Multivariate Study of the Star Formation Rate in Galaxies: Bimodality Revisited

T. Chattopadhyay*,[1] D. Fraix-Burnet,[2] S. Mondal [3]
[1] *Department of Applied Mathematics, University of Calcutta, Kolkata, India*
[2] *Univ. Grenoble Alpes, CNRS, IPAG, F-38000 Grenoble, France*
[3] *Department of Statistics, Bethune College, Kolkata, India*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Subjective classification of galaxies can mislead us in the quest of the origin regarding formation and evolution of galaxies. Multivariate analyses are the best tools used for such kind of purpose to better understand the differences between various objects, in an objective manner. In the present study an objective classification of 362 923 galaxies of the Value Added Galaxy Catalogue (VAGC) is carried out with the help of three methods of multivariate analysis. First, independent component analysis (ICA) is used to determine a set of derived independent variables that are linear combinations of various observed parameters (viz. ionized lines, Lick indices, photometric and morphological parameters, star formation rates etc.) of the galaxies. Subsequently, K-means cluster analysis (CA) is applied on the independent components to find the optimum number of homogeneous groups. Finally, a stepwise multiple regression is carried out on each group to predict and study the star formation rate as a function of other independent observables. The properties of the ten groups thus uncovered, are used to explain their formation and evolution mechanisms. It is suggested that three groups are young and metal poor, belonging to the blue sequence, three others are old and metal rich (red sequence). The remaining four groups of intermediate ages cannot be classified in this bimodal sequence: two belong to a pronounced mixture of early and late type galaxies whereas the other two mostly contain old early type galaxies. The above result is indicative of a continuous evolutionary scenario of galaxies instead of two discrete modes, blue and red, so far suggested by previous authors. Some of our groups occupy the transition region with different quenching mechanisms. This establishes the elegance of a multivariate analysis giving rise to a sophisticated refinement over subjective inference.

Key words: galaxies: general – galaxies: evolution – methods: statistical

1 INTRODUCTION

Investigating the formation and evolution of galaxies is becoming a complicated process with the increased availability of huge database as a result of instrumental improvements. A good understanding of the underlying physical process requires synthetic databases simulated numerically which are replica of real life databases. Hubble's subjective classification based on galaxy morphology ignores many significant observables e.g. kinematics, chemical composition etc.

According to various studies, classical formation of galaxies have been proposed to follow five major trajectories: (i) the monolithic collapse model, (ii) the major merger model, (iii) the multiphase dissipational collapse model, (iv) accretion and (v) in situ hierarchical merging. But no model uniquely explains the formation of all galaxies.

The historical and still most common approach to the classification of galaxies, is based on physical criteria, like apparent traits (i.e. morphology, emission line properties etc.) or more or less understood processes (starbursts, AGNs etc.). With the advent of multi-wavelengths and multivariate data bases, the goal of many studies has been to find the parameters that best characterize the established classes. One such example is given by the BPT diagrams (Baldwin et al. 1981; Veilleux & Osterbrock 1987) which tries to split different kinds of ionizing sources using a few emission lines.

The so-called bimodality of galaxies is a counter example to this traditional approach. The accumulation of observations have shown that for several properties, galaxies show two distribution peaks that do not easily match pre-

* tchatappmath@caluniv.ac.in

established physically motivated classes. Extragalactic studies have now entered the statistical era, with the complexity of galaxies challenging our approaches to understand them.

In this context, one is tempted to apply multivariate partitioning analysis to find homogeneous groups, not focusing on only one aspect of the physics of galaxies, so that the formation and evolutionary history can be demonstrated satisfactorily. One basic tool is principal component analysis and it has been used by many authors (e.g. Whitmore 1984; Watanabe et al. 1985; Cabanac et al. 2002; Chattopadhyay & Chattopadhyay 2006; Peth et al. 2015) but this is not an appropriate clustering (unsupervised classification) tool. Some attempts have been made by K-means cluster analysis (Ellis et al. 2005; Chattopadhyay & Chattopadhyay 2007; Chattopadhyay et al. 2007, 2008, 2009a,b; Sánchez Almeida et al. 2010; Fraix-Burnet et al. 2010, 2012; De et al. 2016). Though sophisticated statistical techniques are being developed steadily, multivariate approaches are not widely used across the astronomical community (see a review in Fraix-Burnet et al. 2015).

Some observations are as follows: a partitioning of objects into robust groups is possible when the parameters are independent as well as physically significant. Observationally the information is usually summarized into broad-band fluxes (magnitudes), slopes (colors), medium-band and line fluxes (Lick indices). Multivariate partitioning groups objects according to their similarity. They show a descriptive diversity but cannot explain it. We need numerical simulation to understand and explain the diversity.

In the present work, we have taken a large dataset from SDSS data archive including various observables regarding morphology, chemical composition and kinematics and used multivariate statistical techniques to explore and explain the underlying diversities. The present work is fundamental in several ways since we use:

- a large dataset of galaxies,
- a larger number of observables at a time,
- a more sophisticated statistical technique, Independent Component Analysis (ICA) which takes into account the discrimination through the independent components instead of parameters,
- a method (ICA) that is applicable to non-Gaussian data,
- star formation rate, used as support of the evolutionary status, is explored with a stepwise regression technique.

A brief description of the data set is given in Sect. 2. The methods are described in Sect. 3, The results and discussion are included under sections Sect. 4 and 5 respectively, before the conclusion in Sect. 6.

2 DATA SET

The NYU Value-Added Galaxy Catalog (VAGC Blanton et al. 2005; Padmanabhan et al. 2008; Abazajian et al. 2009) is a cross-matched collection of galaxy catalogs maintained for the study of galaxy formation and evolution[1]. It is based on the Sloan Digital Sky Survey Data Release 7 (SDSS-DR7[2]).

In the raw table, 2,506,754 objects are available. We have selected only galaxies, which disregards QSOs and stars, ending up with 865 333 entries. We have then restricted the sample to $z < 0.2$ and a good signal to noise ratio S/N > 10. This leaves us with 362 923 galaxies.

After eliminating redundant properties, the parameter set used in the present analysis consists of 49 parameters which covers photometry, spectroscopy, morphology, chemical composition and kinematics. Star formation rates and specific star formation rates are also included. All these parameters are described in Table A1 and details are given on the source website[3].

3 STATISTICAL ANALYSES

3.1 Shapiro-Wilk test

The non Gaussian nature of the data set has been explored by the statistical test Shapiro-Wilk test (1965) in which the test statistic is defined by W = $\sum_{i=1}^{n} a_i x_i^2 / \sum_{i=1}^{n} (x_i - \bar{x})^2$, where n is the number of observations, x_i's are ordered sample values and a_i's are constants generated from the order statistics of a sample from normal distribution. In the present situation a multivariate extension has been used. The p value of the test is less than 2.17×10^{-13}, which is very small to confidently reject the null hypothesis. Therefore, the present data set is found to be non Gaussian in nature.

3.2 Independent Component Analysis

We have already mentioned that Principal Component Analysis (PCA) has been applied by many authors (Murtagh & Heck 1987; Brosche 1973; Whitmore 1984, etc) but it is not appropriate for clustering and classification. Also it is applicable for Gaussian data which is not the present case. On the other hand, Independent Component Analysis (ICA) is applicable to non Gaussian data set like the present situation and it is also a dimension reduction technique like PCA. In addition of being uncorrelated, the components here found are also independent, i.e., it reduces the number of observable parameters p to a number m (m << p) of new parameters such that these m parameters are mutually independent. Mathematically speaking, let X_1, X_2, X_3, ..., X_p be p random vectors (here p parameters, p = 49) and n (here 362923) be the number of observations of each X_i, (i = 1, 2, 3, ..., p).

