




Research Article

Multi-layered characterisation of hot stellar systems with confidence

Souradeep Chattopadhyay^{1,2} , Steven D. Kawaler³ , and Ranjan Maitra¹ 

¹Department of Statistics, Iowa State University, Ames, IA 50011, USA, ²Department of Mechanical Engineering, Iowa State University, Ames, IA 50011, USA and

³Department of Physics and Astronomy, Iowa State University, Ames, IA 50011, USA

Abstract

Understanding the physical and evolutionary properties of Hot Stellar Systems (HSS) is a major challenge in astronomy. We studied the dataset on 13 456 HSS of Misgeld & Hilker (2011, MNRAS, 414, 3 699) that includes 12 763 candidate globular clusters using stellar mass (M_s), effective radius (R_e) and mass-to-luminosity ratio (M_s/L_v), and found multi-layered homogeneous grouping among these stellar systems. Our methods elicited eight homogeneous ellipsoidal groups at the finest sub-group level. Some of these groups have high overlap and were merged through a multi-phased syncytial algorithm motivated from Almodóvar-Rivera & Maitra (2020, JMLR, 21, 1). Five groups were merged in the first phase, resulting in three complex-structured groups. Our algorithm determined further complex structure and permitted another merging phase, revealing two complex-structured groups at the highest level. A nonparametric bootstrap procedure was also used to estimate the confidence of each of our group assignments. These assignments generally had high confidence in classification, indicating great degree of certainty of the HSS assignments into our complex-structured groups. The physical and kinematic properties of the two groups were assessed in terms of M_s , R_e , surface density and M_s/L_v . The first group consisted of older, smaller and less bright HSS while the second group consisted of brighter and younger HSS. Our analysis provides novel insight into the physical and evolutionary properties of HSS and also helps understand physical and evolutionary properties of candidate globular clusters. Further, the candidate globular clusters (GCs) are seen to have very high chance of really being GCs rather than dwarfs or dwarf ellipticals that are also indicated to be quite distinct from each other.

Keywords: methods: statistical – astronomical instrumentation, methods, and techniques – methods: data analysis – astronomical instrumentation, methods, and techniques – galaxy: globular clusters: general – the galaxy

(Received 5 February 2022; revised 13 May 2022; accepted 17 May 2022)

1. Introduction

Over the past many decades, astronomers have identified commonalities in clusters of stars ranging from small groups to large galaxies. Hot stellar systems (HSS) are a class of celestial objects consisting of globular clusters, nuclear star clusters, compact elliptical galaxies, giant elliptical galaxies, ultra compact dwarf elliptical galaxies, nuclear star clusters and so on. These HSS are very important to understand processes such as the formation of stars or black holes, evolution of galaxies and so on. Indeed, the physical properties of these objects have been directly linked to galaxy interactions and have been extensively studied. One of the most useful concepts (Burstein et al. 1997; Bernardi et al. 2003; Kormendy et al. 2009; Misgeld & Hilker 2011) in studying these objects is the set of fundamental plane relations (Brosche 1973). These planes are typically constructed with parameters such as luminosity, surface brightness, stellar magnitude or central velocity dispersion and help understand important properties of these stellar systems. Different HSS sub-groups typically have different fundamental planes, photometric properties and other types of interpretations (Kormendy 1985; Djorgovski 1995; McLaughlin & van der Marel 2005; Meylan et al 2001; Forbes et al. 2008; Harris, Pritchett, & McClure 1995; Webbink 1985; Ichikawa, Wakamatsu,

& Okamura 1986; Kormendy et al. 2009; Bender et al. 1992). But the origins of these relations, the underlying evolutionary processes involved and the statistical reliability of these results are not well understood and remain as significant challenges. A huge volume of data has been collected and catalogued in the last fifteen years to effectively address these challenges. For example Jordán et al. (2008) catalogued and studied 12 763 candidate globular clusters (GCs) from the Virgo Cluster Survey during the eleventh Hubble Space Telescope observation cycle. A more comprehensive catalogue was compiled by Misgeld & Hilker (2011) which included 693 additional HSS along with those of Jordán et al. (2008). Chattopadhyay & Karmakar (2013) pointed out that several studies have previously compared stellar systems such as globular clusters and dwarf spheroidals using two-point correlations between different projections of the fundamental plane of galaxies. Given our lack of understanding about HSS groups, their evolution and relationships, an important step forward is a firm understanding of which groups are demonstrably different in observational parameter spaces, and which may be more closely related. These important questions can be answered by modern multivariate statistical analysis methods, and one effort to do so was by Chattopadhyay & Karmakar (2013) who used k -means clustering with the jump statistic (Sugar & James 2003) and found five and four homogeneous groups in the 693 non-candidate and the larger 13 456 HSS datasets.

Clustering (Kettenring 2006; Xu & Wunsch 2009; Everitt 2011) is an unsupervised learning technique that groups observations

Corresponding author: Ranjan Maitra, email: maitra@iastate.edu

Cite this article: Chattopadhyay S, Kawaler SD and Maitra R. (2022) Multi-layered characterisation of hot stellar systems with confidence. *Publications of the Astronomical Society of Australia* 39, e029, 1–11. <https://doi.org/10.1017/pasa.2022.25>

without a response variable. While there are many kinds of clustering algorithms, most of them can broadly be categorised into hierarchical and non-hierarchical approaches. Hierarchical clustering algorithms yield a tree-like grouping hierarchy while non-hierarchical algorithms, such as from k -means or model-based clustering (MBC) methods, typically optimise an objective function using iterative greedy algorithms—these algorithms typically require a specified number of groups. The objective function is often multimodal and requires careful initialisation (Maitra 2009). A detailed review on MBC is provided, for instance, in McLachlan & Peel (2000) or Melnykov & Maitra (2010); additionally, Chattopadhyay & Maitra (2017) reviewed it in the context of astronomical applications. Even though model-based methods are an improvement over traditional hierarchical and non-hierarchical methods, they are still mostly limited to finding regular-structured groups. For example, the most common technique of Gaussian-mixtures MBC (GMMBC) assumes spherical or ellipsoidally structured spreads in their groups. Very often, however, the underlying groups in a dataset are irregular in shape and do not all neatly fit the assumptions underlying MBC. Thus advanced methods are necessary to identify and analyse complex-structured groups. Almodóvar-Rivera & Maitra (2020) proposed a novel method of identifying such groups by using *syncytial clustering* which incorporates the results from standard clustering methods and then merges them to reveal the complex general structure in data. Specifically, their approach reveals a multi-layered group structure that provides insight into not just the groups but also their underlying sub-group structure, which may have regular structure at some level. We analyse the 13 456 HSS of Misgeld & Hilker (2011) that is briefly summarised in Section 2 using syncytial clustering methods (introduced in Section 3) with initial clustering results obtained using t -mixtures MBC (t MMBC). We analyse the HSS using these methods in Section 4. Further, in all such investigations involving astronomical data, parameter estimation in statistical clustering algorithms is accompanied by uncertainty in those estimates and careful assessment of these estimates is often required to judge their results and their impact on the obtained groupings. We analyse this uncertainty by using a nonparametric bootstrap procedure (Efron 1979) to calculate the confidence of classification of each data point into a group. The paper concludes with some discussion about the physical properties of the obtained groups and to pointers for future work.

