SLUG models. Bottom: the PHATTER 22 cluster CMD, where stellar photometry for sources within the cluster aperture (blue) are accompanied by surrounding photometry of field stars (red). CMD fitting results from Wainer et al. (2022) are listed in the lower right, and a stellar isochrone (black) depicting the cluster’s best-fit properties is overplotted.

To characterize these observed trends in completeness on the age–magnitude and age–mass planes, we derive analytic formulae that can be easily applied in future modeling work.

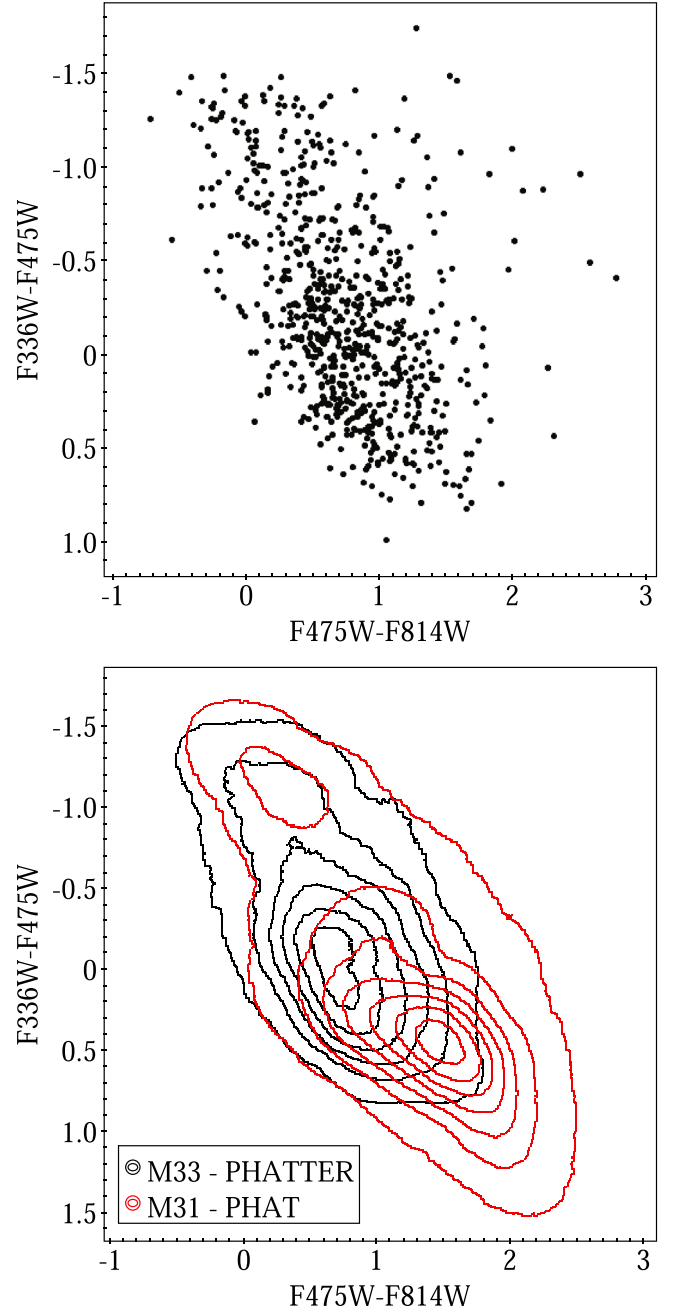


Figure 9. Top: color–color diagram of 729 PHATTER clusters with three-band photometric detections. Bottom: comparison of the color–color distributions for PHATTER clusters in M33 and PHAT clusters in M31 (J15), which shows that the PHATTER clusters are significantly bluer, denoting younger cluster age.

The goal here is to characterize the average completeness properties for the entire catalog, reasoning that detailed studies of specific cluster subpopulations may require a more complex completeness model than the one we present here.

In addition to age, luminosity, and mass, dust extinction, and effective radius (or, surface brightness) may also play a role. We find no need to truncate the cluster effective radius distribution, but choose to omit synthetic clusters in the high A_V tail (>1.5 mag) when fitting for the average model behavior.

We note that the cluster environment also has an impact on completeness (J15). However, because the spatial distribution of half the synthetic clusters was designed to replicate the correlation between young clusters and high N_{MS} regions, we

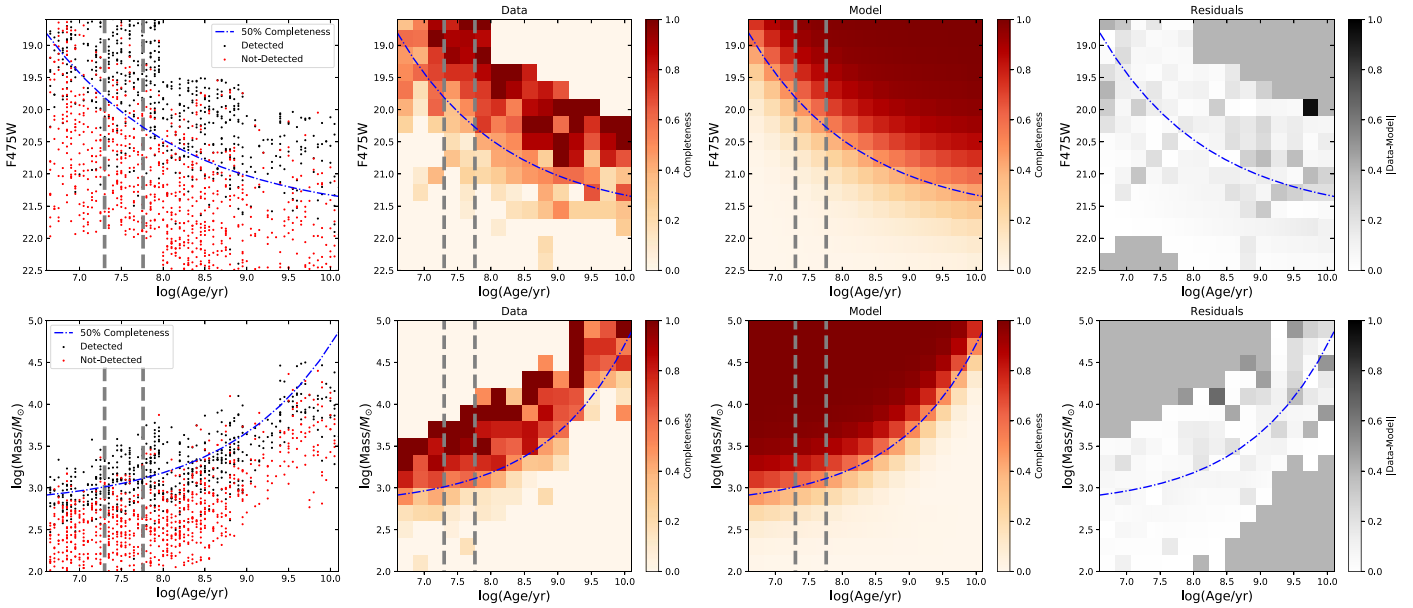


Figure 10. Completeness results based on an analysis of synthetic clusters. The top panels present results as a function of F475W magnitude, the bottom panels as a function of mass. The leftmost panels present results for individual synthetic clusters: detections in black, non-detections in red. The left-middle panels show a binned version of the same results, where the shading denotes the fraction of clusters detected in each bin. The middle-right panels show the completeness model fit, and the rightmost panels showing the data–model residuals. The vertical gray dashed lines in each panel highlight the age bins whose data are presented in Figure 11. The blue dotted–dashed line in each panel represents the fitted exponential 50% completeness model fit to the full synthetic data.

expect our overall distribution of synthetic cluster environments to be comparable to the true cluster sample. We revisit the impact of cluster environment later in this section.

We begin our modeling by characterizing completeness as a function of mass and F475W magnitude for bins of $\log(\text{Age}/\text{yr})$. We use a bin width of 0.23 dex and a sliding two-bin window to improve the number statistics. We use a logistic function to model the completeness curve analytically, as shown in Figure 11, where the black lines show the completeness data and the green line shows the best-fit logistic function. The functional form of the logistic function, C , is given by:

$$C = (1 + e^{-k(x-M_{50})})^{-1}, \quad (3)$$

where k is the slope of the logistic function, M_{50} is the 50% completeness limit, and x is the physical input parameter—either the F475W magnitude or $\log(M/M_{\odot})$.

Based on the initial fits, we find that the slope of the logistic function, k , is quite similar at all ages. Therefore, we fix k to the median fitted value across all age bins (-2.52 for F475W magnitude, 5.84 for mass) and refit.

We find the age dependence of the 50% completeness parameter, M_{50} , is well fit by an exponential function over the full synthetic cluster age range ($6.6 < \log(\text{Age}/\text{yr}) < 10.1$) for both mass and F475W magnitude, following the form:

$$M_{50}(\tau) = a \times e^{b(\tau-\tau_{\min})} + c, \quad (4)$$

where $\tau \equiv \log(\text{Age}/\text{yr})$ and $\tau_{\min}$ is the median τ for the youngest bin, which is 6.71 for the full LGCS synthetic cluster sample and binning used here. We note that in this formulation, the 50% completeness in mass at $\tau_{\min}$ is given by $a+c$. We fit for constants a , b , and c by minimizing χ^2 for the full 2D binned histogram shown in the middle-left panels of Figure 10. We find the best-fit exponential parameter values for mass are $(a, b, c) = (0.1455, 0.7870, 2.7810)$. For F475W magnitude the

best-fit values are $(a, b, c) = (-2.6767, -0.6154, 21.6834)$. The reduced χ^2 for the exponential fit to the mass and F475W magnitude planes is 1.06 and 1.18, respectively. The fitted 50% completeness exponential is presented as a dashed blue line in Figure 10, while the full completeness model (assuming fixed k values) and its residuals relative to the synthetic cluster results are shown in the middle-right and rightmost panels of Figure 10.

