The Astrophysical Journal, 724:678–686, 2010 November 20 doi:10.1088/0004-637X/724/1/678

STATISTICAL ANALYSIS OF DWARF GALAXIES AND THEIR GLOBULAR CLUSTERS IN THE LOCAL VOLUME

Tanuka Chattopadhyay[1], Margarita Sharina[2,3], and Pradip Karmakar[1]
[1] Department of Applied Mathematics, Calcutta University, 92 A.P.C. Road, Calcutta 700009, India; tanuka@iucaa.ernet.in
[2] Special Astrophysical Observatory, Russian Academy of Sciences, N. Arkhyz, KCh R, 369167, Russia
[3] Laboratoire d'Astrophysique de Toulouse-Tarbes, Université de Toulouse, CNRS, 14 Avenue E. Belin, F-31400, France
Received 2010 June 26; accepted 2010 September 13; published 2010 November 4

ABSTRACT

Although morphological classification of dwarf galaxies into early and late types can account for some of their origin and characteristics, this does not aid the study of their formation mechanism. Thus an objective classification using principal component analysis together with *K* means cluster analysis of these dwarf galaxies and their globular clusters (GCs) is carried out to overcome this problem. It is found that the classification of dwarf galaxies in the local volume is irrespective of their morphological indices. The more massive ($M_{V0} < -13.7$) galaxies evolve through self-enrichment and harbor dynamically less evolved younger GCs, whereas fainter galaxies ($M_{V0} > -13.7$) are influenced by their environment in the star formation process.

Key words: galaxies: dwarf – methods: data analysis – methods: statistical

Online-only material: color figures

1. INTRODUCTION

Galaxies with low luminosities and low metallicities having smaller sizes are known as dwarf galaxies. The study of dwarf galaxies is important as massive galaxies are presumed to be formed by the hierarchical merging of dwarf galaxies during the evolution of the early universe (White & Rees 1978; Geisler et al. 2007; Haines et al. 2006). Such objects lose gas easily due to their shallow potential wells. Low surface brightness (LSB) dwarf galaxies are classified primarily into three groups: early-type (dwarf spheroidal, dSph, and dwarf elliptical, dE), late-type (dwarf irregular, dIrr), and transition-type galaxies (Kormendy 1985; Karachentseva et al. 1985; Grebel 1999). There is no sharp border line between these morphological types (see, e.g., Sharina et al. 2008 and references therein, hereafter S08). Also, population gradients exist in early-type dwarf galaxies, and dSphs and dIrrs have exponential surface brightness profiles. All these facts indicate that classification based on morphology and stellar content is not sufficient for studying the formation and evolutionary status of these objects. A more sophisticated classification is essential for studying true formation mechanisms of this class of objects. However, the objects under consideration share one common property. They all harbor globular clusters (GCs) which are older than several Gyr which indicates that early GC formation took place irrespective of morphological type. Thus these GCs can serve as a unique tool for investigating the chemical evolution of the host galaxies. Hence a proper classification of GCs in LSB galaxies is necessary for finding information regarding star formation histories in these dwarf galaxies which is important input for studying the galaxy formation mechanism. Since Hubble's (1922, 1926) tuning fork diagram, few attempts have been made to develop an objective classification for normal and dwarf galaxies using statistical methods and principles (Whitmore 1984; Vaduvescu & McCall 2008; Fraix-Burnet et al. 2006; Chattopadhyay & Chattopadhyay 2006; van den Bergh 2007, 2008; Woo et al. 2008) though some work has been carried out on dE classification and formation scenarios (Marin-Franch & Aparicio 2002; Lisker et al. 2007; Penny & Conselice 2008) in Virgo, Perseus, and Coma clusters of galaxies which is again a consideration of a particular morphological type and therefore not exhaustive. In order to identify the parameters that are mostly responsible for the variation among dwarf galaxies and their GCs and to classify them into homogeneous groups for searching the possible formation mechanism we have to use statistical techniques such as principal component analysis (PCA) and cluster analysis (CA).

In the present paper, in order to study the underlying features of the dwarf galaxy population we have used statistical methods such as PCA, CA, and discriminant analysis. By treating the samples under consideration as representatives of the corresponding underlying population of dwarf galaxies, these methods help us to make an inference regarding the above-mentioned population (and not only for the samples under consideration). As a result, on the basis of the present study we can make some general conclusions which are not feasible on the basis of visual studies.

In Section 2, the data sets are discussed. The methods used are described in Section 3 while results, discussions, and conclusions are summarized in Sections 4, 5, and 6, respectively.

2. DATA SET

Our analysis is based on two data sets of dwarf galaxies and their GCs in the Local Volume (LV).

Data set 1. This data set consists of 60 dwarf galaxies taken from a data set of 104 dwarf galaxies (Sharina et al. 2008; Table 1). The parameters considered, from Sharina et al. (2008, hereafter S08), are distance modulus (μ_0, in mag), morphological index (*T*), mean metallicity of the red giant branches ([Fe/H], in dex), effective color corrected for extinction ($(V-I)_{e0}$, in mag), logarithm of the projected major axis from CNG (log(Diam), in kiloparsec), logarithm of the limiting diameter (log(Dlim), in kiloparsec), limiting *V* and *I* absolute magnitudes within the diameter Dlim corrected for extinction (M_{V0}, M_{I0}), extinction-corrected mean surface brightness within a 25 mag isophote in *V* and *I* magnitudes ($SBV_{25,0}$, $SBI_{25,0}$ in mag arcsec^{-2}), effective surface brightness

in the V band corrected for extinction (SBV_{e0}, in mag arcsec^{-2}), logarithm of effective radius (log(R_e) in kpc), logarithm of model exponential scale length (log h, in kpc), best exponential fitting central surface brightness in the V and I bands corrected for extinction (SBVC0, SBIC0, respectively, in mag arcsec^{-2}). The parameters used from Karachentsev et al. (2004, hereafter CNG) are H I rotational velocity (V_m in km s^{-1}), H I mass-to-luminosity ratio ($M_{\rm H\,I}/L$ in solar units), and tidal index (Θ). The scaling parameters used from Georgiev et al. (2010) are GC specific frequency (S_N), specific luminosity (S_L), specific mass (S_M), specific number ($\hat{T}$), logarithm of specific GC formation efficiency as a function of galaxy luminosity and mass (η_L, η_M), total stellar mass ($M_{*,V}$ in 10^7 $M_\odot$), and H I mass of the host galaxy ($M_{\rm H\,I}$ in 10^7 $M_\odot$), respectively.

During the selection of parameters for PCA and CA the following things were taken into consideration.

1. The parameters must be intrinsic in nature.
2. For parameters that are almost physically similar any one is chosen at random because inclusion of similar parameters is considered redundant in CA.
3. All the parameters should be without missing values as CA cannot be used with parameters that have missing values exceeding 5% for which mean substitution might occur (Little & Rubin 2002).