2. The HSS dataset

The dataset used in our analysis was compiled from different sources by Misgeld & Hilker (2011) and contains measurements on 13 456 HSS of different types: globular clusters (GC), giant ellipticals (gE), compact ellipticals (cE), ultra compact dwarf galaxies (UCD), dwarf globular transition objects (DGTO), nuclear of star clusters (NuSc), bulges of spiral galaxy (Sbul) and nuclei of nucleated dwarf galaxies (dE,N). There are 12 763 HSS in this dataset that are candidate GCs (GC_VCC) from the Virgo Globular Cluster survey meaning it is ambiguous if these systems are GC's or not. Chattopadhyay & Karmakar (2013) primarily excluded these candidate GCs from their analysis, focusing mainly on the 673 stellar systems consisting of confirmed GC's and other different types of stellar systems. We call these 673 systems non-candidate HSS. The main parameters in Misgeld & Hilker (2011) are stellar mass (M_s), effective radius (R_e), mass surface

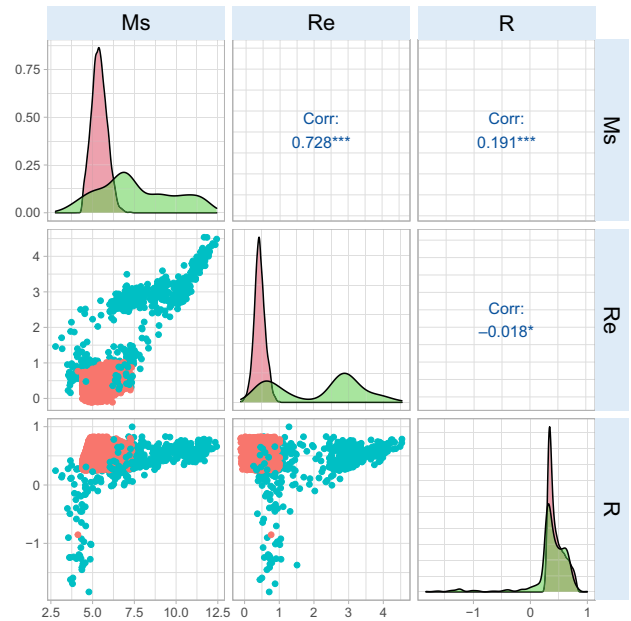


Figure 1. Pairwise scatterplots, estimated densities and correlation coefficients of the logarithm (base 10) of the parameters in the HSS dataset. Here M_s denotes mass, R_e effective radius and R the mass-luminosity ratio. In the scatterplot, orange indicates the non-candidates and blue the candidates. For the density plots the blue curves represent the non-candidates. Note that the calculated correlations are for all 13 456 HSS in the dataset (including the non-candidates and candidates).

density averaged over projected effective radius (S_e) and absolute magnitude in the V band (M_v). Following Chattopadhyay & Karmakar (2013) we use the logarithm (base 10) of M_s , R_e and mass-to-luminosity ratio (M_s/L_v) in solar luminosity ($M_{s,\odot}$) units in our analysis. The M_s/L_v ratio was obtained using the standard magnitude-luminosity relation.

$$\frac{M_s}{M_{s,\odot}} = -2.5 \left(\frac{L_v}{L_{v,\odot}} \right) \quad (1)$$

where $L_{v,\odot}$ denotes the luminosity of the sun. The parameter $\log_{10} S_e$ is taken into account while interpreting the results but, as in Chattopadhyay & Karmakar (2013), is not used in the clustering mechanism. Figure 1 provides pairwise scatterplots of the three parameters for both candidate and non-candidate objects along with density plots of the parameters used in clustering. From Figure 1 we see that the densities of the non-candidates display moderate univariate bimodality, and especially in the context of effective radius and mass-luminosity ratio. There is therefore evidence of grouping, certainly so at the univariate level. Also, the correlation values indicate that M_s and R_e are moderately positively correlated, as expected in stellar objects. But mass-luminosity ratio has a weak linear association with both M_s and R_e . It therefore points to complex structure in the dataset and so in the coming sections, we outline methodology to uncover and understand it from astronomical perspectives.

3. Statistical methodology

In this section we briefly describe our methodology for analysing the 13 456 HSS. Our methodology is a model-based analogue of the syncytial clustering algorithm of Almodóvar-Rivera & Maitra (2020). Our procedure involves an initial clustering of the data

using *t*MMBC followed by merging of the *t*MMBC mixture components using pairwise overlap measure and generalised overlap, calculated using Monte Carlo methods, as described next.

3.1. Initial MBC using a mixture of multivariate-*t* densities

MBC provides a principled way to find homogeneous regular-shaped groups in a given dataset. It scores over classical clustering algorithms like *k*-means due to its ability to better model heterogeneity in groups. As pointed out by Chattopadhyay & Maitra (2017) assuming spherically dispersed homogeneous groups when such assumption is not valid can lead to erroneous results. In MBC (Melnikov & Maitra 2010; McLachlan & Peel 2000) the observations X_1, X_2, \dots, X_N are assumed to be realisations from a *G*-component mixture model (McLachlan & Peel 2000) with probability density function (PDF)

$$f(x; \theta) = \sum_{g=1}^G \pi_g f_g(x; \eta_g) \tag{2}$$

where $f_g(\cdot; \eta_g)$ is the density of the *g*th group, η_g the vector of unknown parameters and $\pi_g = Pr\{x_i \in \mathcal{G}_g\}$ is the mixing proportion of the *g*th group, $g = 1, 2, \dots, G$, and $\sum_{g=1}^G \pi_g = 1$. The component density $f_g(\cdot; \eta_g)$ can be chosen according to the specific needs of the application—the most popular choice is the multivariate Gaussian density (Chattopadhyay & Maitra 2017; Fraley & Raftery 1998, 2002). Another useful family of mixture models proposed by McLachlan & Peel (1998) specifies $f_g(\cdot; \eta_g)$ to be the multivariate-*t* density. That is,

$$f_g(z; \mu_g, \Sigma_g, \nu_g) = \frac{\Gamma(\nu_g/2 + p/2)}{\Gamma(\nu_g/2) \nu_g^{p/2} \pi_g^{p/2} |\Sigma_g|^{1/2}} \left[1 + \frac{1}{\nu_g} (z - \mu_g)^T \Sigma_g^{-1} (z - \mu_g) \right], \tag{3}$$

for $z \in \mathbb{R}^p$, where μ_g denotes the mean vector, Σ_g the scale matrix and ν_g the degrees of freedom, all for the *g*th mixture component, $g = 1, 2, \dots, G$. Our analysis uses, instead of a Gaussian mixture model (GMM), the multivariate-*t* mixture model (*t*MM) (Goren & Maitra 2022) because the multivariate-*t*MM better allows for thicker tails in the component mixture densities. Subsequent steps in clustering involve obtaining maximum likelihood (ML) estimates of the parameters $\eta_g, g = 1, 2, \dots, G$ using the Expectation-Maximisation (EM) algorithm (Dempster, Laird, & Rubin 1977; McLachlan & Krishnan 2008; Chattopadhyay & Maitra 2017; Chattopadhyay & Maitra 2018) and assigning each individual observation based on the maximum posterior probability that it belongs to a given group. We use the Bayesian Information Criterion (BIC) (Schwarz 1978; Chattopadhyay & Maitra 2017) to decide *G*.

3.2. Merging mixture components using overlap

After obtaining the parameter estimates of (2), mixture components can be merged using one of several criteria described in Melnikov (2016). Here we use the pairwise overlap measure (Maitra & Melnikov 2010; Melnikov & Maitra 2011; Melnikov, Chen, & Maitra 2012). The pairwise overlap takes value in [0, 1] and provides us with a sense of the distinctiveness between groups obtained using a particular clustering method, with values closer to 1 indicating that the two groups are poorly separated. We now provide details on how to obtain these overlap measures.

3.2.1. Pairwise overlap between simple groups

A simple group (Almodóvar-Rivera & Maitra 2020) is obtained from the initial classifications, in our case by *t*MMBC. For a *G*-component *t*MM (2), the probability that an observation *X*, originally from a *m*th component distribution (that is, belonging to the *m*th group) gets misclassified to the *n*th group is

$$\omega_{n|m} = P(\pi_n f_n(x; \mu_n, \Sigma_n, \nu_n) > \pi_m f_m(x; \mu_m, \Sigma_m, \nu_m)) \tag{4}$$

where $f_g(x; \mu_g, \Sigma_g, \nu_g)$ has density as in (3). The pairwise overlap (Maitra & Melnikov 2010) between the *m*th and *n*th group is then defined as

$$\omega_{mn} = \omega_{n|m} + \omega_{m|n} \tag{5}$$

For illustration, we display in the appendix (Figure A.1) sample two-component three-dimensional distributions with varying degrees of overlap.

Unlike in the case of GMM, the misclassification probability $\omega_{n|m}$ cannot be readily calculated using analytical methods, so we use Monte Carlo methods as follows:

1. Generate *M* realisations $x_i^*, i = 1, 2, \dots, M$ from the density $f_m(x; \mu_m, \Sigma_m, \nu_m)$.
2. $\omega_{n|m}$ can then be approximated as

$$\hat{\omega}_{n|m} = \frac{1}{M} \sum_{i=1}^M I\{\pi_n f_n(x_i^*; \mu_n, \Sigma_n, \nu_n) > \pi_m f_m(x_i^*; \mu_m, \Sigma_m, \nu_m)\} \tag{6}$$

where $I(\cdot)$ is the indicator function, that in this case is 1 for a given *i* if $\pi_n f_n(x_i^*; \mu_n, \Sigma_n, \nu_n) > \pi_m f_m(x_i^*; \mu_m, \Sigma_m, \nu_m)$ and 0 otherwise.