Let X = AS, where S = $[S_1, S_2, S_3, ..., S_p]'$ is a random vector of hidden components S_i, (i = 1, 2, 3, ..., p) such that S_i's are mutually independent and A is a non singular matrix. Then the objective is to find S by inverting A, i.e., S = A^{-1}X or S = WX. Using ICA we find the unmixing matrix W such that any two functions $g_1(S_i)$ and $g_2(S_j)$, i ≠ j has covariance zero i.e. the ICs are independent (for more details

[1] http://sdss.physics.nyu.edu/vagc/

[2] http://classic.sdss.org/dr7/

[3] http://wwwmpa.mpa-garching.mpg.de/SDSS/DR7/raw_data.html

see Comon 1994; Chattopadhyay et al. 2013, and references therein).

Presently there is no good method available for the determination of the optimum number of ICs. In this work, the optimum number of ICs have been chosen by the optimum number of Principal Components (PCs) (Albazzaz & Wang 2004), to find m (m << p) (Babu et al. 2009; Chattopadhyay & Chattopadhyay 2007; Fraix-Burnet et al. 2010; Chattopadhyay et al. 2010). In the present situation, we have first performed PCA to find the significant number of ICs. In PCA the maximum variation with significantly high eigenvalue (viz. $\lambda \sim 1$) was found to be almost 90% for nine PCs. Hence, we have chosen nine ICs for cluster analysis (CA).

3.3 K-means cluster analysis

K-means cluster analysis (CA) is a multivariate technique for finding coherent groups in a data set giving information of the underlying structure. In this method, one finds K groups, provided each group contains an object and an object belongs to exactly one group. Details of algorithm and applications are found in MacQueen (1967); Chattopadhyay et al. (2009a, 2010, 2012, 2013); Das et al. (2015),

The number K of groups is an input to the algorithm. The optimum value of K is found as follows. First we have found groups assuming K = 1, 2, 3, ...etc. Then a measure, called distortion (viz. d_K), which is a function of distances between the data points, is computed by the following formula $d_K = (1/p)min_x E[(x_K - c_K)^{'}(x_K - c_K)]$, which is the distance of x_K vector (data point) from the centroid c_K of the corresponding group. The optimum value of K is that for which the jump $J_K = (d_K^{-p/2} - d_{K-1}^{-p/2})$ is maximum (Sugar & James 2003).

In this study, we have performed a K-means CA with respect to the ICs and have found the optimum number of groups to be K = 10. We name the groups K1-K10.

3.4 Stepwise multiple regression

Multiple regression is the prediction of one dependent variable (response) in terms of other independent variables (predictors). In stepwise regression, a small set of predictors are chosen from a large set still having good predictive ability. The first predictor for entry into the equation is the one with the largest positive or negative correlation with the dependent variable (here SFR and specific SFR). The statistics used in the entry of such a variable is

$$F_{change} = \frac{R^2_{change}(n-p+1)}{q(1-R^2)}$$

where n is the number of observations, p is the number of parameters, q is the number of parameters entered at the step and $R^2_{change} = R^2$ - $R^2_{(i)}$, where $R^2_{(i)}$ is the square of multiple correlation coefficient when all independent variables except *ith* one are in equation and R^2 is the square of the multiple correlation coefficient.

Table 1. Parameters with highest correlation coefficients (in parentheses) for each Independent Component.

IC	Most influencial parameters
ICA1	EW(OII_3729)(0.83), Lick_Fe5015 (0.53)
ICA2	v$_{disp}$ (0.97), J (-0.7)
ICA3	σ_{balmer} (-0.60)
ICA4	EW(OIII_5007) (0.98)
ICA5	Lick_Fe5015 (0.83)
ICA6	Sersic_r90_R(-0.99)
ICA7	σ_forb (-0.60), σ_{balmer} (0.42)
ICA8	Sersic_amp_R (-0.99)
ICA9	EW(NII_6584)(0.79), EW(Hα)(0.75), EW(SII_6731)(0.72)

Table 2. Distribution of galaxies in the ten groups found with K-means.

Group	Number of galaxies
K1	79576
K2	109188
K3	12630
K4	17336
K5	40045
K6	10552
K7	16325
K8	1375
K9	68050
K10	7846

4 RESULTS

4.1 Properties of the ICs

Performing Independent Component Analysis for the VAGC data set, we have taken nine significant ICs. Then we have found the parameters with highest correlations for each component. They have been listed in Table 1.

From Table 1, it is clear the nine ICs represent five kinds of properties: 1) velocity dispersion (ICA 2, ICA 3, ICA 7), 2) ionisation (ICA 9, ICA 4), 3) metallicity (ICA 1, ICA 5), 4) surface brightness (ICA 8) and 5) structural properties (ICA 6). Thus complete description of the physics of the galaxies can be reduced to this five independent characteristics by means of ICA. Hence these nine independent components are used instead of the initial 49 parameters for Cluster Analysis (CA). This is the goal for a dimension reduction technique like ICA.

4.2 Properties of the galaxies in the ten groups

The cluster analysis divided the galaxies into ten groups, K1-K10. The distribution of galaxies within these groups is given in Table 2. It appears that the four groups K2, K1, K9, K5, in decreasing importance, already gather 82% of the objects (52% with K2 and K1 only), the K8 group being very small.

The boxplots (Fig. B1) summarize the statistics of each parameter for the ten groups. It is interesting to note that the dispersions are nearly always relatively small, indicating that the groups found by the cluster analysis are quite

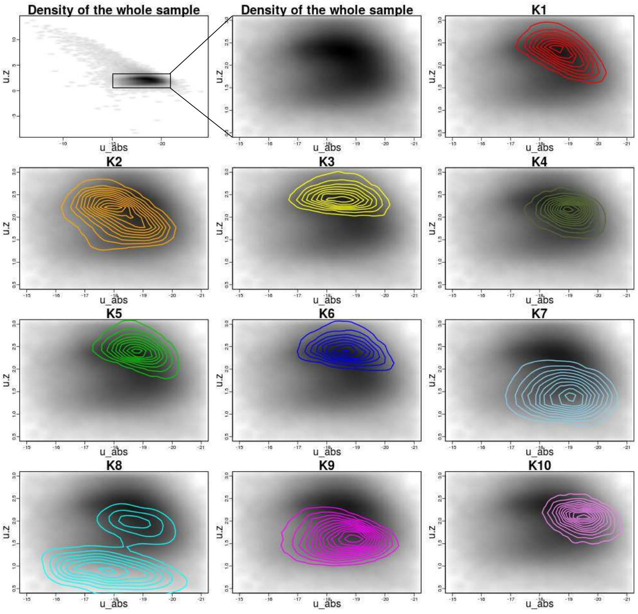

Figure 1. Colour magnitude (u-z vs. U) diagrams of the whole sample along with the classified groups K1 to K10.

homogeneous. This is particularly striking for the biggest groups, K1 and K2. There are often large overlaps, but also clearly separated properties distributions between groups in many instances.

The properties of the groups taken together are clearly distinct for some parameters, such as v$_{disp}$, J or NaD_abs, while they look similar for others, such as EW(Hα_{abs}) or H-K.

The small group K8 stand out in many parameters, most spectacularly in EW(OIII_5007). This group is however close to the groups K7 and the big K9 for many properties, such as Ca_K_abs, Sersic_n_R or Lick_G4300.

Instead of describing in great details the relative properties group by group, we concentrate in the following on some more general physical interpretation that some particular plots are known to provide and that are familiar to astronomers.

4.3 Color-magnitude diagram

The well known bimodality of galaxies is seen on the color magnitude diagrams in Fig. 1 with a crescent shape of the distribution of the whole sample, with the so-called red and blue branches (or sequences).

The groups K3, K5 and K6 clearly belong to the red branch, while K7 to K9 are essentially on the blue one. The groups K1 and K2 span both branches, including the region in between often called the green valley. Groups K4 and K10 are part of this green valley. The very small group K8 appears peculiar, with a very blue part, and 25% of its galaxies belonging to the red branch.