With respect to point 1 above we excluded μ_0 and T which are not intrinsic properties of dwarfs. We included $SBV_{25,0}$, $SBI_{25,0}$, M_{V0}, M_{I0}, log R_e, and log h. $SBVL$, $SBIL$, MV_{25}, MI_{25}, $R_{V,25}$, $R_{I,25}$ were not included because of point (2). We have not considered the remaining parameters except for Θ with respect to point (3), as they have missing values exceeding 5%. However, once the dwarf galaxies were classified we used these parameters to study the properties of the dwarf galaxies in more detail. Thus, of all the parameters mentioned above only 13 from Sharina et al. (2008), excluding μ_0 and T, together with Θ from Karachentsev et al. (2004), were directly used for PCA and CA as the sample is without any missing values with respect to these 14 parameters. Thus, we followed a very standard procedure for PCA and CA of a sample of astronomical objects. In order to have a sample where the values of all the parameters corresponding to each dwarf galaxy are available, we had to drop observations corresponding to 44 dwarf galaxies and, as a result, we had a sample size of 60. The sample is not complete as there are many more galaxies in the LV which have yet to be observed. In this sense, no catalog is complete. The question remains of whether or not the sample is representative of the original one. Regarding this point, all of the two-point correlations discussed in the previous paper (S08) are still in place after selection of the present sample. It is important to note that it contains all transitional type galaxies (dSph/dIrr, $T = -1$) from the original sample, so all morphological types are well represented in this sense. A list of dwarf galaxies considered in data set 1 is given in Table 1.

Data set 2. This set consists of 100 GCs in the LV dwarf galaxies (Sharina et al. 2005). Three candidates, Sc 22-2-879, Sc 22-100, and Sc 22-4-106, were removed since they had subsequently been identified as galaxies and not GCs (Da Costa et al. 2009). Also, the parameters of the GCs in UGC4115, KK65, and UGC3755 were recalculated using the current distances 7.727, 8.017, and 7.413 Mpc, respectively (Tully et al. 2006). The parameter set consists of the logarithm of the half-light radius (log(r_h) in parsecs), apparent axial ratio (e), integrated absolute magnitude ($V0$, in mag) corrected for extinction, integrated absolute $(V-I)_0$ color (corrected for Galactic extinction, in mag), projected distance from the host

Table 1
List of Dwarfs in Data Set 1

Name	R.A. (2000)	Decl. (2000)
E349-031	00 08 13.3	−34 34 42.0
E410-005	00 15 31.4	−32 10 48.0
E294-01	00 26 33.3	−41 51 20.0
KDG2	00 49 21.1	−18 04 28.0
E540-032	00 50 24.3	−19 54 24.0
UGC685	01 07 22.3	16 41 02.0
KKH5	01 07 32.5	51 26 25.0
KKH6	01 34 51.6	52 05 30.0
KK16	01 55 20.6	27 57 15.0
KK17	02 00 09.9	28 49 57.0
KKH34	05 59 41.2	73 25 39.0
E121-20	06 15 54.5	−57 43 35.0
E489-56	06 26 17.0	−26 15 56.0
KKH37	06 47 45.8	80 07 26.0
UGC3755	07 13 51.8	10 31 19.0
E059-01	07 31 19.3	−68 11 10.0
KK65	07 42 31.2	16 33 40.0
UGC4115	07 57 01.8	14 23 27.0
DDO52	08 28 28.5	41 51 24.0
D564-08	09 02 54.0	20 04 31.0
D565-06	09 19 29.4	21 36 12.0
KDG61	09 57 02.7	68 35 30.0
KKH57	10 00 16.0	63 11 06.0
HS117	10 21 25.2	71 06 58.0
UGC6541	11 33 29.1	49 14 17.0
NGC3741	11 36 06.4	45 17 07.0
E320-14	11 37 53.4	−39 13 14.0
KK109	11 47 11.2	43 40 19.0
E379-07	11 54 43.0	−33 33 29.0
NGC4163	12 12 08.9	36 10 10.0
UGC7242	12 14 07.4	66 05 32.0
DDO113	12 14 57.9	36 13 08.0
DDO125	12 27 41.8	43 29 38.0
UGC7605	12 28 39.0	35 43 05.0
E381-018	12 44 42.7	−35 58 00.0
E443-09	12 54 53.6	−28 20 27.0
KK182	13 05 02.9	−40 04 58.0
UGC8215	13 08 03.6	46 49 41.0
E269-58	13 10 32.9	−46 59 27.0
KK189	13 12 45.0	−41 49 55.0
E269-66	13 13 09.2	−44 53 24.0
KK196	13 21 47.1	−45 03 48.0
KK197	13 22 01.8	−42 32 08.0
KKs55	13 22 12.4	−42 43 51.0
14247	13 26 44.4	−30 21 45.0
UGC8508	13 30 44.4	54 54 36.0
E444-78	13 36 30.8	−29 14 11.0
UGC8638	13 39 19.4	24 46 33.0
KKs57	13 41 38.1	−42 34 55.0
KK211	13 42 05.6	−45 12 18.0
KK213	13 43 35.8	−43 46 09.0
KK217	13 46 17.2	−45 41 05.0
CenN	13 48 09.2	−47 33 54.0
KKH86	13 54 33.6	04 14 35.0
UGC8833	13 54 48.7	35 50 15.0
E384-016	13 57 01.6	−35 20 02.0
KK230	14 07 10.7	35 03 37.0
DDO190	14 24 43.5	44 31 33.0
E223-09	15 01 08.5	−48 17 33.0
IC4662	17 47 06.3	−64 38 25.0

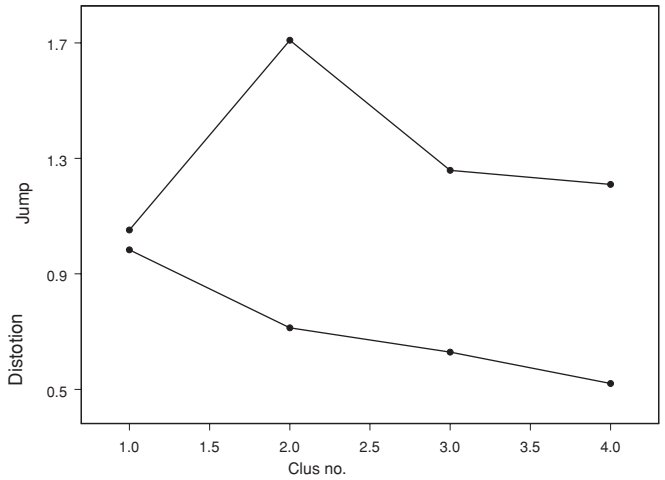

Figure 1. Distortion and jump curves for the classification of dwarf galaxies in the LV.