The pairwise overlap provides us with a sense of the distinctiveness between groups obtained using a particular clustering method. As mentioned earlier, these values lie inside [0, 1] with values near zero indicating perfect separation between groups and those closer to unity indicating poor separation. Melnikov & Maitra (2011) used methods developed in Maitra (2010) to define the generalised overlap $\tilde{\omega}$ to summarise the $G \times G$ matrix Ω of pairwise overlaps which is given by $(\lambda_\Omega - 1)/(G - 1)$ where λ_Ω is the largest eigenvalue of Ω . Typically, smaller values of $\tilde{\omega}$ indicate distinctive groupings while larger values indicate more overlap between groups.

3.2.2. Pairwise overlap between composite groups

Let \mathcal{G}_p and \mathcal{G}_q be two composite groups having probability densities $\sum_{h \in \mathcal{G}_p} \pi_h f_h(x; \mu_h, \Sigma_h, \nu_h)$ and $\sum_{r \in \mathcal{G}_q} \pi_r f_r(x; \mu_r, \Sigma_r, \nu_r)$ respectively, possibly formed by merging one or more components of (2). The probability of an observation *X* actually from \mathcal{G}_p being misclassified to \mathcal{G}_q is

$$\omega_{\mathcal{G}_q|\mathcal{G}_p} = P\left(\frac{\sum_{r \in \mathcal{G}_q} \pi_r f_r(x; \mu_r, \Sigma_r, \nu_r)}{\sum_{h \in \mathcal{G}_p} \pi_h f_h(x; \mu_h, \Sigma_h, \nu_h)} > 1\right) \tag{7}$$

where $f_g(x; \mu_g, \Sigma_g, \nu_g)$ has density given by (3). The pairwise overlap between \mathcal{G}_p and \mathcal{G}_q , denoted by $\omega_{\mathcal{G}_p\mathcal{G}_q}$, is then computed as $\omega_{\mathcal{G}_q|\mathcal{G}_p} + \omega_{\mathcal{G}_p|\mathcal{G}_q}$. The misclassification probability $\omega_{\mathcal{G}_q|\mathcal{G}_p}$ can be estimated as follows:

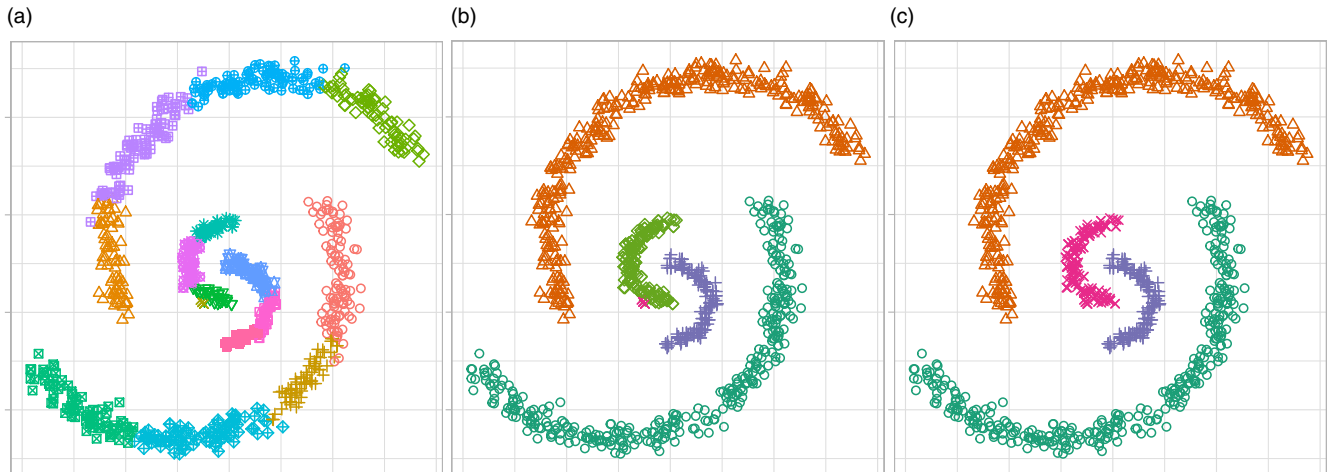


Figure 2. From left to right: (a) Original Gaussian mixture model clustering solution according to BIC. (b) Five clusters obtained after first phase merging. (c) Final clustering solution obtained after second phase merging.

1. Generate M samples x_i^* , $i = 1, 2, \dots, M$ from the density $\sum_{h \in \mathcal{G}_p} \pi_h^* f_r(x; \mu_h, \Sigma_h, \nu_h)$, where π_h^* is a standardised probability obtained as $\pi_h^* = \pi_h / \sum_{h \in \mathcal{G}_p} \pi_h$.
2. $\omega_{\mathcal{G}_q | \mathcal{G}_p}$ can then be approximated as

$$\hat{\omega}_{\mathcal{G}_q | \mathcal{G}_p} = \frac{1}{M} \sum_{i=1}^M I \left\{ \frac{\sum_{r \in \mathcal{G}_q} \pi_r f_r(x_i^*; \mu_r, \Sigma_r, \nu_r)}{\sum_{h \in \mathcal{G}_p} \pi_h f_h(x_i^*; \mu_h, \Sigma_h, \nu_h)} > 1 \right\}, \quad (8)$$

where $I(\cdot)$ is the indicator function, as before.

3.3. The model-based syncytial clustering algorithm

Our syncytial clustering algorithm consists of three phases, the t MMBC phase, the initial overlap calculation phase and the merging phase:

1. *The initial clustering phase:* This phase fits a G -component t MM to the data using the EM algorithm (McLachlan & Peel 2000), with G chosen using BIC (Schwarz 1978).
2. *Merging Phase:* This phase gets triggered only if the generalised overlap $\hat{\omega}^{(1)}$ of the original clustering solution is not negligible, that is if $\hat{\omega}^{(1)} \not\approx 0$ and at least one pairwise overlap of the original clustering solution is greater than $\hat{\omega}^{(1)}$. Specifically this phase involves the following steps:
 - (a) Find the pairwise overlaps between i th and j th group ω_{ij} for the $G(G-1)/2$ pairs of groups using the method described in Section 3.2.1. Also find the generalised overlap $\hat{\omega}$.
 - (b) Merge the i th and j th groups if $\omega_{ij} > \kappa \hat{\omega}$, with κ as discussed shortly, and relabel the merged groups, as needed.
 - (c) Calculate the updated pairwise overlaps of the newly formed composite groups using the method described in Section 3.2.2. Also calculate $\hat{\omega}$ of the new clustering solution. Repeat Step (b) with the updated pairwise overlaps and the generalised overlap.
3. *Termination Phase:* Terminate the merging in Step 2. if the generalised overlap of the current phase clustering solution, or its change, is negligible (in this paper defined to be $\leq 10^{-3}$): that

is, if $\hat{\omega}^{(l)} \approx \hat{\omega}^{(l-1)}$ or $\hat{\omega}^{(l)}$ where $\hat{\omega}^{(l-1)}$ is the generalised overlap of the previous phase clustering solution.

Selection of κ : The parameter κ determines the propensity of merging, and hence the characteristics of the composite groups formed at each merging phase. With larger values of κ , few pairs merge at each phase while smaller values of κ mean that many components are merged simultaneously at each phase. A data-driven approach proposed by Almodóvar-Rivera & Maitra (2020) for selecting κ , that we also adopt, runs the algorithm for several values of κ and uses the final clustering solution with the smallest terminating $\hat{\omega}$.

Choice of mixture model: The description of our algorithm uses t MMBC in the initial clustering phase, however, our algorithm applies to other mixture models with appropriate modifications. Indeed, in our paper, we have used t MMBC (which by default, includes GMMBC, when the degrees of freedom for each mixture component are infinite), with degrees of freedom also estimated, and then decided on the initial clustering solution with the higher BIC.