Even though the correspondence between our groups and the rough usual division in red, green or blue regions of the plot is satisfactory, the multivariate clustering offers more subtle and objective categories of galaxies that can be investigated further by analyzing other properties.

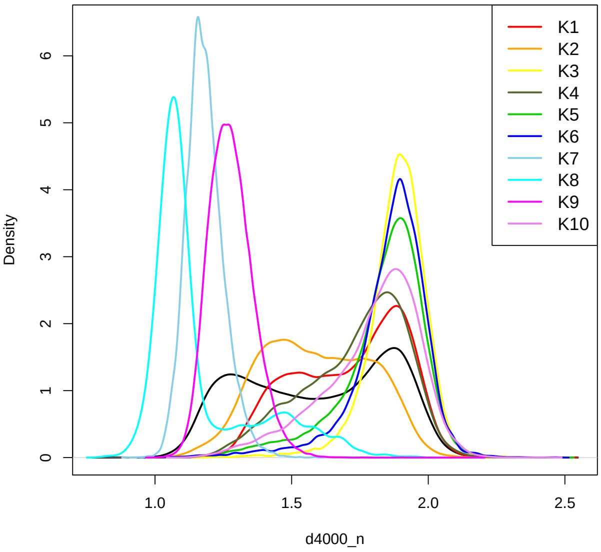

Figure 2. Density distribution of D4000_n. The black line corresponds to the whole sample. Thirty four galaxies with D4000_n> 2.5 are not shown.

4.4 D4000_n

In Fig. 2, a density plot of D4000_n (or Dn_4000 used in some literature) is shown for the ten groups and the whole sample. The limit D4000_n > 1.55 is generally used to define quenched galaxies (e.g. Kauffmann et al. 2003; Haines et al. 2016).

A bimodality is clearly seen in Fig. 2 with two peaks: groups K7, K8, K9 have smaller values of D4000_n, hence contain younger stellar populations compared to K3, K4, K5, K6 and K10 groups which have peaks at larger values of D4000_n, K4, K10 having a longer tail toward smaller values. The two biggest groups K1 and K2 show a larger span but K1 tends to peak at high values and K2 at low values. There is a hint of a bimodality, which is more obvious on the distribution of D4000_n for the whole sample, indicating a mixture population of old and young populations.

However, this bimodality does not match well with the one seen on the color-magnitude diagram (Fig. 1). For instance, the group K7 seems to extend in the red branch, while the D4000_n distribution is entirely below 1.6. This shows that the age of the stellar population alone does not explain the blue, red and green valley categories.

4.5 EW(Hδ_{abs})

At the same time Hδ in absorption is detected in a galaxy spectrum when the massive hot stars just finish their evolution on the main sequence i.e., at least 0.1-1 Gyr after a star burst is truncated. Thus EW(Hδ_{abs}) gives a measure of the age of the youngest stellar population in a galaxy and it is a widely used indicator to determine the mean stellar ages as well (Worthey & Ottaviani 1997). It is a better age estimator than Balmer absorption lines due to the presence of Balmer emissions from HII regions, AGN

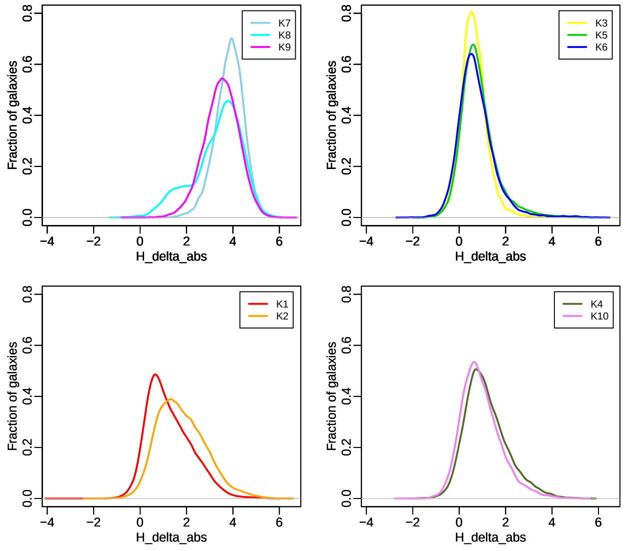

Figure 3. Fraction of galaxies in EW(Hδ_{abs}) distribution in the young star forming groups (viz. K7, K8, K9), old quiescent groups (viz. K3, K5, K6) and mixture groups (viz. K1, K2, K4, K10).

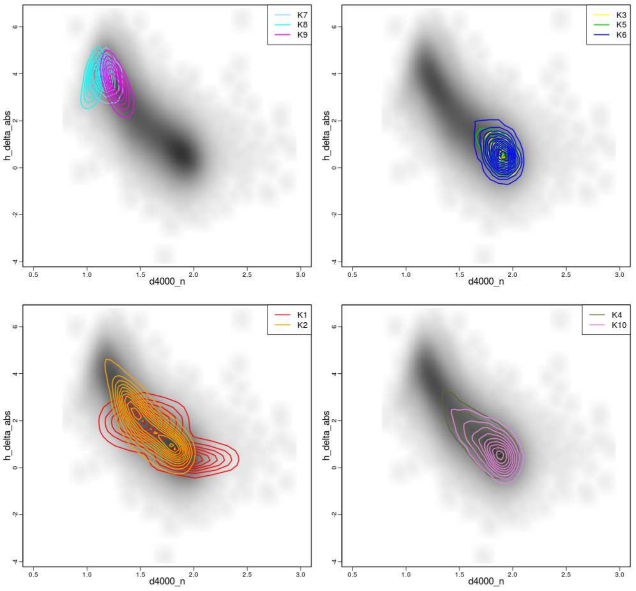

Figure 4. D4000_n vs EW(Hδ_{abs}) plot for the classified groups K7, K8, K9 (top left), K3, K5, K6 (top right), K1, K2 (bottom left) and K4, K10 (bottom right). The grey points in the background represent the whole sample.

and/or planetary nebulae causing age estimates spuriously high (Trager et al. 2000; Prochaska et al. 2007). EW(Hδ_{abs}) remains unaffected by the above factors (Osterbrock 1989; Worthey & Ottaviani 1997).

We plot the fraction of galaxies vs EW(Hδ_{abs}) (viz. Fig. 3) and EW(Hδ_{abs}) vs D4000_n (viz. Fig. 4) for the groups K7, K8, K9 taken together, K3, K5, K6 taken together, K1, K2 and K4, K10 taken together. The figures clearly show that K7, K8, K9 have younger populations, K3, K5, K6 have older populations, K1, K2 have widely mixture populations and K4, K10 have been dominated by older ones.

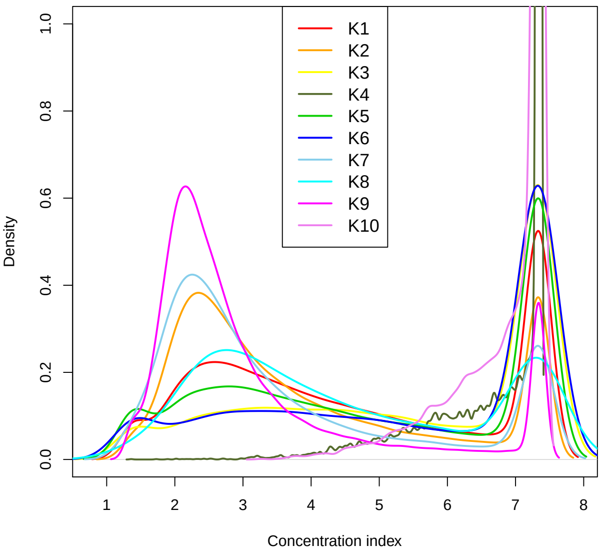

Figure 5. Density distribution of the concentration index C of galaxies in the groups for K1 to K10.

4.6 Morphology

In Fig. 5, we plot the density distribution of the concentration index C = Sersic_r90_R / Sersic_r50_R for all the ten groups. Late-type morphologies are characterized by a low C (< 2.6, Strateva et al. 2001). Note the strong dichotomy with two peaks at $C \simeq 2.5$ and 7.5.