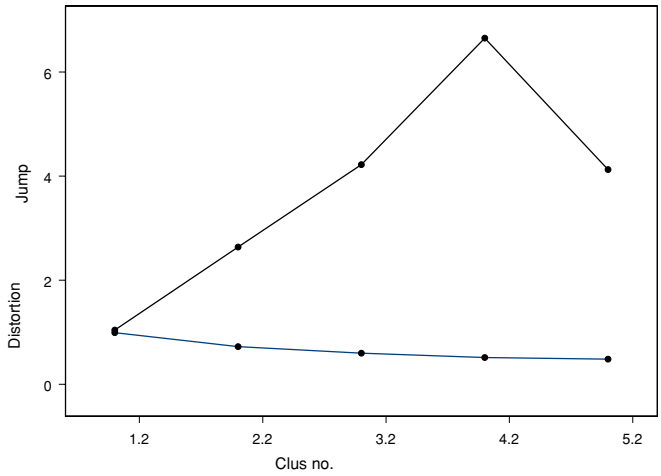

Figure 2. Distortion and jump curves for the classification of GCs in the LV.

galaxy ($d_{\rm proj}$, in kpc), central surface brightness in the V and I bands (μ_{V0}, μ_{I0} in mag arcsec^{-2}), and the logarithms of King core radius and tidal radius (log(r_c), log(r_t), respectively, in parsec).

3. METHOD

PCA is a very common technique used in data reduction and interpretation in multivariate analysis. We are interested in discovering which parameters in a data set form coherent subgroups that are relatively independent of each other. The specific aim of the analysis is to reduce a large number of parameters to a smaller number while retaining maximum spread among experimental units. The analysis therefore helps us to determine the optimum set of parameters causing overall variation in the nature of the objects under consideration. PCA has been discussed and used by various authors (Babu et al. 2009; Chattopadhyay & Chattopadhyay 2006, 2007; Whitmore 1984; Murtagh & Heck 1987).

CA is the art of finding groups in data. Over the last 40 years different algorithms and computer programs have been developed for CA. The choice of a clustering algorithm depends both on the type of data available and on the particular purpose.

In the present study, we have used the K means partitioning algorithm (MacQueen 1967) for clustering. This method constructs K clusters, i.e., it classifies the data into K groups which together satisfy the requirement of a partition such that each group must contain at least one object and each object must belong to exactly one group so at most, there are as many groups as there are objects ($K <= n$). Two different clusters cannot have any object in common and the K groups together add up to the full data set. Partitioning methods are applied if one wants to classify the objects into K clusters where K is fixed (which should be optimally selected). The aim is usually to uncover a structure that is already present in the data. The K means is probably the most widely applied partition clustering technique.

To perform K means clustering we have used the MINITAB package. Under this package, cluster centers have been chosen on the basis of the group average method which makes the process almost robust. This method was developed by Milligan (1980).

By using this algorithm we first determined the structures of sub-populations (clusters) for varying numbers of clusters taking $K = 2, 3, 4$, etc. For each such cluster formation, we computed the values of a distance measure $d_K = (1/p)\min_x E[(x_K - c_K)'(x_K - c_K)]$ which is defined as the distance of the x_K vector (values of the parameters) from the center c_K (which is estimated as the mean value), where p is the order of the x_K vector. Then the algorithm for determining the optimum number of clusters is as follows (Sugar & James 2003). Let us denote by d'_K the estimate of d_K at the Kth point. Then d'_K is the minimum achievable distortion associated with fitting K centers to the data. A natural way of choosing the number of clusters is to plot d'_K versus K and look for the resulting distortion curve (see bottom panels of Figures 1 and 2). This curve is always monotonically decreasing. Initially, one would expect much smaller drops for K greater than the true number of clusters because past this point adding more centers simply partitions within groups rather than between groups. According to Sugar & James (2003), for a large number of items, the distortion curve will exhibit a sharp "jump" when transformed to an appropriate negative power ($p/2$) (if we plot K versus transformed d'_K). We calculated the jumps in the transformed distortion as $J_K = (d'^{-\mathbf{p/2}}_K - d'^{-\mathbf{p/2}}_{K-1})$.

The optimum number of clusters is the value of K associated with the largest jump. The largest jump can be determined by plotting J_K against K and the highest peak will correspond to the largest jump (see top panels of Figures 1 and 2).

It is well known that both the PCA and CA methods are parameter dependent and the parameters considered should be responsible for the variation of the objects under consideration. In the present situation, all of the parameters of that type are taken into consideration. As we have to depend on the available data only, it was not possible for us to consider many unobserved parameters whose inclusion might have improved the classification, e.g., inclusion of central velocity dispersion. But the question is whether the classification is robust or not given the parameters and sample. In this respect, a discriminant analysis is performed (Johnson & Wichern 1998) to verify the acceptability of the classification by computing misclassification probabilities for the different dwarfs and GCs. If the original classification is robust then every dwarf or GC should again be classified as a member of the same class to which it was classified before. Tables 2 and 3 show the results of a discriminant analysis. The fractions of correct classifications are 0.983 and 0.97, respectively, which imply that the classifications are almost robust. Since in the present situation we have only one sample, it is difficult to say whether the same results will be

Table 2
Discriminant Analysis for Dwarf Galaxies in LV

	Number of members	
DA Clusters	G1	G2
$G1^*$	35	1
$G2^*$	0	24
Total	35	25

Notes. G1, G2 are the groups found by K means and $G1^*$ and $G2^*$ are the groups to which dwarfs are assigned by the discriminant analysis. $N = 60$, $N_{\rm correct} = 59$, proportion correct = 0.983.

Table 3
Discriminant Analysis for GCs in the LV Dwarf Galaxies

	Number of members			
DA Clusters	GC1	GC2	GC3	GC4
$GC1^*$	23	0	0	0
$GC2^*$	0	17	0	1
$GC3^*$	1	0	37	1
$GC4^*$	0	0	0	20
Total	24	17	37	22

Notes. GC1, GC2, GC3, and GC4 are the groups found by K means and $GC1^*$, $GC2^*$, $GC3^*$, and $GC4^*$ are the groups to which GCs are assigned by the discriminant analysis. $N = 100$, $N_{\rm correct} = 97$, proportion correct = 0.97.

obtained for other samples as well. It can only be inferred that if the present sample is representative of the underlying population of dwarf galaxies, then the results obtained in this paper are generally true.

4. DWARF GALAXIES OF THE LOCAL VOLUME

For PCA, first we computed a correlation matrix with all 14 parameters for data set 1 and took any one of the two physically similar (e.g., absolute magnitudes in V and I bands) highly correlated (correlation > 0.7) parameters. Following this method, eight parameters were selected for PCA. They were Θ, [Fe/H], M_{V0}, $SBV_{25,0}$, $(V-I)_{e0}$, SBV_{e0}, $\log(R_e)$, and $\log h$. For these eight parameters, PCA gave four principal components with eigenvalues greater than or equal to 1 and at the same time almost 87.7% overall variation. We considered the above four principal components with eigenvalues greater than or equal to 1 and computed the correlations of the parameters appearing in each principal component with the corresponding principal component. We considered significant parameters to be those for which the correlation was greater than 0.65. Thus, following this procedure, the significant parameters are M_{V0}, SBV_{e0}, and $\log(R_e)$ (from the first principal component), Θ, $(V-I)_{e0}$ (from the second principal component) and [Fe/H] (from the third principal component). The fourth component contributes no parameters with such a high correlation.