3.4. An illustrative example

We illustrate our model-based syncytial clustering algorithm on the synthetic 2D *Bananas-Arcs* dataset in Almodóvar-Rivera & Maitra (2020). From Figure 2 it is evident that this dataset has four complex-structured clusters. We fit both GMMBC and t MMBC to this dataset using the methods in Chattopadhyay & Maitra (2017) and (2018) and selected the best clustering solution as well as the number of groups using BIC. For this dataset, GMMBC with 15 groups provided the best solution (Figure 2a) with $\hat{\omega} = 8.2 \times 10^{-3}$. The first of the merging phases (with the best $\kappa = 3$) provided the five clusters of Figure 2b with $\hat{\omega} = 1.4 \times 10^{-4}$. Therefore this merged solution provides a more distinctive grouping than the Gaussian mixture model solution. The second phase of merging resulted in the four groups of Figure 2c with $\hat{\omega} = 1.368 \times 10^{-4}$. Since the generalised overlap of this phase is nearly indistinguishable from that obtained at the end of the first phase of merging, the merging terminates here. This example clearly illustrates

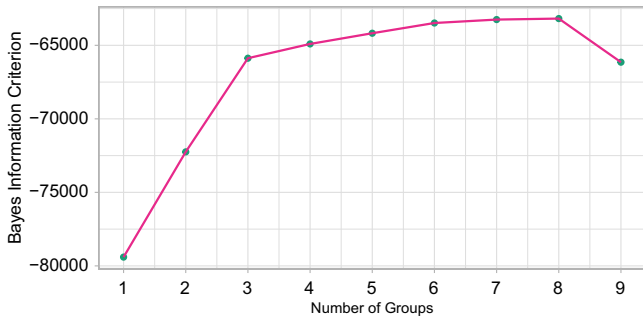


Figure 3. BIC for each G upon performing G -component t MMBC with 13 456 HSS from Misgeld & Hilker (2011).

the effectiveness of syncytial clustering over regular methods in capturing complex structure and irregular-shaped clusters. Additionally, the multi-staged solution provides us with the ability to understand each of these complex structures in terms of their component (simpler) homogeneous groups. Our data-driven approach results in a terminating $\dot{\omega}$ that is much smaller than the original t MMBC solution indicating a preference for the complex-structured solution in providing distinctive groups. We now analyse the 13 456 HSS using our syncytial clustering algorithm.

4. Characterisation of HSS

4.1. Cluster analysis

4.1.1. The initial clustering phase

We performed G -component t MMBC, for $G \in \{1, 2, \dots, 9\}$ on M_s , R_e and M_s/L_v , all in the \log_{10} -scale, of the 13 456 HSS. Figure 3 indicates that a eight-component t MMBC provides an optimal fit, as per BIC. The eight-component t MMBC solution was better, as per BIC, than the optimal GMMBC solution. Our results here are different from that of Chattopadhyay & Karmakar (2013) who, upon using k -means with the Jump statistic (Sugar & James 2003), found five groups when excluding the candidate GCs of Jordán et al. (2008) and four groups on the full HSS dataset.

Figure 4 shows the pairwise overlaps of the eight groups obtained using t MMBC. A 3D scatterplot of the eight t MMBC groups is also provided in Figure 5. Table 1 provides the group means and standard deviations for the eight groups. The composition of the eight groups in terms of the different kinds of stellar objects are in Table 2. Since t MMBC assumes ellipsoidally dispersed groups (with fatter tails) some of the groups exhibit substantial overlap (Figure 4). These large overlap values point to the possibility of complex group structure in the data that is not accounted for by using t MMBC (or GMMBC). We therefore explore if we can use our algorithm to reveal this potentially complex structure.

4.1.2. The merging phases

The first phase: Based on the pairwise overlap map of Figure 4, and for $\kappa = 1$ (determined to be the solution with the lowest terminating $\dot{\omega}$), Groups 1-4, 7 and 8 all merge into one group. Thus, at the end of this merging phase, we have three groups and a generalised overlap of $\dot{\omega} = 0.13$. To facilitate easy reference, we relabel the erstwhile Groups 5 and 6 as the merged (first phase) Groups (i)-5 and (ii)-6 respectively and the merged entity as Group (iii), with the component individual groups as Group (iii)-1 through Group (iii)-4 and Group (iii)-7 and Group (iii)-8. A

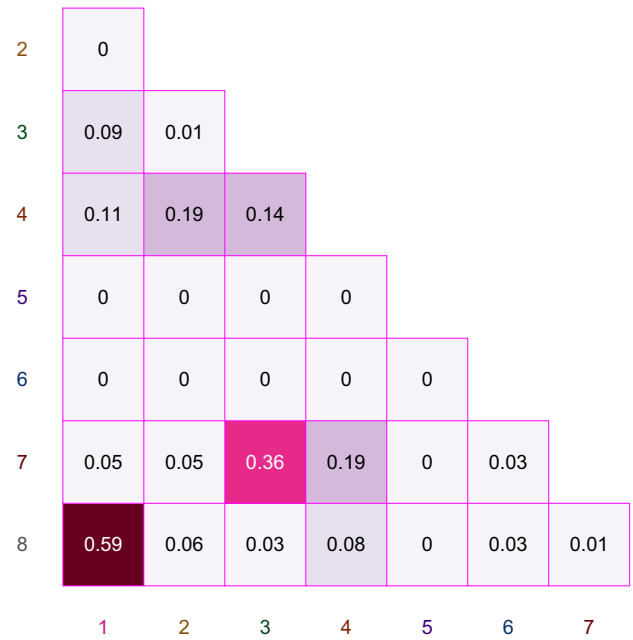


Figure 4. Pairwise overlap measures between any two groups obtained by our eight-component t MMBC solutions.

3D scatterplot of the three first-phase merged groups is provided in Figure 6, while Tables 3 and 4 summarise the mean parameter values and the types of objects in the three merged groups. The pairwise overlap between Groups (i) and (ii) is 0.01 and between Groups (ii) and (iii) is 0.1. the overlap between Groups (i) and (iii) is negligible. Figure 6 immediately points out how the first phase mergers based on the pairwise overlaps have been able to capture the non-ellipsoidal structure present in the data. But the pairwise overlaps and the generalised overlap (and the logic of our method) suggest that there might be additional structure which might have been missed in the first phase merge. So we proceed with another merging phase.

The second phase: In the second phase, Groups (ii)-6 and (iii) merge to yield two complex-structured groups. We label the newly-merged group as Group II—with component merged sub-groups as Groups II-(ii) and II-(iii), and correspondingly down the hierarchy—and the original group as Group I (with sub-groups classification of Group I-(i) and Group I-(i)-5, to indicate the three-stages of classifications). A 3D scatterplot of the two new groups is given in Figure 7, with numerical summaries provided in Tables 5 and 6. The scatterplot clearly depicts that the second merge has captured the complex general-shaped group structures in the dataset well and also obtained to a good extent the two well-separated groups. The pairwise overlap between these two groups is 0.0013. Since the pairwise overlap between Groups I and II is very close to 10^{-3} the algorithm terminates at this stage.

4.2. Measuring uncertainty in the obtained groupings through bootstrapping

We assess the uncertainty in the parameter estimates in each phase and the corresponding classification using a nonparametric bootstrap technique where we repeatedly resample from the dataset with replacement and analyse each resample to obtain estimated

Table 1. Mean and standard deviation (in brackets) of parameter values for each of the eight groups obtained using *t*MMBC.

| Group | M_s | R_e | M_v | M_v/L_v | S_e |
|-------|---------------|---------------|-----------------|----------------|---------------|
| 1 | 6.184 (0.539) | 0.331 (0.194) | -9.107 (1.510) | 0.609 (0.119) | 4.723 (0.321) |
| 2 | 5.134 (0.361) | 0.641 (0.117) | -6.432 (0.844) | 0.630 (0.113) | 3.053 (0.211) |
| 3 | 5.744 (0.294) | 0.408 (0.057) | -8.609 (0.715) | 0.369 (0.026) | 4.131 (0.3) |
| 4 | 5.100 (0.326) | 0.470 (0.153) | -6.905 (0.783) | 0.406 (0.044) | 3.362 (0.387) |
| 5 | 8.817 (1.939) | 3.032 (0.501) | -16.049 (4.546) | 0.465 (0.153) | 1.954 (1.282) |
| 6 | 5.803 (1.442) | 0.991 (0.345) | -9.711 (2.515) | -0.013 (0.707) | 3.023 (1.342) |
| 7 | 5.313 (0.445) | 0.475 (0.164) | -7.651 (1.090) | 0.320 (0.022) | 3.565 (0.452) |
| 8 | 5.556 (0.363) | 0.348 (0.120) | -7.577 (0.987) | 0.593 (0.098) | 4.062 (0.405) |