The mass completeness behavior in the fiducial age range of $7.0 < \log(\text{Age}/\text{yr}) < 8.5$ has useful application in follow-up science analyses of young clusters, so we derive an alternate completeness model specific to this fiducial age range. Adopting a fixed median k of 6.02 and $\tau_{\min}$ of 7.09, we find the best-fit M_{50} exponential parameters are $(a, b, c) = (0.0303, 1.9899, 2.9770)$, with a reduced χ^2 of 0.82. We recommend this alternative completeness model for any applications that exclude old clusters.

We note that careful inspection of the binned F475W magnitude data in Figure 10 (top middle-left panel) and its model residuals (top right panel) show that the youngest age bins do not neatly follow the large scale trends of the age–magnitude plane. The fact that the age–mass plane is more well behaved at these same ages suggests this difference is likely due to stochasticity impacting the F475W magnitudes due to small number statistics of the brightest cluster members (e.g., Fousneau & Lançon 2010; Beerman et al. 2012). As a result, we recommend use of the mass-based completeness model whenever feasible, as we find it more reliable than our luminosity model in describing these youngest clusters.

5.1. Additional Completeness Dependencies

In addition to age, mass, and F475W magnitude, the spatial distribution of the cluster’s member stars is another physical property that effects completeness. Here, we use the synthetic cluster’s effective radius, R_{eff} , to parameterize the cluster size

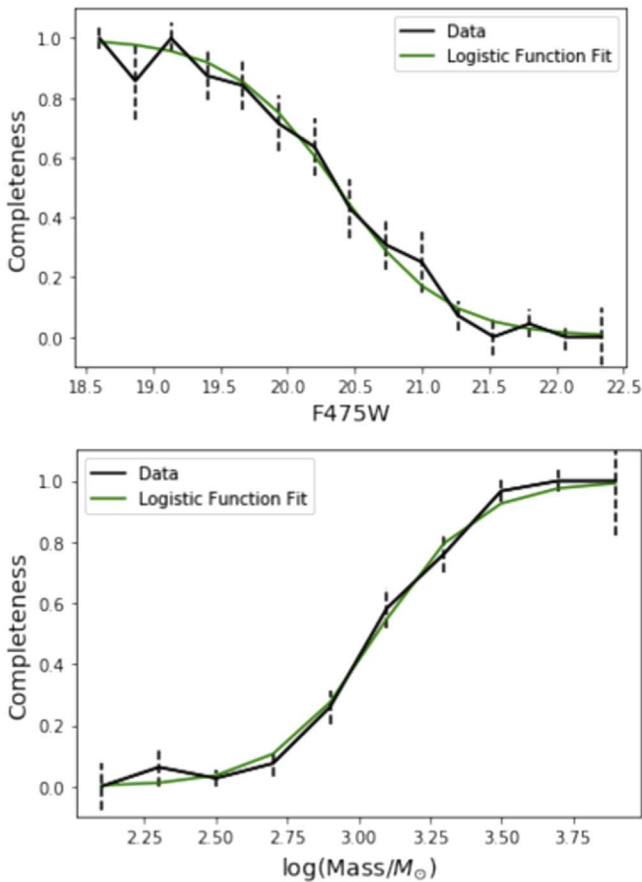


Figure 11. Logistic function fits to catalog completeness as a function of F475W magnitude (top) and mass (bottom) for synthetic clusters with ages $7.3 < \log(\text{Age}/\text{yr}) < 7.77$. Dashed lines represent the Agresti–Coull binomial proportion confidence interval for each bin.

and central density. For synthetic clusters with $R_{\text{eff}} > 3$ pc (33% of sample), we find a 50% completeness limit mass that is 0.2 dex higher than for smaller clusters. This shows that larger, diffuse clusters are systematically more challenging to detect due to the reduced central density of cluster members, which tends to reduce the cluster’s contrast against the field background. As discussed in Section 2.4, we sampled the synthetic cluster radii from the observed cluster radii of the PHAT M31 cluster sample, but biased the distribution toward larger objects. If we remove synthetic clusters with $R_{\text{eff}} > 3$ pc, the mass completeness presented here improves by a median of 0.05 dex. We note that for our final cluster sample, only 14% of the clusters have $R_{\text{eff}} > 3$ pc.

Additionally, local cluster environment plays a role in detection and catalog completeness. At a basic level, when the contrast between a cluster and field is reduced due to an increasing density of nearby field stars of similar color, cluster detection becomes more difficult. For the PHATTER images, we can characterize the local stellar density according to the number of main-sequence stars per LGCS search image, N_{MS} , calculated from counts of PHATTER photometric sources selected according to their optical CMD location. N_{MS} ranges between $2.0 < \log(N_{\text{MS}}) < 3.75$ across the PHATTER footprint. As N_{MS} increases, we see a systematic 0.7 dex increase in the 50% mass completeness limit for the young blue star clusters, whose members are also predominately main-sequence stars. As such, we recommend that future population

analyses take care to account for spatial variation in catalog completeness due to stellar density, especially for trends with respect to galactocentric radius. In particular, we point to the investigation of environmental influence on catalog completeness performed by Wainer et al. (2022) as an example.

5.2. Completeness Comparison: PHATTER versus PHAT

We find that the 50% mass and luminosity completeness limits for the PHATTER M33 cluster catalog are worse relative to the similar PHAT M31 cluster catalog presented by J15; fractional completeness is lower in M33 at a fixed mass or F475W magnitude. At younger ages, the 50% mass completeness limit for the M33 catalog is ~ 0.3 dex higher than what was found for M31, and correspondingly ~ 0.5 mags brighter in F475W magnitude.

While M33’s larger distance can account for a 0.2 mag difference in luminosity, we believe the completeness differences are primarily due to the PHATTER footprint’s central disk location and M33’s higher star formation surface density, which together lead to a higher average density of young field stars (i.e., high N_{MS}) and worse catalog completeness. Thanks to the use of an analogous N_{MS} definition by J15 for PHAT cluster work, we can confirm that the PHATTER median N_{MS} is a factor of 5 (0.7 dex) larger than the PHAT median N_{MS} . We also note that unlike for PHAT, using an alternative F475W₋₃ metric, which subtracts the contribution of the three brightest stars from the integrated F475W magnitude, does not remove the age-dependent trend in luminosity completeness.

6. Cluster SED Fitting

In this section we discuss our method for deriving the cluster properties (age, mass, and extinction) from their integrated light photometry. We use the public source code Stochastically Lighting Up Galaxies (SLUG; Krumholz et al. 2015a) to build a set of 10^7 model star clusters that we use to estimate the M33 cluster properties in Section 6.1. We discuss fitting of various filter combinations and discuss the reliability of this integrated light fitting in Section 6.2 based on comparison to CMD-based results for a similar sample of M31 clusters presented in Appendix B. We derive SLUG-based estimates for the cluster sample and present the results in Section 6.3.

6.1. Building the SLUG Cluster Library

SLUG is a stellar population synthesis code that incorporates stochastic modeling of stellar mass and luminosity distributions. More information and details about SLUG can be found in Krumholz et al. (2015a), and examples of its use include Krumholz et al. (2015b, 2019). Using SLUG, we build a grid of 10^7 model star clusters assuming Padova stellar evolution models that include thermally pulsing AGB stars (Girardi et al. 2000), which are distributed with Starburst99 (Vázquez & Leitherer 2005). We use a Kroupa (2001) stellar IMF which spans from $0.01 M_{\odot}$ to $120 M_{\odot}$ with the “stop after” sampling method, which allows for some of the more massive stars to be included in our simulated sample (Krumholz et al. 2015a). We generate models with SLUG by drawing ages from a t^{-1} distribution over a range of 10^6 to 10^{10} years, encompassing the majority of clusters in the PHATTER catalog. We draw cluster masses from an M^{-2} distribution and draw dust extinction values from a lognormal centered at an A_V of 1 mag, width of

Table 2
Detection Statistics for Passband Combinations

Passband	N(Detections)
F275W+F336W+F475W+F814W+F110W+F160W	349 (28.7%)
F336W+F475W+F814W+F110W+F160W	414 (34.1%)
F275W+F336W+F475W+F814W	612 (50.4%)
F336W+F475W+F814W	729 (60.0%)

Note. The number and percentage of PHATTER clusters with photometric detections in each of the listed combinations of three, four, five, and six filter passbands.

0.33 mag, and min/max breakpoints of 10^{-6} and 5 mag, respectively. We apply extinction according to a Milky Way extinction curve (Fitzpatrick 1999), compute photometry in Vega magnitudes for the six HST filters (ACS: F475W and F814W; WFC3: F275W, F336W, F110W, and F160W), and convert from model magnitudes to observed magnitudes using an adopted distance modulus of 24.67 (de Grijs & Bono 2014).

6.2. Filter Selection and Fitting

The reliability of SLUG cluster property determinations depends on the number of filters we are able to include in the fitted SEDs. Therefore, we begin by analyzing the fraction of clusters with good photometry in various combinations and numbers of filters for the PHATTER sample. Specifically, we choose filter combinations in increasing order of photometric detectability, and present the results for the M33 sample in Table 2.