For CA we took the above six significant parameters and used the method assuming $K = 1, 2, 3$, etc. The optimum number of coherent groups by the above method was obtained at $K = 2$ (viz., G1 and G2). The "distortion" and "jump" curves are shown in Figure 1. The mean values with standard errors for some parameters and significant correlations with their p values are shown in Table 4 for groups G1 and G2, respectively.

In a multivariate situation the role of all parameters is important for classification as they are correlated with each other

Table 4
Mean Values and Correlations of the Significant Parameters for the Two Groups of Dwarfs in the LV

Parameters	G1	G2
Number	35	25
Θ	-0.631 ± 0.200	0.880 ± 0.287
[Fe/H]	-1.8394 ± 0.0658	-1.6675 ± 0.0586
M_{V0}	-14.060 ± 0.190	-11.742 ± 0.177
SBV_{e0}	22.730 ± 0.150	24.102 ± 0.128
$(V-I)_{e0}$	0.7298 ± 0.0257	0.9750 ± 0.0391
$\log R_e$	-0.2827 ± 0.0331	-0.3842 ± 0.0357
V_m	26.12 ± 3.23	11.11 ± 3.69
$M_{\rm H\,I}$	6.571 ± 2.40	0.516 ± 0.249
M_{*V}	6.69 ± 3.06	0.41 ± 0.155
$\hat{T}_b$	52.6 ± 15.7	238.1 ± 96.4
$\log(\eta_L)$	-4.264 ± 0.224	-4.568 ± 0.329
$\log(\eta_M)$	-4.616 ± 0.301	-4.912 ± 0.47
Correlations	r p	r p
$(\log \eta_L, \Theta)$	0.096 0.805	0.987 0.002
$(\log \eta_M, \Theta)$	0.345 0.364	0.956 0.011
(S_L, Θ)	0.102 0.793	0.889 0.044
(S_M, Θ)	0.200 0.606	0.896 0.040
$(\hat{T}_b, \Theta)$	0.460 0.213	0.854 0.065
([Fe/H], M_V)	−0.553 0.001	−0.290 0.160
$(M_{V0}, (V-I)_{e0})$	−0.224 0.195	−0.155 0.459

but sometimes one or two parameters may play a significant role over the others when there are large variations among the values of those parameters. In the present situation, the magnitude (M_{V0}), tidal index (Θ), and effective surface brightness (SBV_{e0}) play such a role (Table 3). Although in terms of magnitude it is possible to find a single cut at $M_{V0} = -13.7$ irrespective of the other two (Figure 3), if we consider Figure 5 it is clear that no such single cut is available in terms of Θ. As such, the classification is based mainly on the three major parameters M_{V0}, Θ, and SBV_{e0} and not solely on the magnitude. In such a multivariate setup we discuss the marginal situations (i.e., the effect of a single parameter) in order to display the results graphically so that one can visualize the underlying scenario. For example, Figure 3 is a two-dimensional projection of the original six-dimensional situation.

Further, on the basis of PCA and CA, we divided the objects into certain groups with respect to certain parameters. The parameter ranges for different groups are dependent on one another and determined by various factors such as the range of the parameters and the size of the sample. Hence the nation that on the basis of a magnitude less than or greater than −13.7 we can get two different groups is not necessarily always true. But the idea that is likely to be retained for different samples is that in most of the situations there will be two significant classes whose distributional natures are likely to be the same as that of the present situation. In a case where the ranges of the sample parameters are close to the present situation one may expect a similar cut in the value of the magnitude.

4.1. Globular Clusters in the Dwarf Galaxies of the Local Volume

For PCA, we used the parameters $\log r_c$, $\log r_h$, e, $V0$, $(V-I)_0$, $d_{\rm proj}$, μ_{I0}, μ_{V0}, and $\log r_t$. The computed correlation matrix did not show a high correlation for physically similar parameters. So all nine of these parameters were considered for PCA. The number of principal components with eigenvalues close to 1 was four for a total variation of 83.3%. For these four

Table 5
Mean Values of the Significant Parameters for the Three Groups of GCs in the LV

Parameters	GC1	GC2	GC3	GC4
Number	24	17	37	22
$V0$	-5.8765 ± 0.0954	-6.041 ± 0.170	-7.481 ± 0.123	-8.263 ± 0.194
μ_{V0}	21.075 ± 0.097	19.903 ± 0.140	20.443 ± 0.141	18.127 ± 159
$\log(r_h)$	0.8863 ± 0.0235	0.7024 ± 0.0287	1.0753 ± 0.0153	0.7820 ± 0.0321
$\log(r_t)$	1.5556 ± 0.0578	1.4077 ± 0.0672	1.7965 ± 0.0557	1.6179 ± 0.0463
$\log(r_c)$	0.5441 ± 0.0366	0.3342 ± 0.0403	0.7244 ± 0.0197	0.3696 ± 0.0377
$\log(r_t/r_c)$	1.0115 ± 0.0747	1.0735 ± 0.0873	1.0721 ± 0.0625	1.2483 ± 0.0500
e	0.1542 ± 0.0208	0.1647 ± 0.0284	0.1081 ± 0.0166	0.10 ± 0.01
$(V-I)_0$	1.1996 ± 0.0340	0.5800 ± 0.0617	0.8035 ± 0.0430	0.8332 ± 0.0436
$d_{\rm proj}$	1.014 ± 0.105	0.563 ± 0.112	1.354 ± 0.171	0.740 ± 0.144
Age (Gyr)	*	*	5.0 ± 1.32	7.2 ± 1.5
Z/H	*	*	-1.167 ± 0.230	-1.3 ± 0.307
(α/H)	*	*	0.2 ± 0.0632	0.18 ± 0.0490

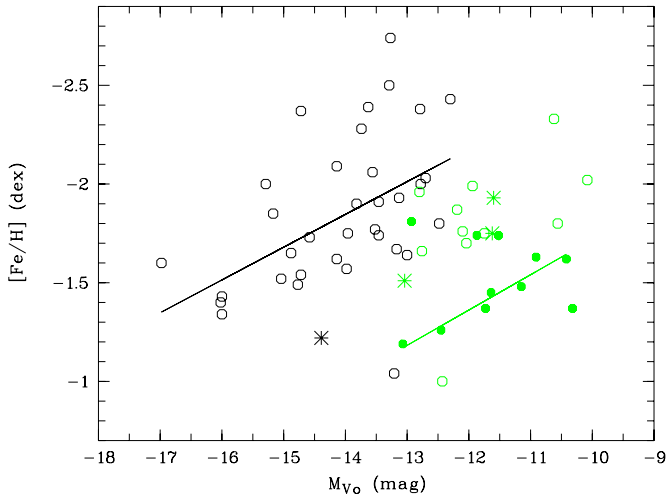

Figure 3. Absolute luminosity in the V-band (M_{V0}) vs. metallicity ([Fe/H]) diagram for the two groups, G1 and G2, of dwarf galaxies found as a result of CA in the LV. The black circles (dIrrs) and one black diamond (dIrr/dSph) are for group G1 and green symbols are for G2. dSphs, dIrrs, and dIrrs/dSphs are shown as dots, circles, and asterisks, respectively. The best-fit line is for G1 galaxies removing the top two black circles as outliers. The green line is from S08.