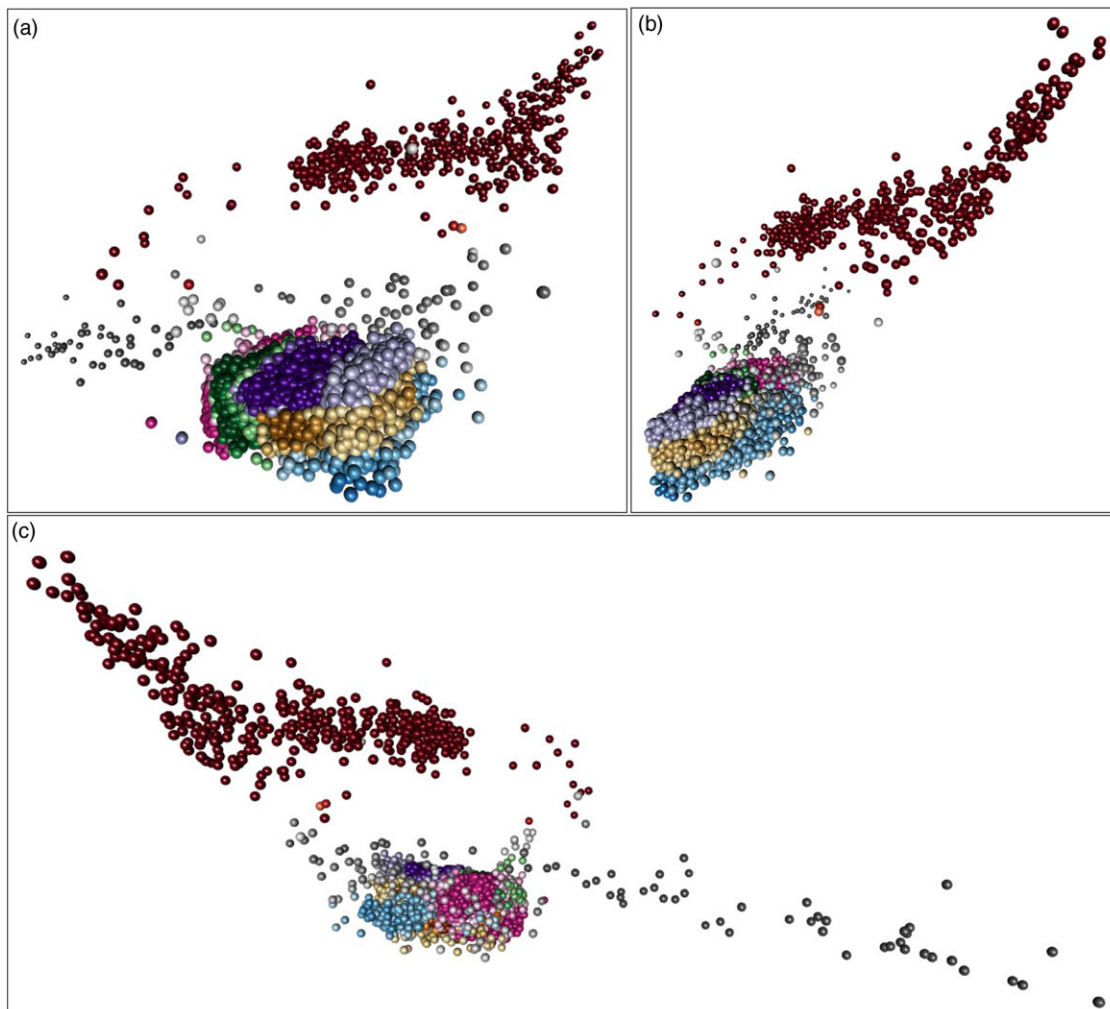


Figure 5. Three viewing angles of the scatterplot of the full dataset with different colours representing different groups, and intensity of the colour indicating the underlying confidence in that particular grouping. Darker shades indicate higher confidence of classification of that particular object.

properties of the parameters of interest. Refer to Efron (1979) for a detailed study on the bootstrap.

In our scenario, the parameters of interest are the group memberships. Here, a well-done cluster analysis is expected to produce the same group memberships for most of the observations at each bootstrap replicate. We now describe the steps to obtain the

confidence of classification of each observation for the original clustering solution using *t*MMBC.

- (i) First we perform *t*MMBC on the original 13 456 HSS as described in Section 2. This matches the analysis so far and yields eight *t*MMBC groups.

Table 2. Types of objects in each of the eight groups obtained by tMMBC.

| | Group 1 | Group 2 | Group 3 | Group 4 | Group 5 | Group 6 | Group 7 | Group 8 |
|--------|---------|---------|---------|---------|---------|---------|---------|---------|
| cE | 1 | | | | 7 | | | |
| dE | | | | | 212 | | | |
| dE,N | 11 | | 4 | | | 15 | 14 | 1 |
| DGTO | | | | | 4 | | 1 | |
| Dwarf | | | | | 22 | 2 | | |
| GC | 1 | 1 | 5 | 17 | 1 | 58 | 65 | 5 |
| GC_VCC | 482 | 872 | 1516 | 2814 | | 10 | 3660 | 3409 |
| gE | | | | | 150 | | | |
| NuSc | 6 | | | | 1 | 12 | | 3 |
| Sbul | | | | | 18 | | | |
| UCD | 14 | | | | 3 | 30 | 8 | 1 |
| Total | 515 | 873 | 1525 | 2831 | 418 | 127 | 3748 | 3419 |

Table 3. Mean and standard deviation (in brackets) of parameter values for each of the three groups after phase one merging. The third group is the merged group.

| Group | M_ζ | R_e | M_v | M_v/L_v | S_e |
|--------|---------------|---------------|-----------------|----------------|---------------|
| (i)-5 | 8.817 (1.939) | 3.032 (0.501) | -16.049 (4.546) | 0.465 (0.153) | 1.954 (1.282) |
| (ii)-6 | 5.803 (1.442) | 0.991 (0.345) | -9.711 (2.515) | -0.013 (0.707) | 3.023 (1.342) |
| (iii) | 5.404 (0.467) | 0.438 (0.160) | -7.557 (1.159) | 0.450 (0.139) | 3.731 (0.556) |

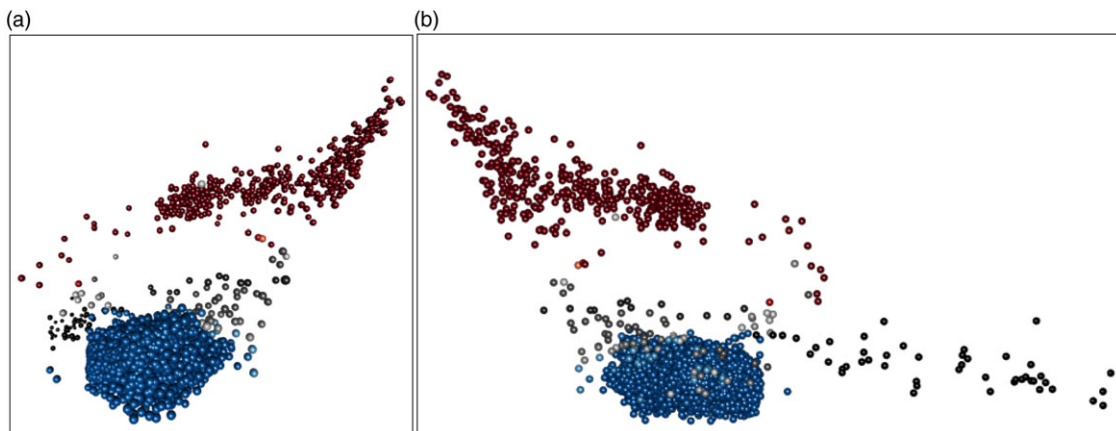


Figure 6. Two viewing angles of the scatterplot of the full dataset after the first stage merging. Here, different colours representing different groups: intensity of the colour indicates the degree of underlying confidence in that particular grouping. Darker shades indicate higher confidence of classification of that particular object.

- (ii) Sample, with replacement, 13 456 HSS from the dataset and again cluster the resampled data using the classifications obtained in Step (i) as the initial estimates for the groupings. Repeat this procedure B times to obtain B sets of classification estimates. Let us denote these B sets by C_1, C_2, \dots, C_B . Also let C_{ij} be 1 if the classification of the j th data point in the i th bootstrap replicate is the same as the original classification of the j th datapoint and 0 otherwise, where $i = 1, 2, \dots, B$ and $j = 1, 2, \dots, 13\,456$.
- (iii) The classification probability for the j th HSS is obtained as

$$p_j = \frac{1}{B} \sum_{i=1}^B C_{ij} \tag{9}$$

where $j = 1, 2, \dots, 13\,456$. In our analysis, we took $B = 1\,000$, that is, we resampled 1 000 times from the dataset.