We compute SLUG fits for each filter combination listed in Table 2 and the corresponding sample of clusters that are detected in all of the combination’s selected passbands. The SLUG model grid is trimmed for each passband combination to omit the magnitudes of any filters that are not selected. Once the cluster fits from each filter combination are compiled, we adopt the fit for each cluster that results from the filter combination with the largest number of filters. This ensures that we are not fitting incomplete SEDs and that we obtain fits for a maximum number of total clusters.

We execute each iteration of SLUG fitting, and process each set of results using a fixed set of parameters and assumptions. We adopt the following settings that relate to the specifics of the fitting process: a photometric bandwidth of 0.02, a physical properties bandwidth of 0.05, and a Gaussian kernel for probability density function (PDF) estimation. And same as for the underlying set of models, we assume a t^{-1} age prior, M^{-2} mass prior, and lognormal A_V prior. The code returns marginalized PDFs for age, mass, and dust extinction of each cluster, from which we can derive the 16th, 50th, and 84th percentile values. These percentiles yield median estimates for each physical property accompanied by an associated 1σ uncertainty. We also flag and exclude highly uncertain fits, such that any cluster with a 16th to 84th percentile range greater than 1.2 dex in age or 1.3 dex in mass is identified by an error flag in the fitting results.

We present an example of a fitted cluster SED in the middle panel of Figure 8 for PHATTER 22. The 100 best-fit SEDs from the SLUG library are plotted along with the observed SED, which show good agreement between models and observations. The median, 16th, and 84th percentiles of the marginalized posterior PDF for cluster age, mass, and dust

extinction, computed over the full library of model SEDs, are derived using functions from the `cluster_slug` package that is included as part of the SLUG code. We find good agreement between SED and CMD fitting results for PHATTER 22.

6.3. SLUG Results

We derive cluster properties using SLUG for 729 objects with detections in at least one of the filter combinations listed in Table 2; we report the fitting results in Table 3. We note that the limited number of fitted clusters (729 out of 1214; 60%) is due to a minimum three-filter (F336W, F475W, and F814W) detection criteria for SED fitting. As a result, the completeness of this fitted sample of clusters is worse than the overall catalog completeness, and is biased toward younger and brighter clusters.

Excluding flagged cases with broad PDFs, the median ages from the cluster PDFs range from $6.08 < \log(\text{Age}/\text{yr}) < 8.91$ with a median value of 7.96 for our cluster sample. The median cluster PDF masses range from $2.14 < \log(M/M_\odot) < 4.59$ with a median value of 3.29. The median 16th to 84th percentile range in mass is 0.46 dex, and the median 16th to 84th percentile range in age is 0.41 dex.

To gauge the reliability of the SLUG fits for the PHATTER cluster sample, we compare newly derived SLUG results to high-quality CMD-based cluster fits for a similar sample of clusters in M31. We find that the masses are reliably determined via SLUG integrated light fitting, but that SLUG age and dust results suffer from large uncertainties and artifacts. In particular, the SLUG fits tend not to reliably recover ages for clusters younger than ~ 100 Myr, which instead are often fit with older ages; see Appendix B for full details of the comparison analysis and results. Due to these results, we recommend that the CMD-based fits for the younger clusters in the PHATTER cluster sample presented in Wainer et al. (2022) should be preferred over the SLUG fits reported here. At older ages, there are fewer resolved stars, making the SLUG age estimates the better (and sometimes the only) option.

7. Discussion

7.1. Comparison to Existing M33 Cluster Catalogs

We crossmatch the full PHATTER candidate list (clusters, galaxies, emission regions, and remaining ancillary objects; see Appendix A) to five different catalogs from the literature: Sarajedini & Mancone (2007), San Roman et al. (2009, 2010), Sharma et al. (2011), and Corbelli et al. (2017). These references are chosen to facilitate three types of comparisons: to comprehensive catalogs (Section 7.1.1), to HST-based catalogs (Section 7.1.2), and to infrared catalogs (Section 7.1.3). We compile the crossmatching results in Table 4 where we list identifiers (and object classes, where relevant) for each catalog, as well as additional alternate names and accompanying references. These matches are based on a one arcsecond matching radius between cataloged positions, after applying a mean astrometric offset of $0''.609$ to the Sarajedini & Mancone (2007) cluster sample before crossmatching.

Table 3
SLUG Results

ID	Error Cut Flag	Filters Available	log(Mass/ $M_{\odot}$)			log(Age/yr)			A_V		
			P16	P50	P84	P16	P50	P84	P16	P50	P84
3	F	4	2.47	3.23	3.43	7.72	8.25	8.41	0.24	0.43	0.69
5	F	6	3.99	4.07	4.16	8.41	8.49	8.57	0.23	0.41	0.57
7	F	6	3.97	4.05	4.12	8.40	8.49	8.58	0.38	0.58	0.74
8	F	6	4.53	4.59	4.65	8.17	8.23	8.28	0.20	0.26	0.32
11	F	6	2.45	3.46	3.56	7.71	8.31	8.38	0.10	0.23	0.32

Note. The error cut flag identifies cases where fits have large uncertainties (16th to 84th percentile range >1.2 dex in age or >1.3 dex in mass) and should be excluded from uses where uncertainties are not factored in explicitly.

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

Table 4
Literature Crossmatching Results

ID	SM07 ID	SM07 Class ^a	SR10 ID	SR10 Class ^b	SR09 ID	S11 ID	C17 ID	C17 Class ^c	Alternate Names & References ^d
1	...	...	1849	3	157	...	...	...	...
2	391	Unknown	2039	0	...	...	...	...	CS U80
3	...	...	2025	0	...	...	...	...	...
4	...	...	1441	−1	...	...	...	...	...
5	372	Cluster	1959	3	...	...	...	...	CBF 58; MKKSS 50; CS U91

Note. Literature references: SM07 (Sarajedini & Mancone 2007), SR10 (San Roman et al. 2010), SR09 (San Roman et al. 2009), S11 (Sharma et al. 2011), and C17 (Corbelli et al. 2017).

^a Sarajedini & Mancone (2007) Classes: cluster, stellar, galaxy, unknown.

^b San Roman et al. (2010) Classes: −1 = galaxy; 0 = unknown extended object; 1 = candidate cluster; 2 = highly probable cluster; and 3 = confirmed cluster.

^c Corbelli et al. (2017) Classes: b = associated with clouds, no optical counterpart; c1 = associated with clouds, coincident H_{α} , and mid-infrared peaks; c2 = associated with clouds: coincident H_{α} , mid-infrared, and UV peaks; c3 = not associated with clouds but optically detected; d = ambiguous; e = not associated with clouds, mostly mid-infrared peaks only.

^d Reference Abbreviations for Alternate Names (from Sarajedini & Mancone 2007): Hilt (Hiltner 1960), MD (Melnick & D’Odorico 1978), CS (Christian & Schommer 1982), MKKSS (Mochejska et al. 1998), CBF (Chandar et al. 1999, 2001), BEA (Bedin et al. 2005), SBGHS (Sarajedini et al. 2007).

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

7.1.1. Primary Catalog Comparisons: Sarajedini & Mancone (2007) and San Roman et al. (2010)

We compare the PHATTER cluster catalog to two key M33 cluster catalogs in the literature: Sarajedini & Mancone (2007) and San Roman et al. (2010). We focus on these catalogs for our primary literature comparison due to their comprehensive compilation of published M33 cluster catalogs and their complete, uniform spatial coverage, respectively.

We summarize the recovery of clusters from the optically selected catalogs of Sarajedini & Mancone (2007) and San Roman et al. (2010) in Figure 12. We recover a large fraction of the cataloged cluster candidates that fall within the spatial coverage of the PHATTER survey. Specifically, for objects that Sarajedini & Mancone (2007) classify as “cluster” and lie within the PHATTER footprint, 89% are present in the PHATTER cluster catalog. We also classify 85% of their “unknown” objects as clusters.

For the portion of the San Roman et al. (2010) catalog that falls within the PHATTER footprint, 92% of their “confirmed star cluster” (class 3) objects and 67% of their “highly probable star cluster” (class 2) objects are present in the PHATTER cluster catalog. Most of the remaining class 2 and 3 San Roman et al. (2010) objects are recovered by the LGCS search, but lie in a long tail at low- $f_{\text{cluster},W}$ values, as shown in the top left panel of Figure 12. Additionally, 41% of their less certain cluster identifications (class 1 “candidate star cluster” and class

0 “unknown”) are also identified as clusters in this work. More surprisingly, we find that half of the “background galaxy” (class −1) objects that lie within the PHATTER footprint are classified as clusters by the LGCS search. Visual inspection of the HST images of these objects confirm they are in fact clusters, and thus were misclassified by the ground-based San Roman et al. (2010) effort.

Finally, we note that a small fraction of the Sarajedini & Mancone (2007) and San Roman et al. (2010) sources that fall within the PHATTER footprint were not recovered in the LGCS search. We show the class distribution of these objects in the right panels of Figure 12.

7.1.2. Space-based Comparisons: HST Catalogs

We compare the PHATTER cluster catalog to three HST-based cluster catalogs from the literature: Chandar et al. (1999, 2001, hereafter, collectively CBF), Bedin et al. (2005), and San Roman et al. (2009). We note that CBF and Bedin et al. (2005) were crossmatched as members of the Sarajedini & Mancone (2007) compilation, and San Roman et al. (2009) was matched individually. While these literature catalogs were derived from imaging data sets with relatively small spatial footprints, they serve as useful points of comparison for analyzing catalog-specific differences in the visual cluster identification of HST imagery.