(A color version of this figure is available in the online journal.)

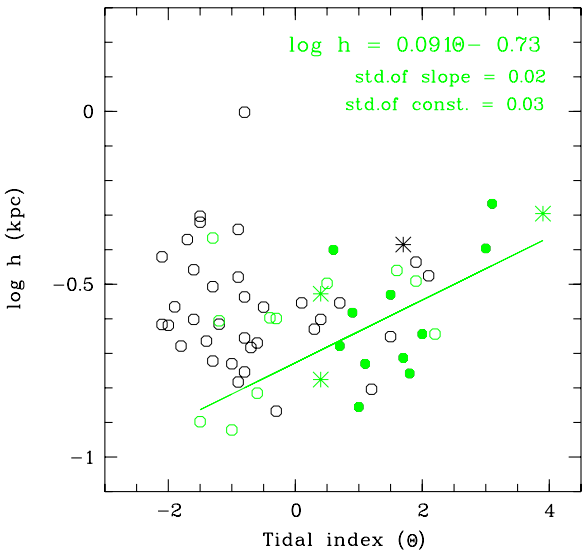

Figure 4. Tidal index (Θ) vs. logarithm of scale length in kiloparsecs for the two groups, G1 and G2, of dwarf galaxies found as a result of CA in the LV. Symbols are the same as in Figure 3. Regression line for G2 was counted removing E443-09 and KKH5 as outliers.

(A color version of this figure is available in the online journal.)

principal components very high correlations occurred only for two parameters so as a rule we considered correlations having values greater than 0.6. Following this, the significant parameters were $\log r_c$, $\log r_h$, e, $V0$, $(V-I)_0$, μ_{I0}, μ_{V0}, and $\log r_t$. Next, a CA was carried out with these eight parameters (standardized) and the optimum number of classes was found to be at $K = 4$. The "distortion" and "jump" curves are shown in Figure 2. The mean values for some parameters are listed in Table 5.

5. DISCUSSION

5.1. Dwarf Galaxies

Two groups, G1 and G2, of dwarf galaxies in the LV have been found as a result of CA, irrespective of their morphological classification (viz., T). In G1, 3% are dIrr/dSphs and 97% are dIrrs whereas in G2 52% are dIrrs, 43% are dSphs, and 5% are dIrr/dSphs (one galaxy). The groups have many distinct properties as seen from Table 4. G1 contains brighter galaxies of larger size with a larger amount of H I mass having a high degree of rotation, whereas G2 consists of fainter galaxies of smaller size and are almost devoid of H I mass having an insignificant amount of rotation. A luminosity–metallicity (viz., [Fe/H] versus M_{V0}) diagram (Figure 3) shows a significant correlation (viz., Table 2, $r \sim -0.553$, $p = 0.001$; two galaxies on top were removed as outliers) together with the best-fit line for the galaxies in G1. The slope of this relation is identical to the one found for dSphs and dIrrs in the Local Group and beyond (Dekel & Silk 1986; Skillman et al. 1989; Smith 1985; S08). The zero point is shifted by ∼4 mag. Note that if we consider G2 in total, such a correlation is absent (viz., Table 2, $r \sim -0.290$, $p = 0.160$). This may indicate that formation of dwarf galaxies is governed by self-enrichment whereas some processes lead to fading during the formation and evolution of stars and interaction of the interstellar gas of dwarf galaxies with the intergalactic medium in groups (see, e.g., Grebel et al. 2003 and references therein). Gravitational potentials are not strong, and gas may be blown out by just a few supernovae. Galactic winds lead to a significant loss of metals from dwarf galaxies. Starvation (Shaya & Tully 1984), tidal, or ram pressure gas stripping affect galaxies in dense galaxy

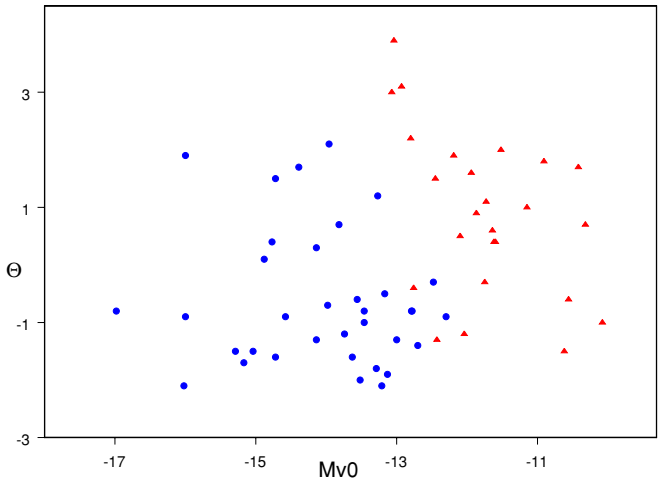

Figure 5. Tidal index (Θ) vs. Mass (M_{V0}) diagram for the two groups, G1 and G2, of dwarf galaxies found as a result of CA in the LV. The blue solid circles are for group G1 and red solid triangles are for G2.

(A color version of this figure is available in the online journal.)

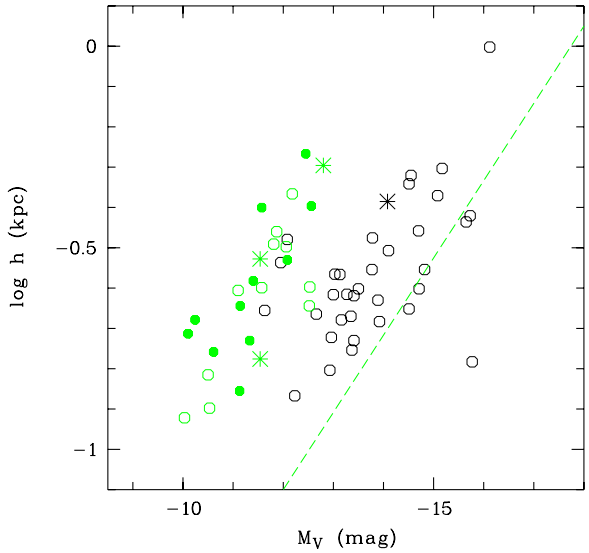

Figure 6. Absolute magnitude in the *V* band, corrected for Galactic extinction, vs. projected distance of a GC from a center of a galaxy for the two groups, G1 and G2, of dwarf galaxies found as a result of CA in the LV. Symbols are the same as in Figure 3. The dotted line is a line of equal central surface brightness for spiral galaxies from S08.

(A color version of this figure is available in the online journal.)

groups or cluster environments. The complex behavior of the luminosity–metallicity in G1 and G2 also might be accounted for by multiple bursts of star formation of short duration in dwarf galaxies of small sizes (Carraro et al. 2001; Hirashita 2000). The presence of H I rotation in G1, and the almost complete absence of gas in G2 supports the above picture.

Figure 4 shows the tidal index versus logarithm of the scale length for the sample galaxies. The so-called tidal index was introduced by Karachentsev & Makarov (1998). It is the maximum logarithm of the local mass densities produced by neighbors of a galaxy. It is seen that for galaxies with a tidal index larger than zero, scale lengths increase with increasing tidal index. This means that neighbors influence the thickening of galactic disks irrespective of morphological types. G2 is more affected by tidal interaction than G1.