In each merging phase, the confidence of classification is obtained by relabeling each of the B bootstrap replicates to the solution obtained from the original clustering results and then repeating Steps (i)-(iii) of our algorithm with the relabelled bootstrap replicates. Table 7 gives the number of HSS in each class having confidences in seven intervals (0, 0.30], (0.30, 0.60], (0.60, 0.85], (0.85, 0.90], (0.90, 0.95] and (0.95, 1], at the two merging stages. The fact that the majority of HSS have a high confidence of classification in both the stages provides reassurance in our clustering results for the 13 456 HSS.

We end our discussion in this section by noting that we could have also obtained an estimate of the confidence of classification

Table 4. Types of objects in each of the three groups after phase one merge.

| | Group (i)-5 | Group (ii)-6 | Group (iii) |
|--------|-------------|--------------|-------------|
| cE | 6 | | 1 |
| dE | 212 | | |
| dE,N | | 15 | 30 |
| DGTO | 4 | | 1 |
| Dwarf | 22 | 2 | |
| GC | 2 | 58 | 94 |
| GC_VCC | | 10 | 12 753 |
| gE | 150 | | |
| NuSc | 1 | 12 | 9 |
| Sbul | 18 | | |
| UCD | 3 | 30 | 23 |
| Total | 418 | 127 | 12 911 |

Table 5. Mean and standard deviation (in bracket) of parameter values for each of the two groups after phase two merging. The second group is the new merged group.

| Group | M_s | R_e | M_v | M_v/L_v | S_e |
|-------|---------------|---------------|-----------------|---------------|---------------|
| I | 8.617 (1.939) | 3.032 (0.501) | -16.049 (4.546) | 0.465 (0.153) | 1.954 (1.282) |
| II | 5.408 (0.487) | 0.443 (0.172) | -7.577 (1.199) | 0.445 (0.161) | 3.724 (0.573) |

Table 6. Types of objects in each of the two groups after the phase two mergers.

| | Group I | Group II |
|--------|---------|----------|
| cE | 6 | 1 |
| dE | 212 | |
| dE,N | | 45 |
| DGTO | 4 | 1 |
| Dwarf | 22 | 2 |
| GC | 2 | 152 |
| GC_VCC | 0 | 12 763 |
| gE | 150 | |
| NuSc | 1 | 21 |
| Sbul | 18 | |
| UCD | 3 | 53 |
| Total | 418 | 13 038 |

based on the posterior probability of classification. However, we use a nonparametric bootstrap technique for estimating the classification probabilities in order to account for the uncertainty in the modelling and to account for possible model misspecification. The choice of the nonparametric bootstrap technique provides us with more robust estimates of the confidence of classification compared to using the posterior probability of classification because it accounts for possible model misspecification in the calculation of the classification confidence.

4.3. Analysis of results

4.3.1. Properties of identified groups

Table 2 gives the numbers of different types of HSS in each of the *t*MMBC groups and Tables 4, 6 give the numbers of different types of HSS in each group at each phase of the algorithm. The means and standard deviations of the parameters for the *t*MMBC solution are provided in Table 1 while those for each group at each of the merging phases is in Tables 3 and 5. Additionally, different colours in Figures 5, 6 and 7 indicate the group to which each HSS belongs, with different shades indicating the confidence of that particular HSS. Darker shades indicate higher confidences for that particular HSS. From the figures and Table 7a and b it is clear that most of the HSS have high confidence coefficients, indicating that the classification method has worked well in identifying the non-ellipsoidal structure that was demonstrated after the initial clustering using *t*MMBC. After the first mergers, the most notable groups are Groups (i)-5 and (iii) which have most of the HSS. Most of the GC_VCCs are in Group (iii) after the first merge and only ten are in Group (ii)-6. The stellar systems in Group (ii)-6 have lower confidence compared to the other two groups due to which inferences about the HSS in this group will be more uncertain compared to the HSS in other groups. From Table 6 we see that after the second phase, all the candidates (GC_VCC) were classified to Group II. All the GCs were also classified to this group, as also a number of UCDs and dE,Ns. These group compositions and means can help us infer about the kinematic properties of these stellar systems. After the final merge, Group I has a larger M_s and a much larger R_e than Group II. The effective magnitude is also significantly larger for the second group. The M_s/L_v ratio is quite similar for both groups. We now interpret our results.

4.3.2. Interpretations

In this section, we understand the physical and evolutionary properties of the objects in each of the two groups after the final merge using stellar mass surface density and absolute magnitude. To analyse the two complex-structured groups at the end of the final merge, we carefully look at the properties of the ellipsoidal groups obtained by *t*MMBC and group structures obtained with the original eight-group clustering solution and after the two mergers, and obtain deeper insight into the evolutionary and physical properties of these celestial objects.

From Table 2 it is evident that Group 7 is the largest group obtained by *t*MMBC and also contains the maximum number of GC_VCCs. Table 1 indicates that the HSS in this group have moderate mass and a low effective radius. The surface density is towards the higher side compared to other groups like Group 5 and this group also has a moderate mass-luminosity ratio. Thus intuition dictates that the HSS in this group are bright and newer compared to other groups like Groups 1, 4 and 8. Indeed, the last group, which is mostly composed of GC_VCCs has a higher mass-luminosity ratio and surface density compared to Group 7 indicating that the HSS in these groups are older compared to that of Group 7 and also brighter. But the HSS in this group are typically smaller compared to those of Group 7 and also heavier which gives some indication as to why the HSS in this group are brighter even though they are older compared to those of Group 7. Group 4 which is composed of only globular clusters (both candidate and non-candidate HSS) have properties similar to those HSS

Table 7. Number of candidate HSS (C) and non-candidate HSS (NC) having confidences in each of the seven intervals for each of the (a) three groups after the first merge and (b) two groups after the second merge. Here $(\alpha, \beta]$ denotes the interval with left endpoint α (not included) and right endpoint β (included). Entries in the table are left blank when there are no members in that group.

| (a) Number of candidates (C) and non-candidates (NC) in having confidences after the first merge. | | | | | | | | | | | | | | |
|---|-----------|---|--------------|---|--------------|---|--------------|---|--------------|---|--------------|---|-----------|--------|
| Group | (0, 0.30] | | (0.30, 0.60] | | (0.60, 0.80] | | (0.80, 0.85] | | (0.85, 0.90] | | (0.90, 0.95] | | (0.95, 1] | |
| | NC | C | NC | C | NC | C | NC | C | NC | C | NC | C | NC | C |
| (i)-5 | | | | | 1 | | | | 1 | | 1 | | 415 | |
| (ii)-6 | | | 8 | 5 | 13 | 2 | 11 | 2 | 28 | | 12 | | 46 | 1 |
| (iii) | | | 5 | 4 | 7 | 3 | 1 | 3 | 3 | 3 | 4 | 3 | 138 | 12 737 |

| (b) Number of candidates (C) and non-candidates (NC) in having confidences after the second merge. | | | | | | | | | | | | | | |
|--|-----------|---|--------------|---|--------------|---|--------------|---|--------------|---|--------------|---|-----------|--------|
| Group | (0, 0.30] | | (0.30, 0.60] | | (0.60, 0.80] | | (0.80, 0.85] | | (0.85, 0.90] | | (0.90, 0.95] | | (0.95, 1] | |
| | NC | C | NC | C | NC | C | NC | C | NC | C | NC | C | NC | C |
| I | | | | | 1 | | | | 1 | | 1 | | 415 | |
| II | | | 13 | 9 | 20 | 5 | 12 | 4 | 31 | 3 | 16 | 3 | 184 | 12 738 |