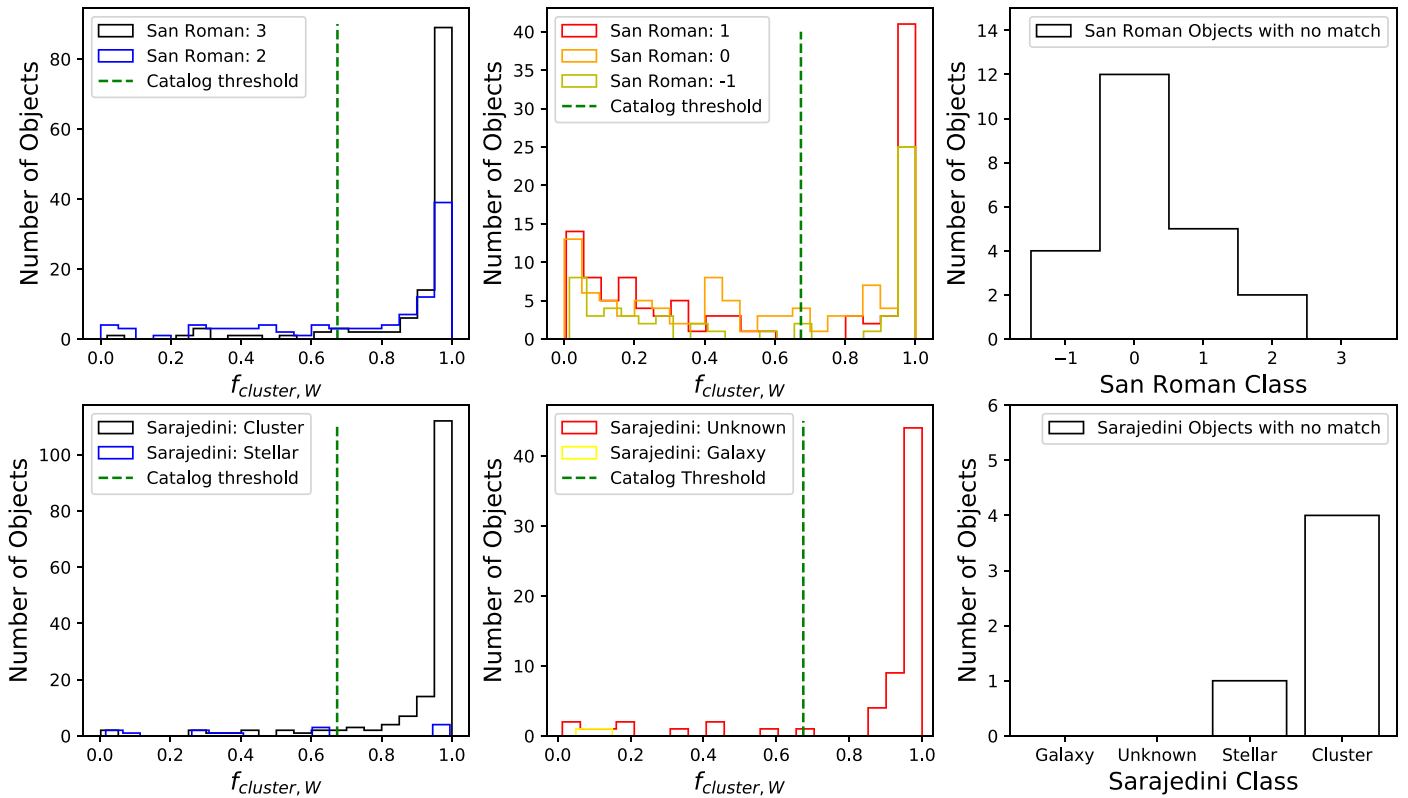


Figure 12. A comparison of LGCS classifications for objects in the San Roman et al. (2010) and Sarajedini & Mancone (2007) catalogs that fall within the PHATTER footprint. Top: left and center panels show $f_{\text{cluster},W}$ distributions for the San Roman et al. (2010) catalog, where object classes include “confirmed star cluster” (class 3), “highly probable star cluster” (class 2), “candidate star cluster” (class 1), “unknown” (class 0), and “background galaxy” (class -1). The dashed line represents the $f_{\text{cluster},W}$ threshold we use for cluster catalog inclusion. The right panel shows the San Roman et al. (2010) classes for objects located in the PHATTER footprint that did not match to a LGCS identification. Bottom: same as top, but for matched (left and center) and unmatched (right) objects from the Sarajedini & Mancone (2007) catalog and its different sets of object classes.

A comparison between the PHATTER catalog and the CBF catalog shows that the PHATTER catalog recovers nearly all previously identified objects (103 out of 110 that lie within the PHATTER footprint). However, the PHATTER catalog includes an additional 445 clusters within the spatial footprint searched by CBF, resulting in a total catalog that is a factor of 5 larger. This discrepancy is expected due to CBF’s use of WPC2 images whose wide field cameras have significantly lower spatial resolution (pixel scale of $0''.1$) than the ACS and WFC3 instruments used by the PHATTER survey (pixel scales of $0''.05$ and $0''.04$, respectively). Lower spatial resolution images make the identification of faint, low-mass clusters much more difficult, leading to worse luminosity and mass completeness limits for the CBF catalog and significantly fewer cluster identifications as a result.

Next we compare the PHATTER catalog to the work of Bedin et al. (2005), who used a single ACS pointing located within the PHATTER survey footprint to identify 33 clusters. Within this spatial region, the PHATTER catalog includes 33 clusters, where 22 entries are shared between the two works. The 11 unmatched clusters from the Bedin et al. (2005) catalog break down into two categories: (1) six objects are explained by object definition differences, where PHATTER categorized these objects as emission regions or loose non-cluster associations; (2) five objects are identified by the PHATTER search, but are excluded from the cluster catalog due to low $f_{\text{cluster},W}$. For the 11 PHATTER clusters not recovered by Bedin et al. (2005), we believe the mismatch is due to their use of a singleband F775W ACS image. These clusters have lower

$f_{\text{cluster},W}$ and fainter m_{F475W} than most of the 22 matched clusters. Faint, low-mass clusters are identified in PHATTER via a small clustering of blue main-sequence stars, and therefore it is expected that these objects would be missed in a search of only red wavelength imagery.

Finally, we compare the PHATTER catalog to the work of San Roman et al. (2009), who searched multiband ACS imagery that partially overlaps with the PHATTER survey footprint and identified 86 clusters in the overlapping region. The PHATTER cluster catalog includes 75 of these previously identified objects, leaving 11 objects that were excluded by the PHATTER catalog selection threshold. Importantly, the PHATTER catalog includes 119 clusters not identified by San Roman et al. (2009), resulting in a total catalog that is a factor of 2.3 larger. Upon examination, this significant difference is due to a more conservative selection threshold, where the PHATTER catalog tends to probe to lower $f_{\text{cluster},W}$ and fainter m_{F475W} . Imagery wavelength may also play a role here, as more than half of the PHATTER overlapping fields were only imaged in two redder bands (F606W and F814W) without bluer F475W coverage.

7.1.3. Infrared Comparisons: Alternative Catalogs

Studies in the literature have also made use of infrared images to assess M33’s cluster population. Sharma et al. (2011) uses Spitzer 24 μm images to construct a catalog of young stellar clusters, where it is assumed these objects are still embedded in their natal molecular clouds. Of the 240 Sharma