Figure 6 shows the absolute magnitude versus logarithm of the scale length for the sample galaxies. The dashed line indicates an $h \sim L^{0.5}$ relation for spiral galaxies. It is seen that the slope of this relation does not change at $M_v \sim -12$ mag as was suggested by S08. We see two sequences of galaxies well divided on the two groups found in our paper. The shift between the two sequences is about 2 mag along the *X*-direction, which is twice as less in comparison with the luminosity–metallicity relation.

One may suggest that the shift in magnitudes between G1 and G2 at the same metallicity (Figure 3) and at the same scale length (Figure 6) is driven by the interplay of different factors. The thickening of disks is produced by interaction with neighbors (tidal, ram pressure stripping) and by disruption of star clusters (Kroupa 2002). The luminosity–metallicity relation is the result of the aforementioned reasons plus the effects of stellar evolution. Since we see the parallel shift according to the absolute magnitude in Figures 3 and 6, one may conclude that G2, which contains all dSphs, evolved from G1 due to reasons such as fading due to cessation of star formation gas outflows produced by supernovae, ram pressure, and tidal stripping.

5.2. Globular Cluster Candidates

Georgiev et al. (2010) have given a conjecture of the formation history of GCs in dwarf galaxies on the basis of stellar and galaxy mass. They investigated the formation of GCs in terms of some observed scaling parameters which were theoretically predicted as a function of galaxy mass on the basis of a model by Dekel & Birnboim (2006). These scaling parameters are specific frequency ($S_N = N_{\rm GC} \times 10^{0.4(M_V+15)}$, where $N_{\rm GC}$ is the number of GCs and M_V is the absolute visual magnitude of the host galaxy), specific mass ($S_M = 100 \times M_{\rm GCs}/(M_* + M_{\rm H\,I})$, where $M_{\rm GCs}$ is the total mass of GCs, M_* is the total stellar mass, and $M_{\rm H\,I}$ is the total H I mass of the host galaxy), specific luminosity ($S_L = 100 \times L_{\rm GCs}/L_V$, where $L_{\rm GCs}$ is the total luminosity of the GCs and L_V is the luminosity of the host galaxy), specific number ($\hat{T} = 10^9\ M_\odot \times N_{\rm GC}/M_b$, where $M_b = M_* + M_{\rm H\,I}$), GC mass and luminosity normalized formation efficiencies (η_M, η_L; related to S_N, S_L, S_M, and $\hat{T}$ through Equations (23)–(26) of Georgiev et al. 2010). According to their model, the star formation process is primarily due to stellar and supernovae feedback when the mass is below $3 \times 10^{10}\ M_\odot$ but is governed by virial shock above this critical mass.

We have computed the correlations of some of these parameters with the tidal index (Θ) for these dwarf galaxies. The correlations show very high values for G2 galaxies ($r \sim 0.9/0.8$, $p < 0.05$, Table 4) contrary to highly insignificant ones ($r \sim 0.09/0.1$, $p \gg 0.05$) for G1 galaxies. This fact indicates that self-enrichment supported by stellar and supernovae feedback plays a very important role in the formation of stellar populations in G1 galaxies but star formation is highly regulated by environment as is evident from high tidal indices, low values of correlations of Θ with scaling parameters, low luminosity–metallicity correlations (Figure 3), and insignificant rotation of H I mass for G2 galaxies. This might be the result of GC formation due to higher velocity collisions in a deep potential well leading to more efficient GC formation. In this respect, Kumai et al. (1993) have suggested that galaxies in a deeper environment (i.e., higher Θ) are more likely to undergo interactions which can increase the random motion of gas clouds within such galaxies. This leads to an increase in S_N (or S_L, S_M, $\hat{T}_b$, $\log(\eta_L)$, $\log(\eta_m)$, etc.) with environment. At

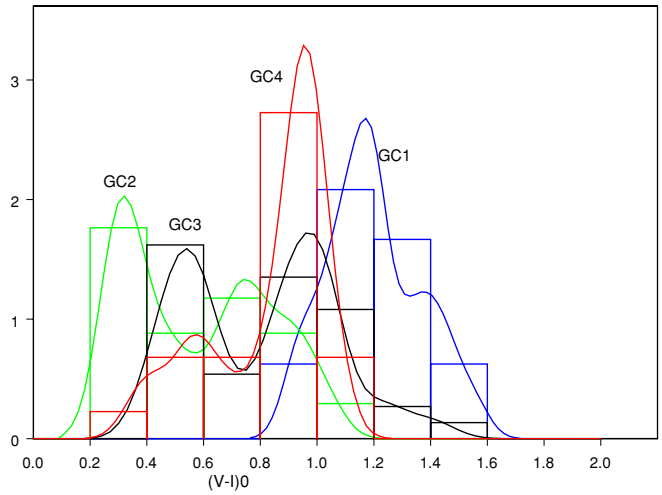

Figure 7. Histograms with density lines of $(V-I)_0$ color for the groups of GCs found as a result of CA in the LV. The blue solid curve is for GC1, the green solid curve is for GC2, the black solid curve is for GC3, and the red solid curve is for GC4.

(A color version of this figure is available in the online journal.)

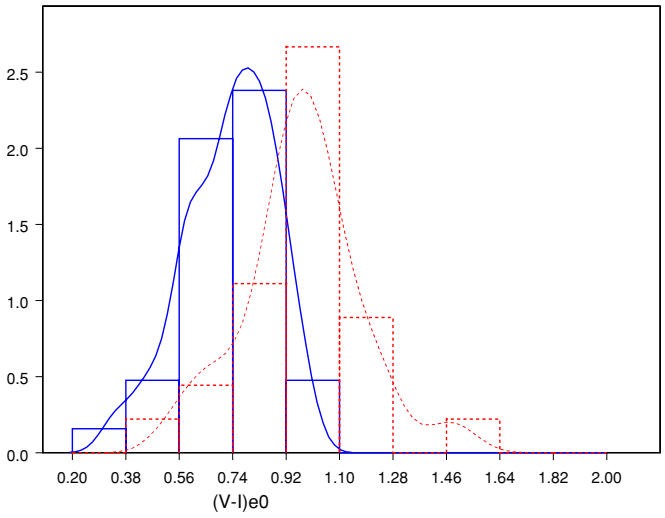

Figure 8. Histograms with density lines of $(V-I)_{e0}$ color for the groups of dwarf galaxies found as a result of CA in the LV. The blue solid line is for G1 galaxies and the red dashed line is for G2.

(A color version of this figure is available in the online journal.)