The groups K4 and K10 are nearly entirely made of early-type galaxies while all the other groups are mixtures of both categories, with a higher fraction of late-type galaxies in K7 and K9 and of early-type ones in K1, K3, K5, K6.

4.7 BPT diagrams

Two emission line ratios were recommended by Baldwin et al. (1981, hereafter BPT) and are often used to discriminate between star forming and composite galaxies and the AGN-dominated galaxies. This diagram is also the basis for more refined empirical classifications based on theoretical population synthesis and photoionization models (Veilleux & Osterbrock 1987; Kauffmann et al. 2003; Kewley et al. 2006; Kewley et al. 2013). Two other similar diagrams, BPT-SII and BPT-OI, were proposed by Veilleux & Osterbrock (1987).

The standard classification scheme on this diagram is based on equations of curves separating the different classes. These cuts are sharp, somewhat arbitrary. Its main goal is to distinguish between different ionization source (basically thermal or non-thermal) while the properties used for

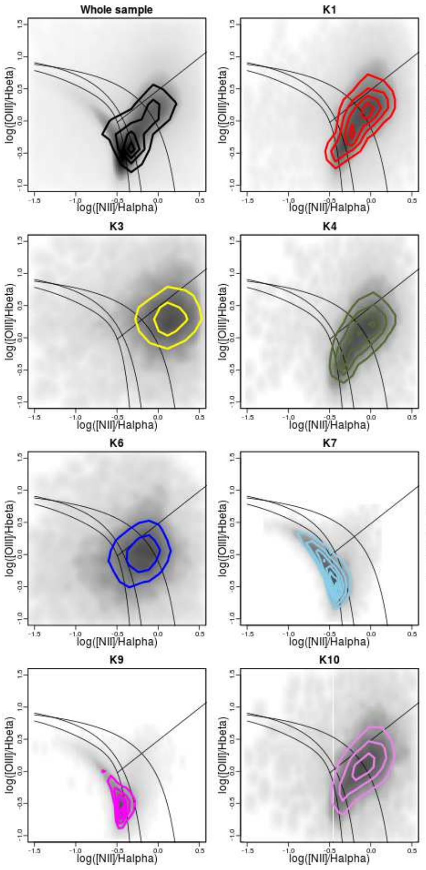

Figure 6. The distribution of galaxies in the groups K1 to K10 in the BPT diagram: [OIII]/Hα vs [NII]/Hβ. Curves are from Kewley et al. (2001); Kauffmann et al. (2003); Stasińska et al. (2006) (right to left respectively).

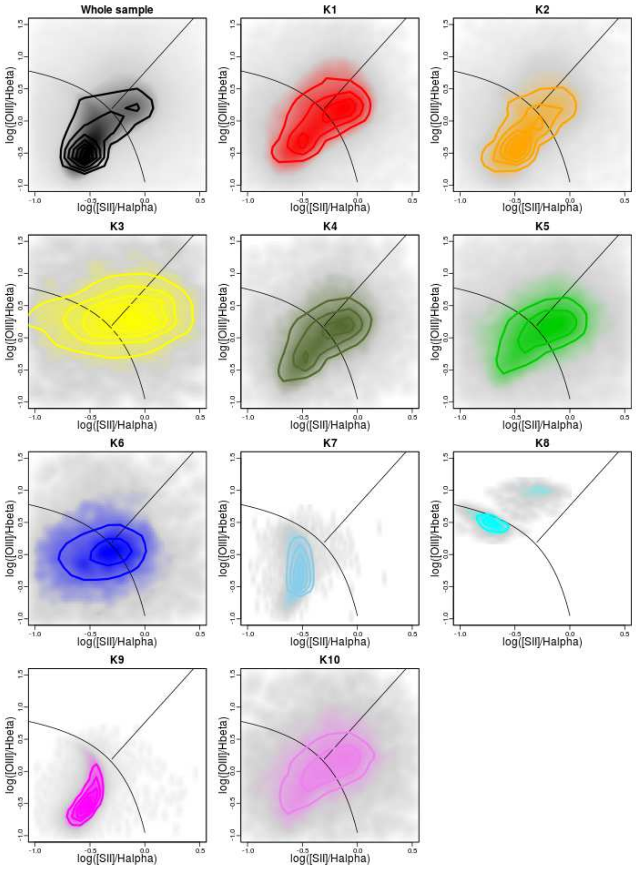

Figure 7. The distribution of galaxies in the groups K1 to K10 in a different version of the BPT diagram: [OIII]/Hβ vs [SII]/Hα. Curves from Kewley et al. (2001).

our clustering analysis are not limited to this peculiar aspect of galaxies. It also appears that the classification based on these diagrams is not clearcut. For instance, "the current LINER classification scheme encompasses two or more types of galaxies, or galaxies at different stages in evolution" (Kewley et al. 2006). Cid Fernandes et al. (2010) proposed a new cut between Seyfert and LINERs to resolve this ambiguous class. They also present an interesting and critical discussion of the classifications based on the BPT diagrams. An important reminder in their discussion is that the star-forming region delimitation is rather arbitrary in the lower part of the diagram where most of the galaxies lie. Also, the upper right part of the diagram is not composed of pure AGNs. Finally, different cuts are proposed by different authors (e.g. Kewley et al. 2001; Lamareille 2010). Interestingly, all these works tend to suggest that more parameters are probably required to fully understand different kinds of galaxies and ionization processes (Richardson et al. 2016).

Indeed, an objective classification with soft frontiers performed with the diagnostics line ratios indicates a somewhat different picture with only four categories, some including for instance sub-divisions like strong or weak AGNs (de Souza et al. 2017). However while being multivariate, this study has a small number and variety of parameters.

An additional limitation of the classifications based on single kinds of ionization processes is that the galaxies are indeed very probably a mixture of different regions (Belfiore et al. 2016).

These limitations on the significance of the cuts should be kept in mind when comparing our multivariate clustering results and diagnostics diagrams (Figs. 6, 7 and 8).

It is clear from the three diagnostics diagrams that K7 and K9 have no AGN, hence are pure star forming galaxies, whereas the very small group K8 is quite peculiar with two peaks, one in the pure star forming region, the other one in the pure AGN zone. Groups K2 and K6 are clearly intermediate between star forming galaxies and AGNs.

All the other groups (K1, K3, K4, K5 and K10) are AGNs (Fig. 6). This interpretation is however less clear from Figs. 7 and 8 since the groups K1, K4, K5, K6 and K10 appear rather as a mixture of star forming and mainly LINERs galaxies. Only K3 seem to be mainly composed of LINERs.

Our results are in the line of the discussion found in the literature as presented above. Our multivariate analysis shows that only some galaxies can be identified with simple properties, most others are more complex and confirms that more parameters than these diagnostics diagrams are required.