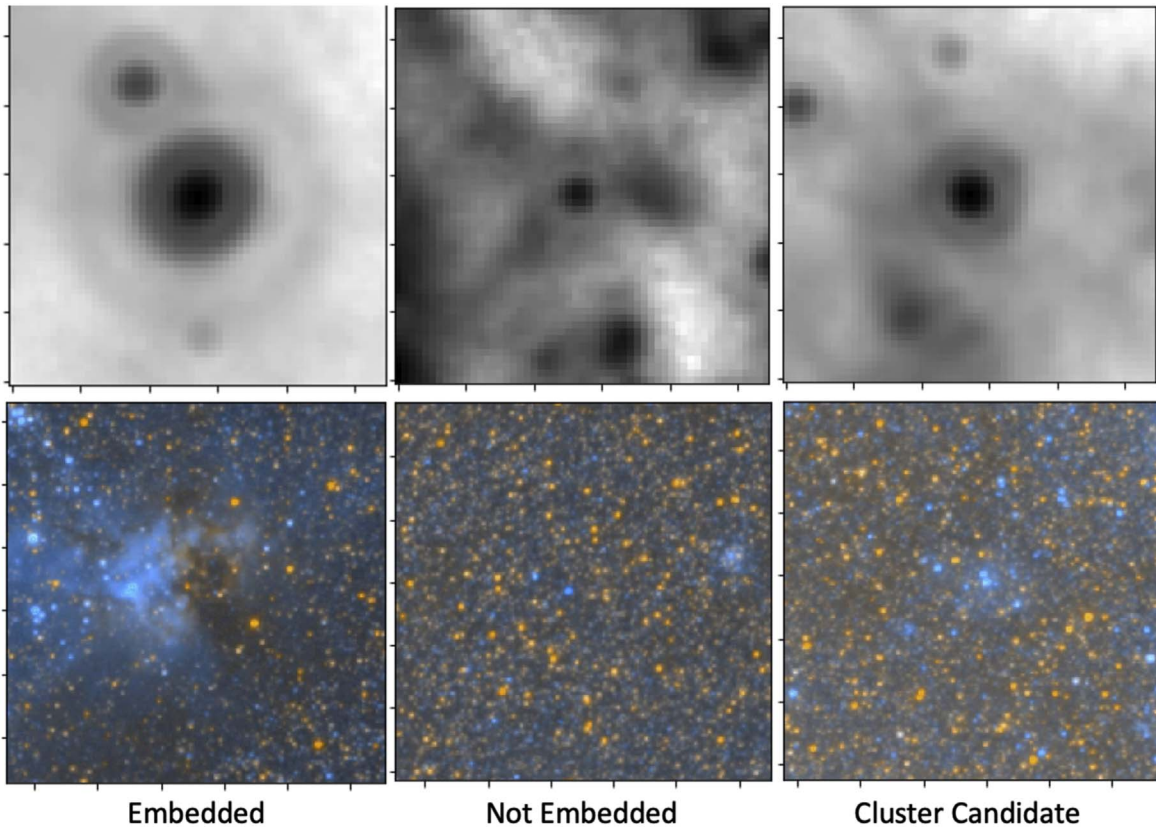


Figure 13. Spatially matched Spitzer $24\ \mu\text{m}$ (top) and HST F475W–F814W (bottom) color cutouts for three Corbelli et al. (2017) and Sharma et al. (2011) objects that represent three scenarios for the objects: an object which shows evidence of being a young embedded star-forming region (left; Corbelli YSCC 225), an object where there is sufficient background visible to rule out an embedded cluster (middle; Corbelli YSCC 215), and a rare object which we classify as a star cluster in the PHATTER catalog (right; Corbelli YSCC 297, PHATTER 529).

et al. (2011) objects that fall within the PHATTER footprint, we only identify 41 crossmatches, of which we only classify five as star clusters while the remainder are mostly low- f_{view} candidates. The poor correspondence between these two catalogs corroborates the conclusions of Sun et al. (2016), who advocate against the use of the Sharma et al. (2011) catalog for star cluster population studies due to contamination by non-cluster objects, as we noted in Section 1.

The catalog of Sharma et al. (2011) was used as a starting point for compiling a sample of young cluster candidates for use in a cross-comparison with CO molecular clouds by Corbelli et al. (2017). Of the 291 objects from Corbelli et al. (2017) within the PHATTER footprint, we match 91 objects and find just 12 to be star clusters identified in our catalog. As with the Sharma et al. (2011) sample, these cluster candidates are generally not associated with optical star clusters, and caution should be taken when using this catalog for the purpose of star cluster population work.

To provide a visual example of optical versus infrared cluster candidates, we present Spitzer $24\ \mu\text{m}$ and HST F475W + F814W color cutout pairs for three Corbelli et al. (2017) and Sharma et al. (2011) objects in Figure 13. The first is an example of a region where there is a bright $24\ \mu\text{m}$ source, and in the color image there is visible extinction, indicative of molecular gas, and an emission region. These indicators confirm the presence of a young embedded star-forming region and potentially (but not certainly) a young star cluster. In the absence of certainty on the presence of a bound star cluster, this object is not identified as a cluster in the PHATTER catalog.

The second example is an object where, even though there is a $24\ \mu\text{m}$ source, the background field in the optical image is fully visible with no sign of dust obscuration. Therefore, we conclude the probability of an embedded cluster is low. The third example is an object that is classified as a cluster in our catalog. Examination of the PHATTER imaging around these positions suggests that for at least 30% of Corbelli et al. (2017)’s objects, we can rule out the presence of an embedded star cluster based on the uniformity of the background at the location of the Corbelli et al. (2017) cluster candidates. Thus it appears that a significant fraction of these candidates are neither young embedded star-forming regions nor optically visible star clusters.

7.1.4. Catalog Comparison Summary

Overall, the PHATTER cluster catalog significantly enhances the population of known clusters in the inner disk region of M33. Out of 1214 total clusters, 810 (67% of the catalog) are identified here for the first time. We compare the luminosity distributions of the PHATTER clusters and previously identified objects from the literature in Figure 14, showing that PHATTER’s sample probes fainter, lower-mass clusters than the previous works. The superior spatial resolution of the HST images facilitates this marked increase, as tight groups of stars can be differentiated from a single, barely or unresolved source. In addition, the ability to identify small groupings of faint resolved stars in the PHATTER images leads to our ability to probe further down the cluster mass function in M33 than ever before.

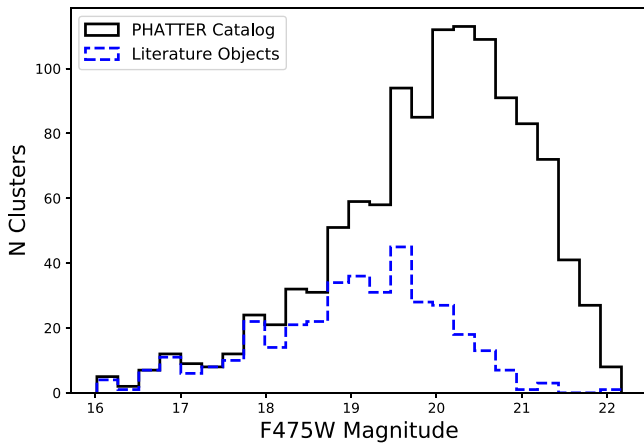


Figure 14. Shown here are the F475W magnitude distributions of PHATTER clusters (black) and previously identified literature objects (blue). Shown in black are 1166 PHATTER clusters with F475W detections. In blue are 391 literature objects included in the PHATTER cluster catalog with F475W detections. Literature objects not represented above: 13 objects lacking F475W detections and 178 objects with low $f_{\text{cluster},W}$ omitted from the PHATTER cluster catalog.

7.2. M33 Catalog in Context: Expectations from the Γ and M_c Correlations

We examine whether the PHATTER sample size of 1214 clusters matches expectations based on our knowledge of M33 and the relation between galaxies and their cluster populations. An easy comparison we can make is to the PHAT M31 catalog (J15) and its 2753 clusters. We might naively expect M31’s sample to outnumber M33’s based on the number of bricks in the survey footprint (23 versus 3) or the total galaxy stellar mass ($6 \times 10^{10} M_\odot$ versus $3 \times 10^9 M_\odot$; Patel et al. 2017). In this context, it seems there should be a much larger than factor of ~ 2 difference in cluster count. However, differences in ongoing star formation and spatial coverage fraction of the HST imaging could plausibly provide an explanation, especially given that both cluster samples are dominated by young star clusters. Therefore, in this section we endeavor to make an SFR-based accounting of the M33 cluster sample to test whether we can explain the sample size differences we see between M33 and M31. We will use observed relations with respect to galaxy Σ_{SFR} to predict cluster formation efficiency (Γ) and Schechter truncation mass (M_c), and use these inputs to infer a prediction for cluster sample size.

To begin, we derive a SFR and Σ_{SFR} measurement for the PHATTER survey region in M33. Following the methodology laid out in Appendix A of Johnson et al. (2017), we use GALEX FUV and Spitzer $24 \mu\text{m}$ images and the Σ_{SFR} prescription from Leroy et al. (2008). We find $\log(\langle \Sigma_{\text{SFR}} \rangle / M_\odot \text{ yr}^{-1} \text{ kpc}^{-2})$ of $-2.04^{+0.16}_{-0.18}$ for M33 within the PHATTER survey footprint. We also extract the total SFR within the survey region, which we find is $0.12 M_\odot \text{ yr}^{-1}$, approximately 46% of M33’s total SFR.

Next, we can use the PHATTER Σ_{SFR} value to derive an estimate of the cluster formation efficiency, Γ , based on the Γ - Σ_{SFR} relation from Johnson et al. (2016). This relation predicts a cluster formation efficiency of 15%. We can also derive an estimate for the cluster mass function by assuming a Schechter function form with a -2 power-law slope and truncation mass, M_c estimate based on the M_c - Σ_{SFR} relation

from Johnson et al. (2017). This relation predicts a Schechter M_c of $4.3 \times 10^4 M_\odot$.

When we combine the determinations of M33 SFR and Γ , we derive a cluster formation rate (CFR) of $0.0180 M_\odot \text{ yr}^{-1}$. Interestingly, this is very similar to the PHAT M31 CFR of 0.0186 ($\Gamma = 6.4\%$ via Johnson et al. 2016 and $\text{SFR} = 0.29$ via Lewis et al. 2015). Although the PHATTER M33 SFR is 2.4 times smaller than the PHAT M31 survey SFR ($0.12 M_\odot \text{ yr}^{-1}$ versus $0.29 M_\odot \text{ yr}^{-1}$), M33’s Σ_{SFR} is almost 4 times higher, which leads to a $2.4\times$ higher Γ (15% versus 6.3%) that cancels out the SFR difference.

Based on the well-matched M31 and M33 CFRs, the relative difference in size between the two cluster catalogs (2753 versus 1214; $2.3\times$) seems unexpected. Note that the variation in M_c ($8.5 \times 10^3 M_\odot$ versus $4.3 \times 10^4 M_\odot$) is only expected to make a few percent difference in the number statistics due to the small number of clusters at the high-mass end, so that does not explain the difference. However, catalog completeness differences likely play a role. M33’s high stellar density within the PHATTER footprint in the central region of the disk leads to a higher mass for the 50% completeness limit than in M31: $\log(M/M_\odot)$ of 3.2 versus 3.0 for a nominal 100–300 Myr age range. However, this 0.2 dex offset only affords a $1.3\times$ – $1.7\times$ correction to M33 N_{clusters} , leaving another factor of $1.3\times$ – $1.7\times$ still unexplained. This remaining discrepancy could be due to our assumption of a constant SFH, or perhaps we will find that Γ is not as high in M33 as predicted.