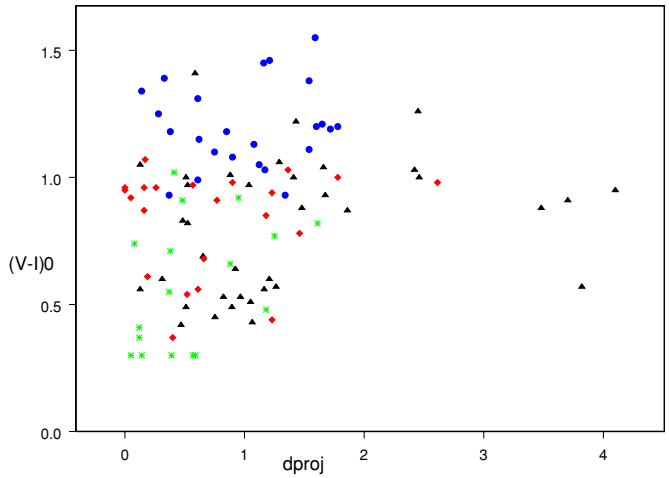

Figure 9. Color $((V-I)_0)$ vs. projected distances ($d_{\rm proj}$) of the four groups of GCs (GC1, GC2, GC3, and GC4). Blue solid circles are for GC1, green stars are for GC2, black triangles are for GC3, and red diamonds are for GC4.

(A color version of this figure is available in the online journal.)

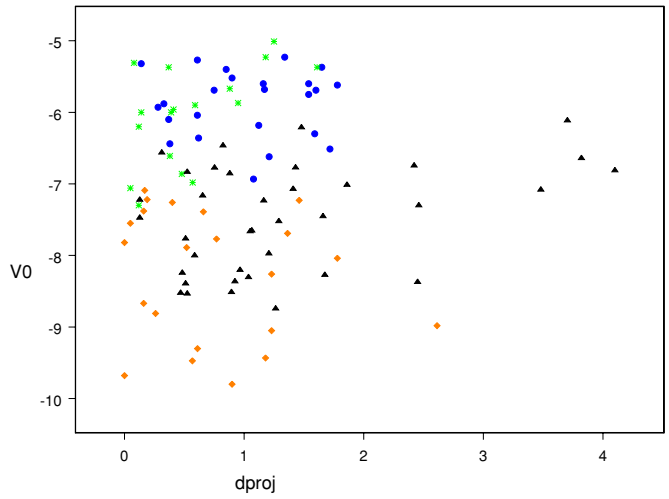

Figure 10. Absolute magnitude $V0$ vs. $d_{\rm proj}$ for the groups of GCs (GC1, GC2, GC3, and GC4). The symbols and colors are same as Figure 9.

(A color version of this figure is available in the online journal.)

the same time, color histograms (Figure 6) of GCs as well as of the dwarf galaxies in G2 show a major star formation episode (largest peak) at $\langle(V-I)_0\rangle \sim 1.0/0.9$ for GCs in GC4 and GC1 and $\langle(V-I)_{e0}\rangle \sim 0.9750$; (viz., Table 4) for G2 galaxies which corresponds to an older burst of star formation (viz., Table 3 $\sim$ Gyr; Sharina et al. 2008; Puzia & Sharina 2008). The above phenomenon can be interpreted as star formation that has been ceased subsequently due to gas stripping or ram pressure sweeping and evaporation which may give rise to a different amount of mass loss as a consequence of the action of the dense environment. Low H I masses as well as low rotation velocities of H I masses for G2 galaxies also support the above scenario. This is in contrast to the low-density environment of G1 galaxies which are free from suffering any external triggering. Hence in a low-density environment a younger burst of star formation is possible (Vilchez 1997). This is also clear from the color profiles of GCs in GC2 and GC3 (Figure 7) and G1 (Figure 8) galaxies which also have peaks at $(V-I) \sim 0.3/0.5$ which correspond to an age less than 1 Gyr (Sharina et al. 2005).

When GCs evolve, their core radii decrease and their tidal radii increase. Thus, the quantity $\log(r_t/r_c)$ increases. When $\log(r_t/r_c) > 2.5$ (Chattopadhyay et al. 2009), the GCs undergo core collapse, i.e., they are dynamically much evolved. Now the values of the above quantity for the four groups of GCs (GC1, GC2, GC3, and GC4) found as a result of CA are 1.0115, 1.0735, 1.0721, and 1.2483, respectively. So, GCs of GC4 are dynamically much evolved compared to those in GC1, GC2, and GC3, respectively. Since we know that the most evolved GCs are the roundest with respect to ellipticities, the GCs of GC4 are more evolved than others. In summary, from the above facts and values of the peaks of the colors in these four groups we can conclude that the GCs of GC4 and GC1 are more evolved than those in GC2 and GC3.

The mean values of $(V-I)_0$ for GC1 and GC4 are similar to the mean value of G2 whereas the mean values of GC2 and GC3 are similar to that of G1 (Tables 4 and 5). G1 galaxies can be considered normal sites for the formation of GCs in GC2

and GC3 which are dynamically less evolved (viz., ⟨Age⟩ ∼ 5 Gyr for GC4, Table 5). On the other hand, G2 galaxies can be considered places for the formation of GCs in GC1 and GC4 which are dynamically more evolved and, hence, older (age ∼ 7.2 Gyr for GC3, Table 5). Though the ellipticities and $\log(r_t/r_c)$ for GC1 do not support the above fact, the higher mean value of color $(V-I)_0$ indicates that GC1 contains redder GCs which is an indication of older age. The GCs in GC1 and GC4 are formed by a mechanism other than self-enrichment (viz., $r(V0,(V-I)_0) \sim -0.199$, $p = 0.350$ for GC1 and $r(V0,(V-I)_0) \sim -0.155$, $p = 0.49$ for GC4). From the histograms of colors (Figure 7) and color versus projected distance (Figure 9) of the groups GC1, GC2, GC3, and GC4 it is clear that the highest peaks of GCs in GC1 and GC4 occur at higher values of $(V-I)_0$. But for GCs in GC2 and GC3 the heights of the peaks at the modes are not very different from one another and they occur at much lower values of $(V-I)_0$ (viz., 0.3/0.5, Table 5). If the correlations between color and projected distance are calculated for GC1, GC2, GC3, and GC4, these are (0.102, $p = 0.635$), (0.402, $p = 0.109$), (0.234, $p = 0.164$), and (0.091, $p = 687$), respectively. So it is clear that for the GCs of GC2 and GC3 the correlations are moderate at a 10% level of significance.

The correlations between magnitude and projected distance of the above four groups are (0.015, $p = 0.945$), (0.560, $p = 0.019$), (0.373, $p = 0.023$), and (−0.180, $p = 0.422$), respectively (Figure 10). This is an indication that the star formation history can be considered to be similar for groups GC1 and GC4. Since the GCs in these groups are more evolved than those in GC2 and GC3, and the tidal indices are higher for those in G2 which are their places of formation, it is very likely to assume that the GCs in the outer parts of GC1 and GC4 are tidally stripped from their host galaxies. This does not hold for GCs in GC2 if their host galaxies (which are considered to be G1) possess a high degree of rotation which is the case as discussed before regarding the rotation of their total H I masses though this is not true for GCs of GC3.