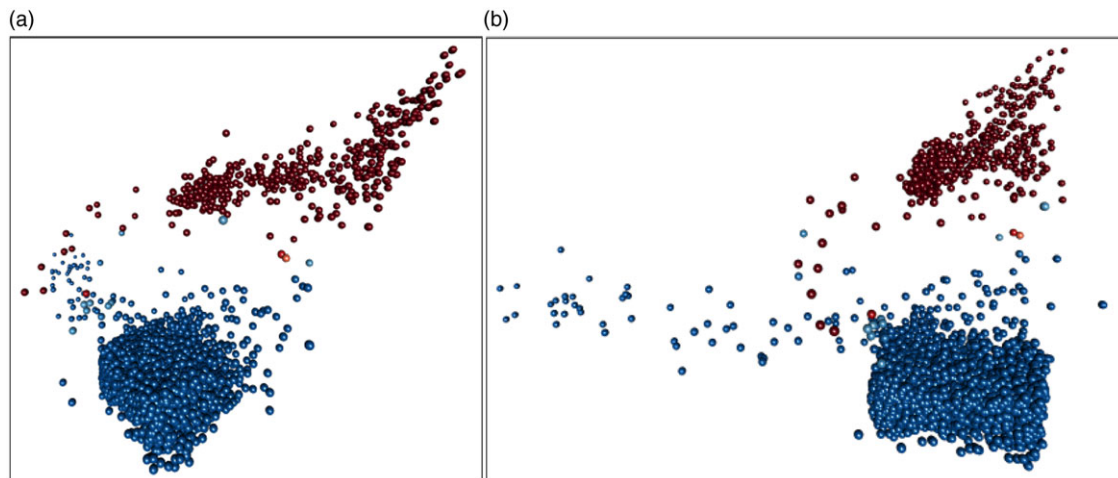


Figure 7. Two viewing angles of the scatterplot of the entire dataset after the final merging stage with different colours representing different groups and intensity of the colour signifying the underlying confidence in that particular grouping. Darker shades indicate higher confidence of classification of that particular HSS.

in Group 7. The same can be said about Group 3. This similarity in terms of kinematic properties also indicates this group having a high overlap as seen in Figure 4. Group 1 also exhibits similarity with Group 8 with respect to kinematic properties. The HSS in Group 2 exhibit kinematic properties that are close to that of Group 4 but are somewhat smaller since they have a higher mass-luminosity ratio than that in Group 4. Thus it is very clear that many of the eight *t*MMBC groups exhibit a good degree of similarity in terms of kinematic properties which satisfies the intuition of many of the groups getting merged to reveal only two complex-structured groups with a low overlap. We now study the kinematic properties of the complex groups obtained at each merging phase.

At the end of the final merge, Group 1 is essentially Group 5 of the *t*MMBC solution. From Table 5, we see that the HSS in this group have higher mass and effective radius compared to the second group. The mass-luminosity ratio for this group is slightly higher compared to the second group which indicates that the stellar systems in this group have probably lost most of their massive stars and are mostly composed of low-mass stars, further indicating that these HSS are typically older than those in the second

group. Since this group is similar to Group 5 from the *t*MMBC solution, and did not change during the merging phases, we conclude that there is no complex structure present in this group as compared to Group II. The conclusions for this group might be a little uncertain for one stellar object that has a low confidence of classification.

Group II is much larger in size than Group I and as mentioned earlier, contains all the GC_VCCs and GCs, dE,Ns and most of the UCDs, with many of the HSS assignments having low confidence of classification, mainly because this group is formed by merging Groups (ii)-6 and (iii). So, it would be more helpful to analyse this group by separately looking at the properties of Groups (ii)-6 and (iii) to get a better understanding of the properties of this group.

For Group (iii) at the end of the first merging phase, it can be inferred that the GC_VCCs in this group (and also Group (ii)-6) are typically smaller HSS compared to that of Group I (or Group (i)-5). Also the HSS in this group have a higher surface density compared to the other groups formed after the first phase, indicating that these objects are typically brighter compared to objects

in the first group. These inferences are a little uncertain for the sixteen HSS with low confidence of classification.

The stellar systems in Group (ii)-6 have slightly higher mass compared to Group (iii). The effective radius is also higher for the HSS in this group which indicates that these stellar systems are larger compared to the objects in Group (iii). The mass-luminosity ratio is significantly low compared to the other two groups formed in the first merging phase, and indicates that these stellar systems are relatively young as compared to other HSS. A marginally low surface density compared to that of Group (iii) indicates that these HSS are not as bright as those in Group (iii) although they are substantially brighter than those in Group (i)-5. These inferences are highly uncertain since a good number of stellar systems in this group have low confidence of classification.

An important aspect of our analysis is how our algorithm is able to capture the actual number of well-separated groups present in this dataset. Previous analyses like those of Chattopadhyay & Karmakar (2013) indicated that there are four spherically dispersed groups but Figures 5, 6, 7 clearly show that the actual number of well-separated groups is two. The additional groups found by Chattopadhyay & Karmakar (2013) (and even *t*MMBC in our initial stage) is due to the inability of *k*-means and *t*MMBC to detect complex group structures. Our algorithm is able to capture this complex structure through an effective merging mechanism in multiple (here, two) phases. Another important aspect is the confidence of classification for each HSS that allows us to assess how well our clustering algorithm performed for this dataset. The fact that most of the HSS have high confidence of classification indicates that our clustering algorithm has been able to reliably capture the complex structure present in the data.

5. Conclusions

Chattopadhyay & Karmakar (2013) performed statistical cluster analysis on 673 HSS from Misgeld & Hilker (2011) to determine the homogeneous groups present in the HSS data. Using the *k*-means clustering algorithm together with the jump statistic (Sugar & James 2003), they arrived at five optimal groups whose properties were explored using fundamental plane relations and other physical parameters. Their main analysis excluded the candidate GCs (GC_VCCs), because of concerns that including them would render the data unfit for clustering. Their analysis hinges on an assumption of homogeneous spherically dispersed groups, that underlies the *k*-means algorithm, and that may not be appropriate and can lead to erroneous results, especially when there is a complex structure present in the underlying statistical groups as found in our analysis. We initially grouped all 13 456 HSS of Misgeld & Hilker (2011) using *t*MMBC and BIC to optimally find eight groups. Motivated from a syncytial clustering technique developed by Almodóvar-Rivera & Maitra (2020), we objectively, in a data-driven manner, merged components of the *t*MMBC solution which revealed two complex-structured groups. Using a nonparametric bootstrap technique, we further determined the confidence of classification for each of the 13 456 HSS at each stage of merging to quantify how correctly the stellar systems have been assigned to their correct groups. We then studied the physical and evolutionary properties of the objects in each of the two groups by analysing their mean stellar mass, surface density and effective radius. We found that Group I (that is, Group 5 from the original *t*MMBC solution) typically consists of large and old

stellar systems. Group II which is made up by merging Groups 1-4, 6-8 of the original *t*MMBC solution consists of younger stellar systems which are brighter than those of Group I and also smaller in size compared to those in Group I. All the GC_VCCs were classified to this group along with all the GCs and some other HSS. Our results point to the fact that the GCs are indeed different from elliptical galaxies which includes the dwarf ellipticals (dEs). These results support earlier conclusions (Kormendy 1985). Our results also confirm that the GC_VCCs are most likely to be GCs and that there is very less chance that these are dwarfs.

A reviewer has asked whether our merging stages can, for instance, result in erasing of correlations between parameters that may have physical meaning, and whether this presents a limitation for our method. We note that correlation only measures linear association (or relationships) and is inadequate for describing other kinds of relationships. The possible reduction or erasure of the correlation between parameters in complex-structured groups formed by merging, is a consequence more of the complex structure of the well-separated groups: note also that relationships (including linear ones) continue to be described, as necessary, at the appropriate sub-group level. Therefore, we do not necessarily consider our methodology to have a limitation here: rather, we feel that it provides a more nuanced understanding of the group structure present in the HSS data. Finally, we note that the goal of clustering is to find different groups of homogeneous observations. How the clustering is done is dependent on the resolution (which is analogous to the merging level) and the properties that we would like our clusters to portray. For instance, if we want ellipsoidally shaped groups, then the eight-component *t*MMBC solution provides a good description of the HSS. At the same time, we can see which of these groups of HSS are closer (or more similar) to each other, at the different resolution levels, and how they form into more generally shaped groups. Our approach here formally helps determine the groups that combine (i.e., are similar) at each level and provides an objective approach to describe where the groups are distinct enough to require separate consideration and understanding of their properties. As we see in Section 4.3.2, this sort of understanding of the underlying complex group structure can better inform our understanding of the characteristics of the different kinds of HSS.

There are a number of issues that merit further investigation. For example it would be useful to explore if the logarithmic transformation used on the parameters is plausible or is detrimental to the objective of finding groups of HSS. It is possible that some other transformation may lead to better-separated groups. Our analysis may also be extended to larger and more comprehensive data sets than in this study, in order to analyse whether similar results hold for the HSS from other sources. Also with additional information like temperature and colour indices of the stars present in the system, it might be possible to obtain deeper understanding of the evolution of these systems possibly through HR diagrams. Apart from this it would also be interesting to analyse how fundamental plane relations hold in conjunction with these results. Thus we see that while our analysis has provided some interesting insights into the different types of HSS, additional issues remain that deserve further attention.