Overall, this exercise shows that the star formation differences between M31 and M33 are likely significant enough to impact cluster populations. We look forward to using robust star formation history fitting (Lazzarini et al. 2022) to inform recent SFR and Σ_{SFR} determinations for M33, and aid future Γ determinations for the PHATTER cluster sample.

7.3. Cluster Affiliated Phenomena: X-Ray Sources & Planetary Nebulae

One of the immediate uses of the PHATTER cluster catalog is to crossmatch it with objects of interest in M33, such as planetary nebulae (PNe), X-ray emitting sources, and other stellar populations. Identifying associations between clusters and these source populations can provide useful information about the source, such as a cluster-based age or additional information that assists in source classification.

We begin by searching for crossmatches between the PHATTER cluster catalog and two M33 X-ray source catalogs created from Chandra (ChASem33 survey; Tüllmann et al. 2011) and XMM-Newton (Williams et al. 2015) observations. We use an initial matching radius of $5''$, but require the X-ray source to fall within the aperture radius of the cluster center (typically $\sim 1.5''$), resulting in two matches—one from each X-ray catalog. The matched Chandra source ChASem33 393 is paired with PHATTER 675, a known globular cluster (ID: 275; Sarajedini & Mancone 2007) whose association with an X-ray source was identified by Tüllmann et al. (2011). The matched XMM-Newton catalog entry Source 716 is associated with PHATTER 29, a previously identified young cluster (ID: 260; Sarajedini & Mancone 2007) with a CMD-estimated age of 10 Myr (Wainer et al. 2022). Given the ages of the clusters, the Chandra source in the globular cluster is likely to be a bright low-mass X-ray binary, and the XMM-Newton source in the young cluster is likely to be a bright high-mass X-ray binary.

We also report the presence of a possible PN associated with a PHATTER cluster. The candidate PN associated with the PHATTER 4 cluster was discovered serendipitously while reviewing sources with outlier optical colors in cluster CMDs. PNe are known to appear as anomalous blue sources in F475W – F814W color due to strong line emission in the F475W bandpass (Veyette et al. 2014). PHATTER 4 has a very uncertain SLUG integrated light age determination, most likely due to the unmodeled contribution of nebular line emission from the PN, however its CMD-fitted age estimate of ~ 1 Gyr (Wainer et al. 2022) is not unexpected for a PN-hosting star cluster.

Given the serendipitous identification of the PN, we conducted a search for additional candidates. We performed a crossmatch of the PHATTER clusters with the PN catalog of Ciardullo et al. (2004), but we found no candidates that lie within the aperture radius from the cluster center. This lack of matches is not unexpected, however, given that a cluster PN would likely have been rejected by the Ciardullo et al. (2004) search of ground-based narrowband imaging due to the presence of coincident continuum emission from the cluster. We also perform a search for other cluster members with anomalous optical colors ($F475W - F814W < -1$), but find no other reliable sources among the cluster members.

We will continue to expand the PHATTER cluster catalog crossmatching and perform analyses of cluster membership to additional source populations in future work. Following on from work conducted for PHAT, cluster membership of AGB stars (Girardi et al. 2020), Cepheid variables (Senchyna et al. 2015), and other populations are ripe for study in the PHATTER data.

8. Summary

We present the results of a crowdsourced visual star cluster search of M33 conducted as part of the LGCS citizen science project using imaging from the PHATTER survey. The resulting catalog of 1214 star clusters has well-characterized completeness properties and a 50% completeness limit of approximately $1500 M_{\odot}$ at an age of 100 Myr. We derive ages and masses from SED fitting of the subset of clusters with multiband detections in the catalog’s integrated aperture photometry. We find the sample is composed primarily of young, low-mass star clusters, although the SLUG-fitted clusters are a biased subsample of the full PHATTER catalog.

This cluster catalog builds upon similar Local Group cluster work in M31 (J15) and significantly increases the number of known Local Group star clusters observed with HST. The PHATTER cluster catalog samples higher Σ_{SFR} galactic properties than M31, which provides leverage for studying how cluster properties like the cluster mass function, cluster formation efficiency, and more depend on star formation intensity. In accompanying work, we use the CMDs of individually resolved stars to fit high-precision ages and masses, and to constrain the mass function of young clusters (Wainer et al. 2022). We also expect the sample will also be useful for calibrating models of stellar evolution (e.g., Girardi et al. 2020) and other future M33 cluster studies.

We enthusiastically thank the ~ 2800 LGCS volunteers who made this work possible. Their contributions are acknowledged individually at <https://authors.clustersearch.org>. This publication uses data generated via the Zooniverse.org platform, the

development of which is funded by generous support, including a Global Impact Award from Google, and by a grant from the Alfred P. Sloan Foundation. Support for this work was provided by NASA through grant number HST-GO-14610 from the Space Telescope Science Institute, which is operated by AURA, Inc., under NASA contract NAS5-26555. L.C.J. acknowledges support through a CIERA Postdoctoral Fellowship at Northwestern University. This material is partially based on work by T.M.W. as a CIERA REU student at Northwestern University, supported by the National Science Foundation under grant No. AST-1757792. Some of the data presented in this paper were obtained from the Mikulski Archive for Space Telescopes (MAST). This research made use of NASA’s Astrophysics Data System (ADS) bibliographic services.

Facilities: HST(ACS, WFC3).

Software: astropy (Astropy Collaboration et al. 2013), DOLPHOT (Dolphin 2000), DrizzlePac (STSCI Development Team 2012; Hack et al. 2013; Avila et al. 2015), SLUG (Krumholz et al. 2015a), TOPCAT (Taylor 2005).

Appendix A Ancillary Catalogs

To accompany the primary PHATTER cluster catalog described in Section 4 and presented in Table 1, we present ancillary catalog results in this Appendix. First, we present an ancillary catalog in Table 5, which contains 3566 candidate identifications with $f_{\text{cluster,W}} < 0.674$ and $f_{\text{view}} \geq 0.1$. These objects fall below the catalog threshold adopted for the primary PHATTER cluster sample, however we publish these results to enable investigators to make their own decisions regarding completeness versus contamination, depending on their science case.

Second, we present object-by-object synthetic cluster results in Table 6. These results can be used to derive catalog completeness for the case where a different catalog selection is adopted.

Third, we identify a sample of background galaxies using a selection criteria of $f_{\text{galaxy}} \geq 0.25$. We present this sample of 203 background galaxies identified by LGCS participants in Table 7. The adopted f_{galaxy} detection threshold was chosen via visual inspection and ensures good sample purity. We find seven crossmatches with San Roman et al. (2010), three crossmatches with Sarajedini & Mancone (2007), and three crossmatches with Corbelli et al. (2017) for the background galaxy sample.

Fourth, we identify a sample of emission regions using a selection criteria of $f_{\text{emission}} \geq 0.25$. We present this sample of 95 emission regions identified by LGCS participants in Table 8. The adopted f_{emission} detection threshold was chosen via visual inspection and ensures good sample purity. We urge caution when interpreting this sample due to the fact that cluster identification, not emission region identification, was the primary goal of the LGCS search. As a result, we do not expect this sample to be systematically complete, and instead recommend that these objects be considered opportunistic identifications. We note that three objects listed here are also included in the PHATTER cluster catalog, but where we agree with the cross-listing between the two classification categories: objects 163, 275, and 652. Also, we find 41 crossmatches with San Roman et al. (2010), 15 crossmatches with Sarajedini & Mancone (2007), and 21 crossmatches with Corbelli et al. (2017) for the emission region sample.

Table 5
Ancillary Catalog

ID	R.A. (J2000)	DECL. (J2000)	R_{ap} (")	R_{eff} (")	f_{view}	f_{cluster}	f_{galaxy}	f_{emission}	$f_{\text{cluster,W}}$	Flags	m_{apcor}	m_{275}	σ_{275}	m_{336}	σ_{336}	m_{475}	σ_{475}	m_{814}	σ_{814}	m_{110}	σ_{110}	m_{160}	σ_{160}
37	23.474286	30.498702	2.10	0.89	0.9333	0.3583	0.4667	0.1083	0.4833	...	−0.31	...	...	...	...	20.48	0.12	18.86	0.01	...	...	...	...
46	23.363899	30.699947	1.90	0.35	0.9667	0.3000	0.5833	0.0833	0.3108	...	−0.01	19.49	0.21	19.28	0.13	19.23	0.06	17.69	0.08	...	...	...	...
54	23.400054	30.709211	2.79	0.69	0.9667	0.2667	0.0000	0.7000	0.3235	...	−0.06	17.33	0.14	17.41	0.07	18.73	0.11	18.56	0.14	17.93	0.12	>18.28	...
57	23.469856	30.730419	3.84	1.48	0.9583	0.3583	0.0000	0.6000	0.4373	...	−0.25	16.49	0.07	16.60	0.11	17.62	0.12	17.24	0.10	...	...	...	...
60	23.527783	30.815464	3.74	0.81	0.9667	0.2500	0.0000	0.7167	0.2552	...	−0.03	16.11	0.02	16.34	0.05	17.48	0.02	17.56	0.08	17.13	0.06	>17.88	...