6. SUMMARY AND CONCLUSIONS

In the present work, a statistical approach for classifying LSB dwarf galaxies and GCs has been developed. For this classification, two statistical techniques are used, viz., PCA followed by *K* means CA together with the criterion for finding the optimum number of homogeneous groups. Through PCA the optimum set of parameters giving maximum variation among the objects is found while the required homogeneous groups are found using CA. The optimum number of groups is found following Sugar & James (2003). For the sample of dwarf galaxies two groups are found primarily indicative of their masses (M_{V0}), tidal indices (Θ), and surface brightnesses averaged over effective radius (SBV_{e0}) but irrespective of their morphological indices.

G1 galaxies are massive ($M_{V0} < -13$) with a larger amount of H I mass having a higher degree of rotation (V_m) and a lower mean value of tidal index in the absence of any correlation ($r \sim 0.09/0.1$, $p \gg 0.05$; viz., Table 4) with scaling parameters. Also moderate mass–metallicity correlations exist among the dwarf galaxies of G1. All these facts indicate that dwarf galaxies of G1 are formed by self-enrichment supported by stellar and supernovae feedback. On the other hand, G2 galaxies are less massive, have an insignificant amount of H I mass with little or no rotation, are devoid of any mass–metallicity correlation, and have high values of tidal indices with significant correlations ($r \sim 0.9/0.8$, $p < 0.05$; viz., Table 4) with the scaling parameters. The above-mentioned characteristics suggest that environment plays a very important role in the star formation scenario of these dwarf galaxies.

Subsequently a classification of GCs in the LV has been carried out and four groups emerged as a result of this classification. A comparison of the color profiles of the GCs in these groups with those of dwarf galaxies suggests that among the four groups, GC1 and GC4, which are dynamically more evolved, may have formed in G2 whereas dynamically less evolved GCs in GC2 and GC3, having no significant self-enrichment, may have formed in galaxies like G1. Also the colors of GCs in GC1 and GC4 bear no correlations with their projected distances while a moderate correlation exists for the GCs of GC2 and GC3. This is also true for magnitude versus projected distance correlations. Thus, the star formation history for the GCs of GC1 and GC4 might be speculated to be different from those in GC2 and GC3.

T.C. thanks the Department of Science and Technology (DST), India for awarding her a major research project for the work. The authors are grateful to Prof. A. K. Chattopadhyay and Emmanuel Davoust for their useful suggestions and help. The authors are also thankful for the suggestions of the referee. M.S. acknowledges partial support from the grant RFBR 08-02-00627.

REFERENCES

Babu, G. J., Chattopadhyay, T., Chattopadhyay, A. K., & Mondal, S. 2009, ApJ, 700, 1768
Carraro, G., Chiosi, C., Girardi, L., & Lia, C. 2001, MNRAS, 327, 69
Chattopadhyay, A. K., Chattopadhyay, T., Davoust, E., Mondal, S., & Sharina, M. 2009, ApJ, 705, 1533
Chattopadhyay, T., & Chattopadhyay, A. K. 2006, AJ, 131, 2452
Chattopadhyay, T., & Chattopadhyay, A. K. 2007, A&A, 472, 131
Da Costa, G. S., Grebel, E. K., Jerjen, H., Rejkuba, M., & Sharina, M. E. 2009, AJ, 137, 4361
Dekel, A., & Birnboim, Y. 2006, MNRAS, 344, 1131
Dekel, A., & Silk, J. 1986, ApJ, 303, 39
Fraix-Burnet, D., Choler, P., & Douzery, E. J. P. 2006, A&A, 455, 845
Geisler, D., George, W., Verne, V. S., & Dana, I. C. 2007, PASP, 119, 939
Georgiev, I. Y., Puzia, T. H., Goudfrooij, P., & Hilker, M. 2010, MNRAS, 406, 1967
Grebel, E. K. 1999, in IAU Symp. 192, The Stellar Content of Local Group Galaxies, ed. P. Whitelock & R. Cannon (San Francisco, CA: ASP), 17
Grebel, E. K., Gallagher, J. S., & Harbeck, D. 2003, AJ, 125, 1926
Haines, C. P., Barbera, F. L., Mercurio, A., Meriuzzi, P., & Busarello, G. 2006, ApJ, 647, L21
Hirashita, H. 2000, PASJ, 52, 107
Hubble, E. P. 1922, ApJ, 56, 162
Hubble, E. P. 1926, ApJ, 64, 321
Johnson, R. A., & Wichern, G. 1998, Applied Multivariate Statistical Analysis (4th ed.; Upper Saddle River, NJ: Prentice-Hall)
Karachentseva, V. E., Karachentsev, I. D., & Boerngen, F. 1985, A&AS, 60, 213
Karachentsev, I. D., Karachentseva, V. E., Huchtmeier, W. K., & Makarov, D. I. 2004, AJ, 127, 2031
Karachentsev, I. D., & Makarov, D. I. 1998, in Proc. IAU Symp. 186, Galaxy Interactions at High and Low Redshift, Kyoto, ed. J. Barnes & D. B. Sanders (Dordrecht: Kluwer), 109
Kormendy, J. 1985, ApJ, 295, 73
Kroupa, P. 2002, MNRAS, 330, 707
Kumai, Y., Hashi, Y., & Fujimoto, M. 1993, ApJ, 416, 576
Lisker, T., Grebel, E. K., Binggeli, B., & Glatt, K. 2007, ApJ, 660, 1186
Little, R. J. A., & Rubin, D. B. 2002, in Statistical Analysis with Missing Data, ed. R. J. A. Little & D. B. Rubin (2nd ed., Wiley Series in Probability and Statistics; New York: Wiley)
MacQueen, J. 1967, in Proc. Fifth Berkeley Symp. on Mathematical Statistics and Probability, Vol. 1 (Berkeley, CA: Univ. California Press), 281
Marin-Franch, A., & Aparicio, A. 2002, ApJ, 568, 174

Milligan, G. W. 1980, Psychometrika, 45, 325
Murtagh, F., & Heck, A. 1987, Multivariate Data Analysis (Dordrecht: Reidel)
Penny, S. J., & Conselice, C. J. 2008, MNRAS, 383, 247
Puzia, T. H., & Sharina, M. E. 2008, ApJ, 674, 909
Sharina, M. E., Puzia, T. H., & Makarov, D. I. 2005, A&A, 442, 85
Sharina, M. E., et al. 2008, MNRAS, 384, 1544
Shaya, E. J., & Tully, R. B. 1984, ApJ, 281, 56
Skillman, E. D., Kennicutt, R. C., & Hodge, P. W. 1989, ApJ, 347, 875
Smith, G. H. 1985, PASP, 97, 1058
Sugar, A. S., & James, G. M. 2003, J. Am. Stat. Assoc., 98, 750
Tully, R. B., et al. 2006, AJ, 132, 729
van den Bergh, S. 2007, AJ, 134, 344
van den Bergh, S. 2008, MNRAS, 385, L20
Vaduvescu, O., & McCall, M. L. 2008, A&A, 487, 147
White, S. D. M., & Rees, M. J. 1978, MNRAS, 183, 341
Whitmore, B. C. 1984, ApJ, 278, 61
Woo, J., Courteau, S., & Dekel, A. 2008, MNRAS, 390, 1453
Vilchez, J. M. 1997, RevMexAA, 6, 30