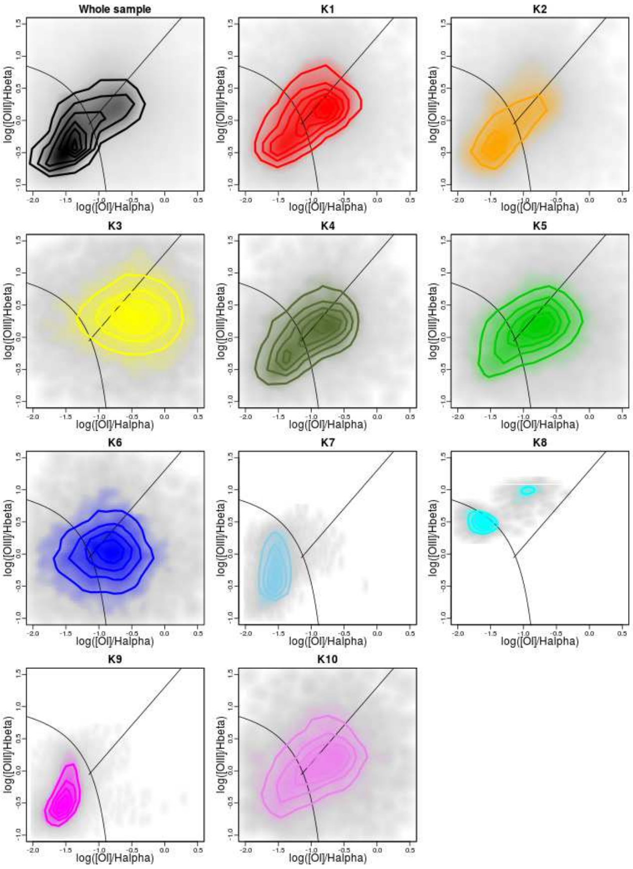

Figure 8. The distribution of galaxies in the groups K1 to K10 in a different version of the BPT diagram: [OIII]/Hβ vs [OI]/Hα. Curves from Kewley et al. (2001, 2006).

4.8 Fe5270 vs Mgb

A galaxy having younger age may be due to lowest value of Lick_Fe5270 and Mgb, but younger age is also sensitive to the colour of the horizontal branch (HB), higher values meaning a bluer HB. Bluer HB in turn can be due to He-enriched stars, hence higher α/Fe values.

We have placed the mean values and the associated standard errors of the ten groups in the scatter plot Lick_Fe5270 vs Mgb (Fig. 9). Also shown are the isochrones of SSP model from Thomas et al. (2011) at 9 (red) and 12 (blue) Gyrs respectively for [α/Fe] = -0.3, 0, 0.3 and 0.5 respectively.

Groups K3, K5, K6, K9 and K10 are compatible with $\alpha/Fe = 0.3$, K1 and K4 may have a slightly lower value, still lower for K2 with $\alpha/Fe \simeq 0.15$, while K7 is compatible with $\alpha/Fe \simeq 0.5$ and K8 with a much higher value. Group means of K7, K8, K9 fall in the low metallicity region, with a high α/Fe, indicating they are young starbursts galaxies.

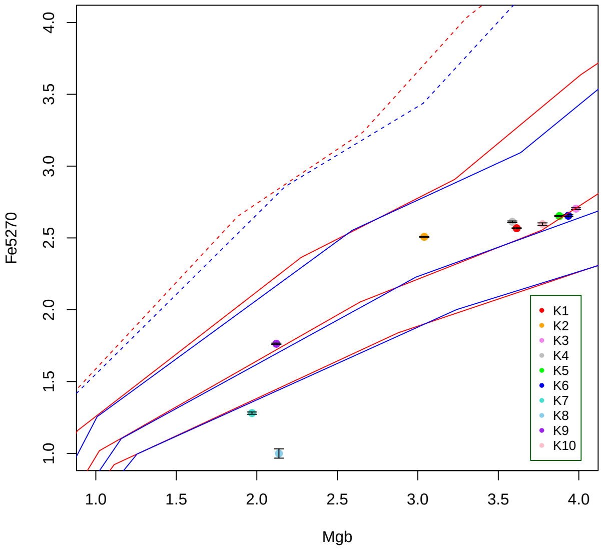

Figure 9. Mean values of Lick_Fe5270 vs Lick_Mgb for K1 to K10. Blue and red lines are for stellar populations models of Thomas et al. (2011) for the age of 9 Gyr (red lines) and 12 Gyr (blue lines). Dotted lines are for α/Fe = -0.3 and solid lines are for α/Fe = 0.0, 0.3, 0.5 respectively from top to bottom.

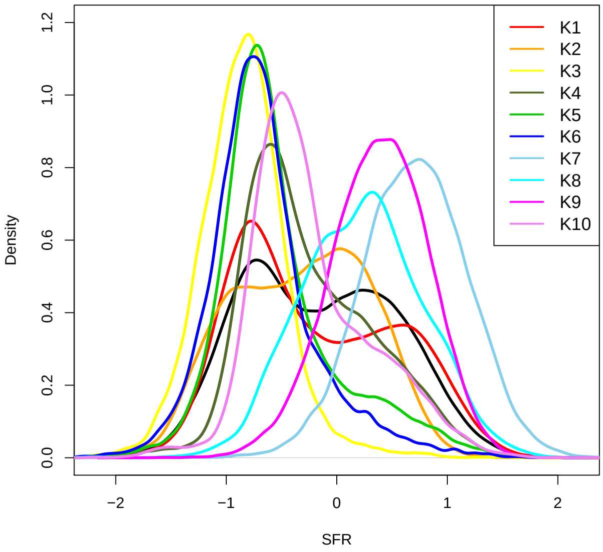

Figure 10. Density of the distribution of the SFR in the ten groups K1 to K10.

4.9 Star Formation Rate

We have performed a stepwise regression in order to check whether we can infer the star formation rate (SFR) and the specific SFR (specSFR) from the data without fitting some SED templates/models or using specific tracers as usually done (Brinchmann et al. 2004; Kennicutt 1998; Kennicutt & Evans 2012). This would be especially suited for very large data bases. Such a machine learning approach has been explored by Stensbo-Smidt et al. (2016) through a k-Nearest-Neighbour regression technique to estimate reliably specSFR and photometric redshifts by using only the u, g, r, i, and z photometric bands. In our case, we are also

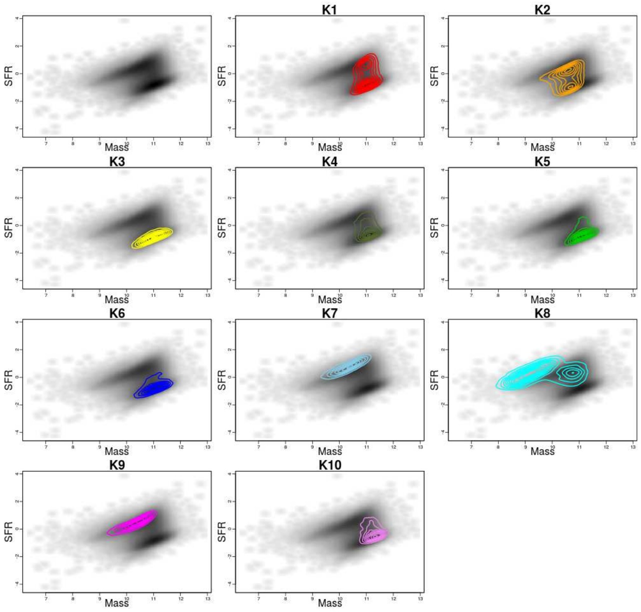

Figure 11. SFR vs Mass for the ten groups K1 to K10.

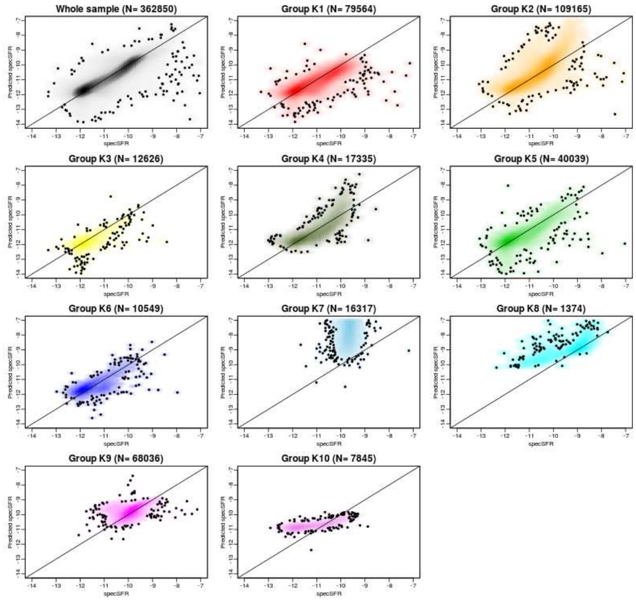

Figure 13. Same as Fig. 12 but for the specific SFR.