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

Table 6
Synthetic Cluster Results

SYNID	R.A. (J2000)	DECL. (J2000)	$\log(M/M_{\odot})$	$\log(\text{Age/yr})$	Z	A_V	$R_{\text{eff,in}}$ (pc)	$F475W_{\text{in}}$	$F814W_{\text{in}}$	f_{view}	f_{cluster}	f_{galaxy}	f_{emission}	$f_{\text{cluster,W}}$	N_{MS}	N_{RGB}	Detected
1	23.638290	30.781798	2.79	7.50	0.0152	0.44	1.09	20.64	20.59	0.0	0.0	0.0	0.0	0.0	3184	163	False
2	23.630015	30.783396	2.86	7.35	0.0152	0.28	7.62	19.93	20.04	0.0	0.0	0.0	0.0	0.0	2186	180	False
3	23.620705	30.785866	3.25	6.85	0.0152	2.75	9.64	20.66	19.43	0.05	0.05	0.0	0.0	0.0397	778	191	False
4	23.629348	30.790771	2.43	6.85	0.0152	0.15	1.17	18.80	15.64	0.1833	0.1167	0.0167	0.05	0.0478	839	120	False
5	23.624151	30.789617	3.22	8.95	0.0152	0.85	2.09	22.44	21.27	0.1475	0.1475	0.0	0.0	0.1740	676	167	False

Note. Synthetic clusters 1–848 are the first batch of randomly distributed tests, and clusters 849–1696 are the second batch of tests spatially distributed to place young clusters in regions with higher N_{MS} . The boolean “detected” column reflects whether the synthetic cluster is selected by the final cluster catalog selection criteria: $f_{\text{cluster,W}} > 0.674$.

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

Table 7
Background Galaxy Catalog

ID	R.A. (J2000)	DECL. (J2000)	R_{ap} (")	f_{view}	f_{cluster}	f_{galaxy}	f_{emission}	$f_{\text{cluster,W}}$	m_{814}	σ_{814}
24	23.647657	30.805648	6.40	0.9833	0.1000	0.8833	0.0000	0.0855	18.03	0.07
37	23.474286	30.498702	2.10	0.9333	0.3583	0.4667	0.1083	0.4833	18.86	0.01
46	23.363899	30.699947	1.90	0.9667	0.3000	0.5833	0.0833	0.3108	17.69	0.08
108	23.557170	30.482706	2.27	0.9500	0.0833	0.8667	0.0000	0.1116	20.16	0.37
126	23.484892	30.491225	2.64	0.9344	0.0656	0.8525	0.0164	0.1497	17.00	0.05

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

Table 8
Emission Region Catalog

ID	R.A. (J2000)	DECL. (J2000)	R_{ap} (")	f_{view}	f_{cluster}	f_{galaxy}	f_{emission}	$f_{\text{cluster,W}}$
54	23.400054	30.709211	2.79	0.9667	0.2667	0.0000	0.7000	0.3235
57	23.469856	30.730419	3.84	0.9583	0.3583	0.0000	0.6000	0.4373
60	23.527783	30.815464	3.74	0.9667	0.2500	0.0000	0.7167	0.2552
114	23.491020	30.818007	2.37	0.9344	0.3771	0.0000	0.5574	0.4062
163	23.495319	30.819708	2.80	0.9344	0.6167	0.0000	0.3167	0.8384

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

Appendix B

SLUG Integrated Light Fitting: Comparison to CMD-based Fits for M31 Clusters

To determine the reliability of SLUG integrated light fits, we use previous CMD-based cluster fits for M31 clusters observed by PHAT (Johnson et al. 2016) and compare to SLUG results for these clusters. Specifically, we identify a sample of clusters with good CMD fits and best-fit $\log(\text{Age/yr}) < 8.5$. To compute the SLUG fits for the M31 clusters, we adopt the same fitting techniques we use for M33 that are described in Section 6. We use the same model grid for fitting, but we adopt a different assumed distance: 24.47 for M31. We report the SLUG fitting results for a sample of 885 M31 star clusters in Table 9.

We observe that a significant number of M31 SLUG fits have large quoted uncertainties. We consider a SLUG fit to have a large error if its 84th percentile minus 16th percentile value is ≥ 1.2 dex in age or ≥ 1.3 dex in mass. This selection identified 122 clusters from the M31 sample with large SLUG uncertainties. We opt to flag these most uncertain fits in the results table and omit them from the CMD versus SLUG comparison below.

Figure 15 compares results in mass, age, and A_V from the cluster's CMD and SLUG fits. The reliability of the integrated light SLUG values for mass is demonstrated on the top panel. The mass values for previous CMD estimates follow a 1:1 relationship with the new SLUG mass values. The same cannot be clearly stated for the SLUG age and A_V values based on our comparison to CMD derived properties.

The comparison of fitted ages reveals a gap in SLUG ages between $\log(\text{Age/yr})$ of 7.0 and 7.5, whereas the previous CMD ages are consistent with a continuous distribution for the last 300 Myr (Johnson et al. 2016). The gap seems to be due to SLUG fitting clusters in this age range with either younger ($< 10^7$ yr) or older ($> 10^{7.5}$ yr) ages. Given this significant gap in the age distribution, similar to a known artifact in deterministic integrated light fitting at this same age that coincides with the age of emergence for evolved supergiant stars (e.g., see Foesneau & Lançon 2010), we do not consider the SLUG ages to be as reliable as the CMD ages. At older ages ($> 10^{8.5}$ yr), however, the SLUG ages become the sole option for age determination due to the difficulty of fitting CMDs in the case where the main-sequence turnoff lies below the completeness limit for the resolve star photometry catalogs.

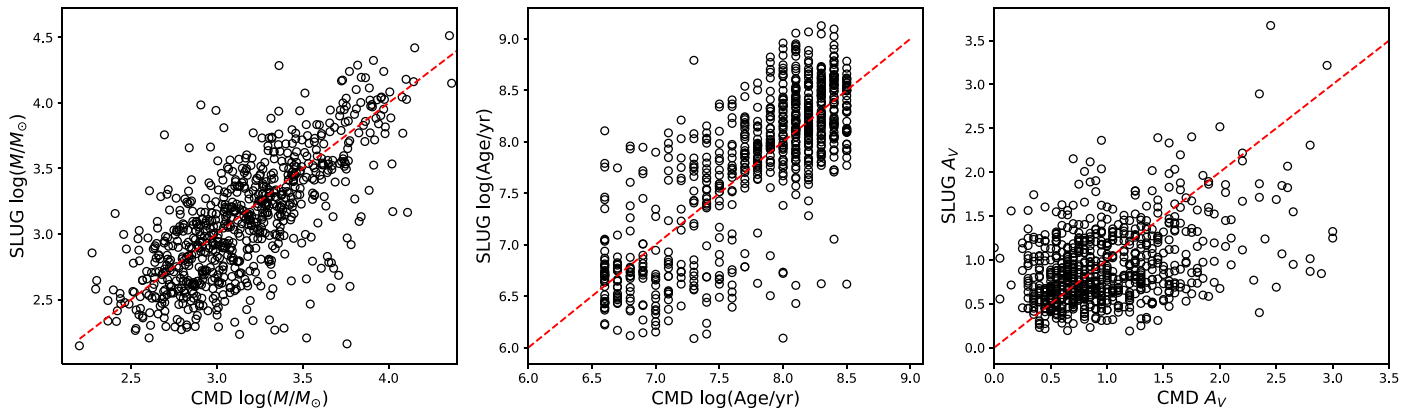


Figure 15. A comparison between the SLUG integrated light fits and CMD-based fits for 763 clusters in M31. The panels show the distribution of results relative to the one-to-one relation (dashed red line) for cluster mass (left), age (middle), and A_V (right).

Table 9
SLUG Results for PHAT M31 Clusters

APID	Error Cut Flag	Filters Available	$\log(\text{Mass}/M_\odot)$			$\log(\text{Age}/\text{yr})$			A_V		
			SLUG ₁₆ CMD ₁₆	SLUG ₅₀ CMD ₅₀	SLUG ₈₄ CMD ₈₄	SLUG ₁₆ CMD ₁₆	SLUG ₅₀ CMD ₅₀	SLUG ₈₄ CMD ₈₄	SLUG ₁₆ CMD ₁₆	SLUG ₅₀ CMD ₅₀	SLUG ₈₄ CMD ₈₄
2	T	4	2.39	3.61	3.78	7.62	8.26	8.70	0.31	0.66	1.16
			3.95	3.98	3.98	8.4	8.4	8.4	1.1	1.15	1.15
5	F	6	3.52	3.62	3.70	8.32	8.55	8.68	0.30	0.54	0.91
			3.40	3.41	3.45	8.3	8.4	8.4	0.44	0.45	0.65
6	F	6	3.66	3.77	3.89	8.17	8.29	8.42	0.39	0.64	0.85
			3.81	3.83	3.83	8.4	8.5	8.5	0.5	0.5	0.65
7	F	4	2.74	3.17	3.34	7.81	8.11	8.29	0.36	0.63	0.98
			3.12	3.16	3.16	7.9	8.2	8.2	0.4	0.45	0.55
14	F	4	3.28	3.77	3.93	7.62	7.99	8.29	0.49	0.81	1.15
			4.08	4.08	4.13	8.2	8.2	8.3	1.15	1.25	1.25

Note. The 16th, 50th, and 84th percentile values extracted for the mass, age, and A_V PDFs from SLUG and CMD fitting results are reported. The filter count of the best available passband combination is listed, along with a boolean flag identifying cases of high uncertainty.