Acknowledgement. We sincerely thank T. Chattopadhyay for providing us with the dataset of Chattopadhyay & Karmakar (2013). We are also very grateful to C. Struck for helpful comments on clarifying and interpreting our

results. Our sincere appreciation also to an Associate Editor and an anonymous reviewer whose thorough and insightful comments greatly improved an earlier version of this article.

References

- Almodóvar-Rivera, I., & Maitra, R. 2020, *JMLR*, 21, 1
 Bender, R., Burstein, D., & Faber, S. M. 1992, *ApJ*, 399, 462
 Bernardi, M., et al. 2003, *AJ*, 125, 1866
 Brosche, P. 1973, *A&A*, 23, 259
 Burstein, D., Bender, R., Faber, S., & Nolthenius, R. 1997, *AJ*, 114, 1365
 Chattopadhyay, S., & Maitra, R. 2017, *MNRAS*, 469, 3374
 Chattopadhyay, S., & Maitra, R. 2018, *MNRAS*, 481, 3196
 Chattopadhyay, T., & Karmakar, P. 2013, *NewA*, 22, 22
 Dempster, A. P., Laird, N. M., & Rubin, D. B. 1977, *JRSSB*, 39, 1
 Djorgovski, S. 1995, *ApJ*, 438, L29
 Efron, B. 1979, *AnS*, 7, 1
 Everitt, B. 2011, *Cluster Analysis* (Chichester, West Sussex, UK: Wiley), doi: 10.1002/9780470977811
 Forbes, D. A., Lasky, P., Graham, A. W., & Spitler, L. 2008, *MNRAS*, 389, 1924
 Fraley, C., & Raftery, A. E. 1998, *CJ*, 41, 578
 Fraley, C., & Raftery, A. E. 2002, *JASA*, 97, 611
 Goren, E. M., & Maitra, R. 2022, *Stat*, 11, e416
 Harris, W. E., Pritchett, C. J., & McClure, R. D. 1995, *ApJ*, 441, 120
 Ichikawa, S. I., Wakamatsu, K. I., & Okamura, S. 1986, *ApJS*, 60, 475
 Jordán, A., et al. 2008, *ApJSS*, 180, 54
 Kettenring, J. R. 2006, *JC*, 23, 3
 Kormendy, J. 1985, *ApJ*, 295, 73
 Kormendy, J., Fisher, D. B., Cornell, M. E., & Bender, R. 2009, *ApJS*, 182, 216
 Maitra, R. 2009, *IEEE/ACM TCBB*, 6, 144
 Maitra, R. 2010, *NeuroImage*, 50, 124
 Maitra, R., & Melnykov, V. 2010, *JCGS*, 19, 354
 McLachlan, G., & Krishnan, T. 2008, *The EM Algorithm and Extensions* (2nd edn.; New York: Wiley), doi: 10.2307/2534032
 McLachlan, G., & Peel, D. 2000, *Finite Mixture Models* (New York: John Wiley and Sons, Inc.), doi: 10.1002/0471721182
 McLachlan, G. J., & Peel, D. 1998, in *Advances in Pattern Recognition: Joint IAPR International Workshops SSPR'98 and SPR'98 Sydney, Australia, August 11–13, 1998 Proceedings*, ed. A. Amin, D. Dori, P. Pudil, & H. Freeman (Berlin, Heidelberg: Springer Berlin Heidelberg), 658
 McLaughlin, D. E., & van der Marel, R. P. 2005, *ApJS*, 161, 304
 Melnykov, V. 2016, *JCGS*, 25, 66
 Melnykov, V., Chen, W.-C., & Maitra, R. 2012, *JSS*, 51, 1
 Melnykov, V., & Maitra, R. 2010, *SS*, 4, 80
 Melnykov, V., & Maitra, R. 2011, *JMLR*, 12, 69
 Meylan, G., et al. 2001, *AJ*, 122, 830
 Misgeld, I., & Hilker, M. 2011, *MNRAS*, 414, 3699
 R Core Team. 2017, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria
 Schwarz, G. 1978, *AnS*, 6, 461
 Sugar, C. A., & James, G. M. 2003, *JASA*, 98, 750
 Webbink, R. F. 1985, in *Dynamics of Star Clusters*, ed. J. Goodman, & P. Hut, Vol. 113, 541
 Xu, R., & Wunsch, D. C. 2009, *Clustering* (NJ, Hoboken: John Wiley and Sons, Inc), doi: 10.1002/9780470382776

Appendix 1. Illustration of the pairwise overlap

In this section, we demonstrate the kinds of 3D distributions that may be obtained with specification of different pairwise overlap measures. Using the R (R Core Team 2017) package MIXSIM (Melnykov et al. 2012), we simulated 1000 observations from two-component GMMs with pairwise overlap $\omega = 10^{-5}, 0.001, 0.01, 0.05, 0.1$, and 0.5 . The simulated data are displayed in Figure A.1. Each component distribution is also specified by means of its 95% ellipsoid of concentration. The figure shows very good separation with $\omega = 10^{-5}$, and even with $\omega = 0.001$, modest separation for $\omega = 0.01$, substantially poor separation for $\omega = 0.05$, and worsening separation with higher values of ω .

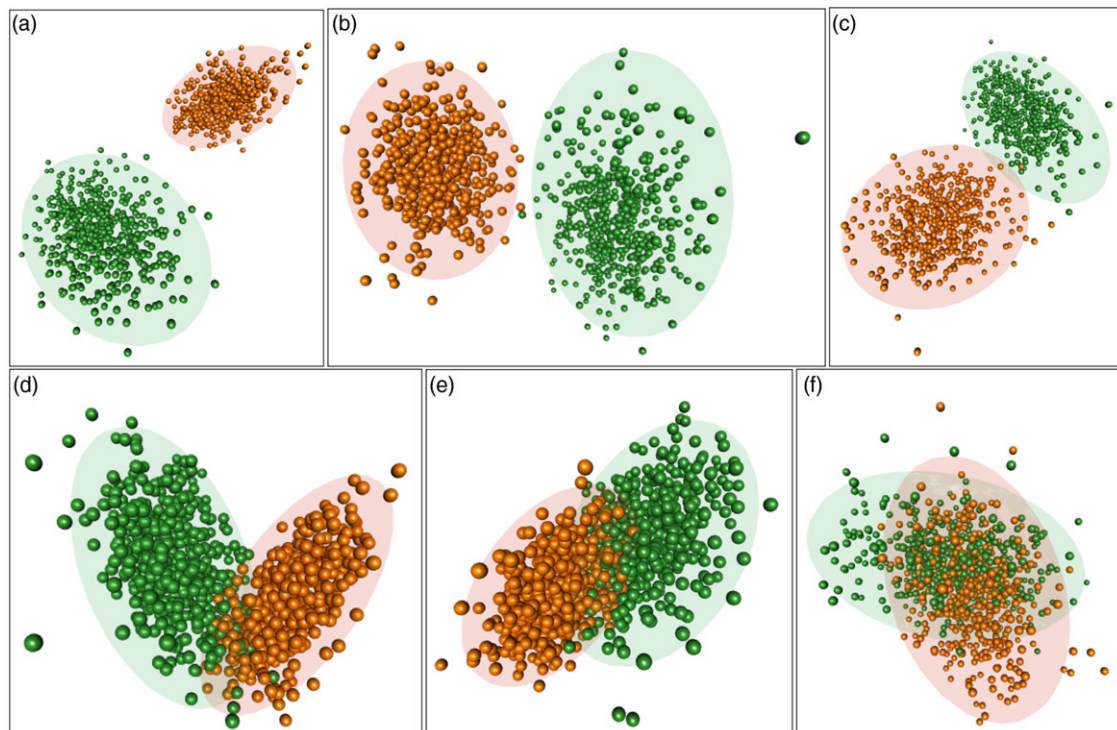


Figure A.1. Realisations from sample two-component Gaussian mixture distributions having pairwise overlap measures of (a) $\omega = 0.00001$, (b) $\omega = 0.001$, (c) $\omega = 0.01$ (d) $\omega = 0.05$, (e) $\omega = 0.1$ and (f) $\omega = 0.5$. For each component, we also provide the 95% ellipsoid of concentrations.