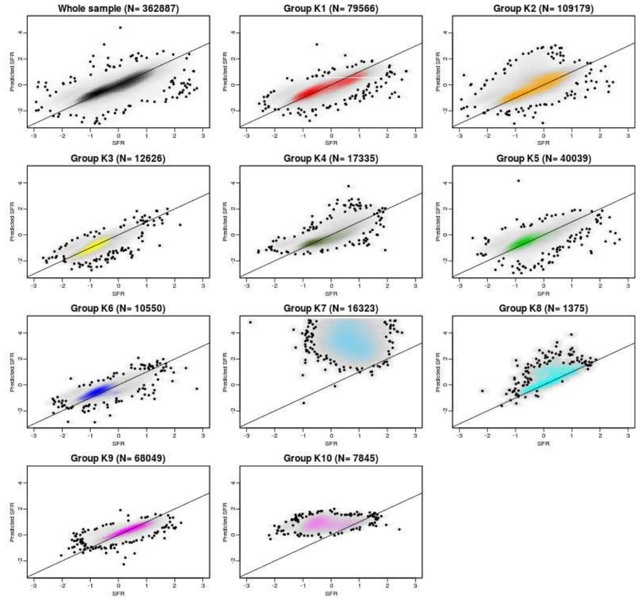

Figure 12. Plot of the SFR predicted by the regression using 45 parameters vs the computed SFR provided in the data set. The regression was performed on the whole sample and on each of the ten groups K1 to K10.

interested to see how much the most influent parameters depend on the groups and how good is the prediction.

The well-known bimodality of the SFR (Kauffmann et al. 2003) is shown in Fig. 10. Our groups are split in low (K1, K3, K4, K5, K6, K10) and high (K2, K7, K8, K9) SFRs. However, K2 is kind of intermediate while the other big group, K1, shows a significant fraction of galaxies with high SFRs. This is also illustrated in Fig. 11: most of the groups clearly belong to either the main sequence (quenched galaxies) or the starburst ones (Renzini & Peng 2015).

The predictive power of our regression analysis is illustrated in Figs. 12 and 13. It clearly appears that the analysis fails for groups K7 and K10, and is not very good for K8 in specSFR. Globally, it is less good for specSFR than for SFR.

The coefficients of the regression equations are given in Tables C1 and C2, the highest values (i.e. the most influential parameters) are indicated in boldface. In both SFR and specSFR, D4000_n dominates largely as expected since it is a known indicator of star formation. However, it is remarkable that it has a noticeably weaker influence in K7 and K9.

For SFR, the U and J magnitudes are important, revealing role of the global mass of the galaxy. In addition, the three colors $U-z$, $J-H$ and $H-K$ are positively correlated with SFR, in contrary to U, J and D4000_n which are negatively correlated.

For specSFR, the photometry has a much weaker influence but not totally negligible such as in K6, K8 and K9. The colors are also less inflent than for SFR, $U-z$ being the strongest one.

The very strong influence of the concentration index C in K4 for both SFR and specSFR is rather surprising and unique to this group.

Hence, it seems somewhat difficult to derive the SFR from a regression analysis for all galaxies with a high accuracy. This is possible only for some groups of galaxies for which the dispersion is slightly lower than for the entire sample. However, it is striking that this seems more problematic for specSFR than for SFR. In any case, some groups clearly show unexpected behaviors which probably tells something on their star formation activity.

5 DISCUSSION

From the results described in the previous section, we can distinguish four categories of groups.

5.1 Groups K1, K2

There two dominating groups which gather about half of our sample: K1 and K2. From the colour-magnitude diagram (viz. Fig. 1), it is clear that K1 and K2 have well defined bimodality. The Sersic_n_R indices of K1 and K2 lie between 2 - 4 which show that some of the galaxies are young and in the formative stages of their bulges and some of them have well developed bulges. The SFR and metallicities in these galaxies are intermediate between highest and lowest values. They differ largely in their light element abundances: K1 has larger value than K2. Regarding EW(Hδ_{abs}) the values lie between 0.5 and 2.5 with pronounced scatter. Groups K1 and K2 have globally intermediate values of SFR and more particularly of specSFR with respect to the other groups. Most of the parameters like metallicity, SFR, EW(Hδ_{abs}), specSFR, D4000_n and C have large scatters which indicate that K1 and K2 have a mixture of old and young populations of active and passive galaxies with well defined bulges for K1 and more pseudo bulges for K2. K1 group has higher velocity dispersion hence more massive galaxies as compared to K2. K1 group is more abundant in helium enriched population than K2, which is a signature of second generation stars. So K1 has slightly older population than K2. This is also evident from D4000_n peaks: K1 has a larger peak at higher D4000_n whereas the opposite is true for K2. Also C values show that the two groups have populations of late type spirals, but in much higher proportion in K2.

While one of the BPT diagram (Fig. 6) seem to show that these two groups are undoubtely AGNs, the other BPT diagrams (Fig. 7 and 8) seem to indicate they are more probably LINERS with an extension in the star forming galaxy region. They probably correspond to both the composite class and the ambiguous galaxies mentioned in Kewley et al. (2006). Hence, in the multivariate space, these galaxies show more subtle similarities than the sole ionization processes.

As a conclusion, K1 and K2 galaxies represent the bulk of galaxies in the Universe, with somewhat average values for most parameters, gathering both high and low mass galaxies, star forming and non-forming as well as intermediate objects.

Another interpretation might be that even with 49 parameters, it is not possible to distinguish sub-classes within these two big groups. This may be due to several reasons: lack of more distinctive information, uncertainties that smears out true differences, or to the fact that galaxies are big and complex systems so that integrated values mix up several peculiarities like AGN emission, multiple star forming or ionization regions etc.

5.2 Groups K4, K10

Groups K4 and K10 are globally similar with mostly average values. They are often close to groups K1 and K2. However, they have remarkable high C together with very high Sersic_r50_R and Sersic_r90_R. Hence they are very big and large galaxies, and thus spheroidals. They occupy the same region in the color-magnitude diagram, corresponding to the green valley. However they have a relatively old population as based on D4000_n, but with a small fraction of younger stellar populations and some star formation ongoing. These galaxies are thus clearly in the transitional phase of being quenched.

5.3 Groups K3, K5, K6

The three groups K3, K5, and K6 are rather specific: they have least specSFR, high velocity dispersion, high metallicity, high Sersic indices and C >> 2.6 for most of their members. All this indicates that they are early type spheroidal galaxies and quenched. They are more concentrated at the centre and massive in nature. There is a high abundance of oxygen (viz. Fig. B1, EW(OIII_4363), especially in K3) which might be due to the explosion of massive supernovae (Pop II objects). They have well developed bulges as seen from their highest Sersic_n_R values (~ 4, Fig. B1). The forbidden line are pronounced in K3, indicating that these galaxies have enough neutral gas. They might be formed by wet mergers of galaxies. K3, K5, K6 have lower EW(Hδ_{abs}) values with small scatters and high D4000_n values. The EW(Hδ_{abs}) values do not vary much in these groups. The BPT diagrams (Fig. 6) and C - parameter (Fig. 5) show that these groups have a significant population of ellipticals.

However the other two BPT diagrams on Fig. 7 and 8 seem to show a mixture of star forming galaxies and LINERS/Seyfert. These may correspond somehow to the composite galaxy class of Kewley et al. (2006), the difference being that here it is a composite class derived from the observations, not from theoretical models (see also Bamford et al. 2008; Zhang & Hao 2016). In addition, our multivariate analysis is better able to reveal some hidden properties in such composite objects.

We may also explain the extension of these groups on the BPT diagrams as galaxies seen at different stages of evolution that indeed belong to the same multivariate class (e.g. Fiorenza et al. 2014).