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

ORCID iDs

L. Clifton Johnson <https://orcid.org/0000-0001-6421-0953>
 Tobin M. Wainer <https://orcid.org/0000-0001-6320-2230>
 Estephani E. TorresVillanueva <https://orcid.org/0000-0001-9961-8203>
 Anil C. Seth <https://orcid.org/0000-0003-0248-5470>
 Benjamin F. Williams <https://orcid.org/0000-0002-7502-0597>
 Meredith J. Durbin <https://orcid.org/0000-0001-7531-9815>
 Julianne J. Dalcanton <https://orcid.org/0000-0002-1264-2006>
 Daniel R. Weisz <https://orcid.org/0000-0002-6442-6030>
 Eric F. Bell <https://orcid.org/0000-0002-5564-9873>
 Puragra Guhathakurta <https://orcid.org/0000-0001-8867-4234>
 Evan Skillman <https://orcid.org/0000-0003-0605-8732>
 Adam Smercina <https://orcid.org/0000-0003-2599-7524>

References

Adamo, A., Ryon, J. E., Messa, M., et al. 2017, *ApJ*, **841**, 131
 Adamo, A., Zeidler, P., Kruijssen, J. M. D., et al. 2020, *SSRv*, **216**, 69

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **558**, A33
 Avila, R. J., Hack, W., Cara, M., et al. 2015, in ASP Conf. Ser. 495, Astronomical Data Analysis Software and Systems XXIV (ADASS XXIV), ed. A. R. Taylor & E. Rosolowsky (San Francisco, CA: ASP), 281
 Bedin, L. R., Piotto, G., Baume, G., et al. 2005, *A&A*, **444**, 831
 Beerman, L. C., Johnson, L. C., Fouesneau, M., et al. 2012, *ApJ*, **760**, 104
 Bohlin, R. C. 2016, *AJ*, **152**, 60
 Bressan, A., Marigo, P., Girardi, L., et al. 2012, *MNRAS*, **427**, 127
 Chandar, R., Bianchi, L., & Ford, H. C. 1999, *ApJS*, **122**, 431
 Chandar, R., Bianchi, L., & Ford, H. C. 2001, *A&A*, **366**, 498
 Christian, C. A., & Schommer, R. A. 1982, *ApJS*, **49**, 405
 Ciardullo, R., Durrell, P. R., Laychak, M. B., et al. 2004, *ApJ*, **614**, 167
 Corbelli, E., Braine, J., Bandiera, R., et al. 2017, *A&A*, **601**, A146
 Dalcanton, J. J., Williams, B. F., Lang, D., et al. 2012, *ApJS*, **200**, 18
 de Grijs, R., & Bono, G. 2014, *AJ*, **148**, 17
 de Meulenaer, P., Narbutis, D., Mineikis, T., & Vansevičius, V. 2015, *A&A*, **581**, A111
 Deustua, S. E., Mack, J., Bajaj, V., & Khandrika, H. 2017, WFC3 Instrument Science Report, **2017-14**
 Dolphin, A. E. 2000, *PASP*, **112**, 1383
 Eisner, N. L., Nicholson, B. A., Barragán, O., et al. 2021, *MNRAS*, **505**, 1827
 Elmegreen, B. G., & Efremov, Y. N. 1997, *ApJ*, **480**, 235
 Fan, Z., & de Grijs, R. 2014, *ApJS*, **211**, 22
 Fitzpatrick, E. L. 1999, *PASP*, **111**, 63
 Fouesneau, M., & Lançon, A. 2010, *A&A*, **521**, A22
 Girardi, L., Bressan, A., Bertelli, G., & Chiosi, C. 2000, *A&AS*, **141**, 371

- Girardi, L., Boyer, M. L., Johnson, L. C., et al. 2020, *ApJ*, **901**, 19
- González-Lópezlira, R. A., Pflamm-Altenburg, J., & Kroupa, P. 2012, *ApJ*, **761**, 124
- Grudić, M. Y., Kruijssen, J. M. D., Faucher-Giguère, C.-A., et al. 2021, *MNRAS*, **506**, 3239
- Hack, W. J., Dencheva, N., & Fruchter, A. S. 2013, in ASP Conf. Ser. 475, *Astronomical Data Analysis Software and Systems XXII*, ed. D. N. Friedel (San Francisco, CA: ASP), 49
- Hiltner, W. A. 1960, *ApJ*, **131**, 163
- Jayasinghe, T., Dixon, D., Povich, M. S., et al. 2019, *MNRAS*, **488**, 1141
- Johnson, L. C., Seth, A. C., Dalcanton, J. J., et al. 2012, *ApJ*, **752**, 95
- Johnson, L. C., Seth, A. C., Dalcanton, J. J., et al. 2015, *ApJ*, **802**, 127
- Johnson, L. C., Seth, A. C., Dalcanton, J. J., et al. 2016, *ApJ*, **827**, 33
- Johnson, L. C., Seth, A. C., Dalcanton, J. J., et al. 2017, *ApJ*, **839**, 78
- King, I. 1962, *AJ*, **67**, 471
- Koch, E. W., Rosolowsky, E. W., Lockman, F. J., et al. 2018, *MNRAS*, **479**, 2505
- Kroupa, P. 2001, *MNRAS*, **322**, 231
- Kruijssen, J. M. D. 2012, *MNRAS*, **426**, 3008
- Krumholz, M. R., Adamo, A., Fumagalli, M., & Calzetti, D. 2019, *MNRAS*, **482**, 3550
- Krumholz, M. R., Fumagalli, M., da Silva, R. L., Rendahl, T., & Parra, J. 2015a, *MNRAS*, **452**, 1447
- Krumholz, M. R., Adamo, A., Fumagalli, M., et al. 2015b, *ApJ*, **812**, 147
- Lazzarini, M., Williams, B. F., Durbin, M. J., et al. 2022, *ApJ*, **934**, 76
- Lee, J. C., Whitmore, B. C., Thilker, D. A., et al. 2022, *ApJS*, **258**, 10
- Leroy, A. K., Walter, F., Brinks, E., et al. 2008, *AJ*, **136**, 2782
- Lewis, A. R., Dolphin, A. E., Dalcanton, J. J., et al. 2015, *ApJ*, **805**, 183
- Marigo, P., Girardi, L., Bressan, A., et al. 2017, *ApJ*, **835**, 77
- Massey, P., Olsen, K. A. G., Hodge, P. W., et al. 2006, *AJ*, **131**, 2478
- Melnick, J., & D’Odorico, S. 1978, *A&AS*, **34**, 249
- Mochejska, B. J., Kaluzny, J., Krockenberger, M., Sasselov, D. D., & Stanek, K. Z. 1998, *AcA*, **48**, 455
- Nidever, D. L., Olsen, K., Walker, A. R., et al. 2017, *AJ*, **154**, 199
- Park, W.-K., & Lee, M. G. 2007, *AJ*, **134**, 2168
- Patel, E., Besla, G., & Sohn, S. T. 2017, *MNRAS*, **464**, 3825
- Pflamm-Altenburg, J., González-Lópezlira, R. A., & Kroupa, P. 2013, *MNRAS*, **435**, 2604
- San Roman, I., Sarajedini, A., & Aparicio, A. 2010, *ApJ*, **720**, 1674
- San Roman, I., Sarajedini, A., Garnett, D. R., & Holtzman, J. A. 2009, *ApJ*, **699**, 839
- Sarajedini, A., Barker, M. K., Geisler, D., Harding, P., & Schommer, R. 2007, *AJ*, **133**, 290
- Sarajedini, A., & Mancone, C. L. 2007, *AJ*, **134**, 447
- Senchyna, P., Johnson, L. C., Dalcanton, J. J., et al. 2015, *ApJ*, **813**, 31
- Sharma, S., Corbelli, E., Giovanardi, C., Hunt, L. K., & Palla, F. 2011, *A&A*, **534**, A96
- Spies, H., Swanson, A., Fortson, L., et al. 2019, *J. Sci. Commun.*, **18**, A04
- STSCI Development Team 2012, DrizzlePac: HST Image Software v2.2.6, Astrophysics Source Code Library ascl:1212.011
- Sun, W., de Grijs, R., Fan, Z., & Cameron, E. 2016, *ApJ*, **816**, 9
- Taylor, M. B. 2005, in ASP Conf. Ser. 347, *Astronomical Data Analysis Software and Systems XIV*, ed. P. Shopbell, M. Britton, & R. Ebert (San Francisco, CA: ASP), 29
- Tüllmann, R., Gaetz, T. J., Plucinsky, P. P., et al. 2011, *ApJS*, **193**, 31
- Vázquez, G. A., & Leitherer, C. 2005, *ApJ*, **621**, 695
- Veyette, M. J., Williams, B. F., Dalcanton, J. J., et al. 2014, *ApJ*, **792**, 121
- Wainer, T. M., Johnson, L. C., Seth, A. C., et al. 2022, *ApJ*, **928**, 15
- Weisz, D. R., Johnson, L. C., Foreman-Mackey, D., et al. 2015, *ApJ*, **806**, 198
- Willett, K. W., Lintott, C. J., Bamford, S. P., et al. 2013, *MNRAS*, **435**, 2835
- Williams, B. F., Wold, B., Haberl, F., et al. 2015, *ApJS*, **218**, 9
- Williams, B. F., Durbin, M. J., Dalcanton, J. J., et al. 2021, *ApJS*, **253**, 53