5.4 Groups K7, K8, K9

The three groups are active sites of star formation as supported by high EW(Hβ_{abs}), EW(Hδ_{abs}), EW(Hγ_{abs}), specSFR, low metallicity and high EW(OIII_5007) and in particular for the signature of a star burst in K8 other than AGN.

They are also the less massive ones of the sample, especially K8. They have higher values of EW(Hδ_{abs}) thus containing young stellar populations but the widths of its distribution are not similar in these groups. There is maximum scatter in K8 and minimum scatter in K7. They all have a high SFR, but the very small group K8 has a particularly high specSFR. The large scatter in EW(OIII_5007) in the latter groups indicates that these galaxies contain gas at high temperature to form new generation of stars. Hence K8 has stellar population which is a mixture of various ages as a result of star burst of recent origin.

Though K7 and K9 are dominated by late type galaxies, they also have small fractions of spheroidals. The galaxies in these groups have a low Sersic_n_R lying between 0 and

2 which indicates that the bulges in the spheroidals are not also very pronounced i.e., these galaxies may be in a formation stage.

The galaxies in the three groups K7, K8 and K9 are thus identified by our multivariate analysis as being in an active stage of their evolution, the only one in our large sample.

5.5 SFR and quenching

Contrarily to most if not all SFR studies, we do not assume or find only two categories of galaxies despite an apparent bimodality in many parameters. Instead, our ten groups do not simply split into the so-called red ant blue sequences or in the green valley, they often show some non negligible extensions. In addition, about half of our large sample belongs to two groups (K1 and K2) that span these three regions of the color-magnitude diagram.

The usual interpretation of the bimodality is that galaxies undergo a dramatic decrease of their SFR at some time, so moving from the blue sequence to the red one through the transitional green valley. This picture is certainly true on average, but ignores specificities of sub-populations of galaxies and different quenching mechanisms.

The apparent bimodality strongly suggests that the transitional phase is generally short. However, the quenching mechanisms are various and their timescales are rather different (e.g. Lian et al. 2016; Fossati et al. 2017).

It can be expected that the fast quenching mechanisms may be somewhat sufficiently violent so that the galaxies move quickly in the multivariate property space, and hence in the color-magnitude diagram as well. This may explain that groups K3, K5 and K6 are confined on the red sequence. Another possibility is that their quenching were not so fast but occurred a long time ago since they have well developed bulges (highest Sersic_n_R) and are the oldest group of our sample (highest $u - z$ and D4000_n).

Groups K4 and K10 are in the transitional phase, being also probably quickly quenched since they have a low SFR with little dispersion (Fig. 11), are easily identified in our multivariate analysis and are confined to the green valley.

Conversely, groups K1 and K2 contain star forming, quenched and transitional galaxies, which can be explained by slow quenching mechanisms, and/or rejuvenating processes creating new starbursts as seen in K1.

For the blue groups K7, K8 and K9 we cannot predict what kind of quenching mechanism will take place.

A more precise picture would require stellar population modelling and fitting, which will deserve a subsequent paper.

The regression analysis of SFR and specSFR fails for groups K7, K10 and also K8, while for K4 the concentration index C has a strong influence. Even if we have no clear explanation at this moment, it is interesting to note that these four groups are all star forming galaxies or in the transitional phase of quenching.

6 SUMMARY AND CONCLUSION

In the present work we have classified a large data set of galaxies with a large number of morphological, photometric and spectroscopic parameters compiled from VAGC/SDSS data archive. We have used two very sophisticated statistical methods e.g. ICA and k-means cluster analysis for finding coherent groups and subsequently used another sophisticated method, Stepwise Multiple Regression to predict the SFR and specSFR in each group as a function of significant galaxy parameters. Unlike other studies we have chosen ICA which is appropriate for a non-Gaussian data set like the present one. We have found ten coherent groups with the following features.

- **K1:** populations of young and old objects, early-type in nature with a small fraction of late-type ones. Some of its galaxies are in formative stages of their bulges. Probably slow quenching mechanisms and rejuvenating processes.
- **K2:** mixture populations of late- (majority) and early-type, young and old, galaxies, some of them are in formative stages of their bulges. Slow quenching mechanisms.
- **K3:** old populations of early type galaxies, might have been formed by wet mergers. Possibly fast quenching mechanisms.
- **K4:** massive elliptical galaxies, mixture populations of young and old objects, early type in nature and some of them are in formative stages of their bulges. In the transitional phase of quenching.
- **K5:** old populations of early type galaxies and/or fast quenching mechanisms.
- **K6:** old populations of early type galaxies and/or fast quenching mechanisms.
- **K7:** mostly starburst young galaxies.
- **K8:** mostly starburst young galaxies, K8 consists of mixture populations of various ages compared to semi passive and passive groups of galaxies K7 and K9 respectively. Includes some true AGNs.
- **K9:** mostly starburst young galaxies.
- **K10:** massive elliptical galaxies, mixture populations of young and old objects, and some of them are in formative stages of their bulges. In the transitional phase of quenching.

The most important aspect of a multivariate technique is to explain (i) the significance of mixture populations e.g. groups K1, K2, K4 and K10, as well as the several groups in the blue (viz. K7, K8, K9) or red (viz. K3, K5, K6) sequences, (ii) the implications of the mixture groups. There can be two aspects; (i) one from their compositional aspect and (ii) the other from their evolutionary aspect. In the former case we can point out in respect e.g. of BPT diagram (viz. Fig. 8) that the AGNs in the group K8 have higher [OIII]/Hβ values (viz. ~ 1.0) as compared to the AGNs the other groups (K1, K2, K3, K4, K5, K6 and K10). On the other hand the trend is opposite for the starburst galaxies i.e. these are continuously decreasing from blue to red via green valley.

In the latter case in respect of the Fig. 11 it is seen that the number of massive galaxies is generally increasing from the blue to the red end along with a decreasing trend in star formation rates.

From the above conjecture we can immediately reach at the conclusion that galaxies are not classified in two distinct groups, blue or red but there is a rather continuous process in the scenario of star formation: galaxies which are starburst galaxies become active and massive by minor and/or major mergers and gradually become passive passing into the red sequence, galaxies harboring AGNs and initially having stronger activities become weaker passing to the red sequence, the green valley actually shows the transition trajectory of this evolution. Using many parameters as in our analysis allows to distinguish galaxies according to the quenching mechanisms of stellar formation.

The multivariate analysis thus unveils this hidden picture automatically, which is not possible to explore otherwise like through scatter diagrams only. The detailed study of this aspect for giving a complete scenario of galaxy formation and subsequent evolution using such huge data base will be carried out in a subsequent paper.

ACKNOWLEDGEMENTS

We thanks Fabrice Lamareille and Emmanuel Davoust for having compiled for us the sample used in this analysis. T.K and DFB are thankful to Indo-French Centre for Applied Mathematics (Bangalore, India) for providing partial support.

Funding for the Sloan Digital Sky Survey (SDSS) has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the U.S. Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS Web site is http://www.sdss.org/.

The SDSS is managed by the Astrophysical Research Consortium (ARC) for the Participating Institutions. The Participating Institutions are The University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, University of Pittsburgh, Princeton University, the United States Naval Observatory, and the University of Washington.

REFERENCES

Abazajian K. N., et al., 2009, ApJS, 182, 543

Albazzaz H., Wang X. Z., 2004, Industrial & engineering chemistry research, 43, 6731

Babu G. J., Chattopadhyay T., Chattopadhyay A. K., Mondal S., 2009, ApJ, 700, 1768

Baldwin J. A., Phillips M. M., Terlevich R., 1981, PASP, 93, 5

Bamford S. P., Rojas A. L., Nichol R. C., Miller C. J., Wasserman L., Genovese C. R., Freeman P. E., 2008, MNRAS, 391, 607

Belfiore F., et al., 2016, MNRAS, 461, 3111

Blanton M. R., et al., 2005, The Astronomical Journal, 129, 2562

Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, Monthly Notices of the Royal Astronomical Society, 351, 1151

Brosche P., 1973, A&A, 23, 259

Cabanac R. A., de Lapparent V., Hickson P., 2002, Astronomy and Astrophysics, 389, 1090

Chattopadhyay T., Chattopadhyay A., 2006, The Astronomical Journal, 131, 2452âĂŞ2468

Chattopadhyay T., Chattopadhyay A., 2007, Astronomy & Astrophysics, 472, 131

Chattopadhyay T., Misra R., Naskar M., Chattopadhyay A., 2007, Astrophysical Journal, 667, 1017

Chattopadhyay T., Mondal S., Chattopadhyay A., 2008, Astrophysical Journal, 683, 172

Chattopadhyay T., Babu J., Chattopadhyay A., Mondal S., 2009a, Astrophysical Journal, 700, 1768

Chattopadhyay A., Chattopadhyay T., Davoust E., Mondal S., Sharina M., 2009b, The Astrophysical Journal, 705, 1533

Chattopadhyay T., Sharina M., Karmakar P., 2010, ApJ, 724, 678

Chattopadhyay T., Sharina M., Davoust E., De T., Chattopadhyay A. K., 2012, ApJ, 750, 91

Chattopadhyay A. K., Chattopadhyay T., De T., Mondal S., 2013, Astrostatistical Challenges for the New Astronomy. Springer-Verlag New York, pp 185–202, doi:10.1007/978-1-4614-3508-2

Cid Fernandes R., Stasińska G., Schlickmann M. S., Mateus A., Vale Asari N., Schoenell W., Sodré L., 2010, MNRAS, 403, 1036

Comon P., 1994, Signal Processing, 36, 287

Das S., Chattopadhayay T., Davoust E., 2015, Publ. Astron. Soc. Australia, 32, e041

De T., Fraix-Burnet D., Chattopadhyay A. K., 2016, Communication in Statistics - Theory and Methods, 45, 2638

Ellis S. C., Driver S. P., Allen P. D., Liske J.b Bland-Hawthorn J., De Propris R., 2005, Monthly Notices of the Royal Astronomical Society, 363, 1257

Fiorenza S. L., Takeuchi T. T., Małek K. E., Liu C. T., 2014, ApJ, 784, 140

Fossati M., et al., 2017, The Astrophysical Journal, 835, 153

Fraix-Burnet D., Dugué M., Chattopadhyay T., Chattopadhyay A. K., Davoust E., 2010, MNRAS, 407, 2207

Fraix-Burnet D., Chattopadhyay T., Chattopadhyay A. K., Davoust E., Thuillard M., 2012, Astronomy and Astrophysics, 545, A80

Fraix-Burnet D., Thuillard M., Chattopadhyay A. K., 2015, Frontiers in Astronomy and Space Sciences, 2

Haines C. P., et al., 2016, preprint (arXiv:1611.07050)

Kauffmann G., et al., 2003, MNRAS, 346, 1055

Kennicutt R. C., 1998, Annual Review of Astronomy and Astrophysics, 36, 189

Kennicutt R. C., Evans N. J., 2012, Annual Review of Astronomy and Astrophysics, 50, 531

Kewley L. J., Dopita M. A., Sutherland R. S., Heisler C. A., Trevena J., 2001, ApJ, 556, 121

Kewley L. J., Groves B., Kauffmann G., Heckman T., 2006, MNRAS, 372, 961

Kewley L. J., Dopita M. A., Leitherer C., Davé R., Yuan T., Allen M., Groves B., Sutherland R., 2013, ApJ, 774, 100

Lamareille F., 2010, A&A, 509, A53

Lian J., Yan R., Zhang K., Kong X., 2016, The Astrophysical Journal, 832, 29

MacQueen J. B., 1967, in Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. pp 281–297, `http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/kmeans`

Murtagh F., Heck A., eds, 1987, Multivariate Data Analysis Astrophysics and Space Science Library Vol. 131, doi:10.1007/978-94-009-3789-5.

Osterbrock D. E., 1989, Astrophysics of gaseous nebulae and active galactic nuclei

Padmanabhan N., et al., 2008, ApJ, 674, 1217

Peth M. A., et al., 2015, preprint (arXiv:1504.01751)

Prochaska L. C., Rose J. A., Caldwell N., Castilho B. V., Concannon K., Harding P., Morrison H., Schiavon R. P., 2007, AJ, 134, 401

Renzini A., Peng Y.-j., 2015, ApJ, 801, L29

Richardson C. T., Allen J. T., Baldwin J. A., Hewett P. C., Ferland G. J., Crider A., Meskhidze H., 2016, MNRAS, 458, 988

Sánchez Almeida J., Aguerri J. A. L., Muñoz-Tuñón C., de Vicente A., 2010, ApJ, 714, 487

Stasińska G., Cid Fernandes R., Mateus A., Sodré L., Asari N. V., 2006, MNRAS, 371, 972

Stensbo-Smidt K., Gieseke F., Igel C., Zirm A., Pedersen K. S., 2016, Monthly Notices of the Royal Astronomical Society, 464, 2577

Strateva I., et al., 2001, AJ, 122, 1861

Sugar C. A., James G. M., 2003, Journal of the American Statistical Association, 98, 750

Thomas D., Maraston C., Johansson J., 2011, MNRAS, 412, 2183

Trager S. C., Faber S. M., Worthey G., González J. J., 2000, AJ, 120, 165

Veilleux S., Osterbrock D. E., 1987, ApJS, 63, 295

Watanabe M., Kodaira K., Okamura S., 1985, The Astrophysical Journal, 292, 72

Whitmore B. C., 1984, Astrophysical Journal, 278, 61

Worthey G., Ottaviani D. L., 1997, ApJS, 111, 377

Zhang K., Hao L., 2016, preprint (`arXiv:1610.03495`)

de Souza R. S., et al., 2017, preprint (`arXiv:1703.07607`)

APPENDIX A: LIST OF PARAMETERS

Table A1. All the parameters of the data set used for the analysis. See http://wwwmpa.mpa-garching.mpg.de/SDSS/DR7/SDSS_line.html for more details.

Parameter	Description
v_{disp}	estimated velocity dispersion from spectrum
U	u absolute magnitude (log of intensity)
J	J absolute magnitude
Sersic_amp_R	The best fit to the variable “A” in band R (nanomaggies/$arcsec^2$): describes the radial distribution of light
Sersic_n_R	The best fit to the Sersic index “n” in band R
sigma_balmer	Velocity dispersion (σ not FWHM) measured simultaneously in all of the Balmer lines in km/s
sigma_forb	Velocity dispersion (σ not FWHM) measured simultaneously in all the forbidden lines in km/s
oii_3729_seqw	The equivalent width of the continuum-subtracted emission line with the other emission lines subtracted off
neiii_3869_seqw	”
oiii_4363_seqw (EW(OIII_4363))	”
oiii_5007_seqw (EW(OIII_5007))	”
hei_5876_seqw	”
oi_6300_seqw	”
h_alpha_seqw (EW(Hα))	”
nii_6584_seqw	”
sii_6731_seqw	”
EW(Hδ_{abs})	Equivalent width in the absorption line
EW(Hγ_{abs})	”
EW(Hβ_{abs})	”
EW(Hα_{abs})	”
Ca_K_abs	”
ca_h_abs	”
na_d_abs	”
Lick_CN2	stellar absorption line (Lick) index
Lick_Ca4227	”
Lick_G4300	”
Lick_Fe4383	”
Ca4455	”
Lick_Fe4531	”
Lick_c4668	”
Lick_Hb	”
Lick_Fe5015	”
Lick_Mgb	”
Lick_Fe5270	”
Lick_Fe5335	”
Lick_Fe5406	”
Lick_Fe5709	”
Lick_Fe5782	”
Lick_NaD	”
Lick_Hd_A	”
D4000_n	The break in the spectrum at 4000 Å
SFR	Star Formation Rate
specSFR	specific Star Formation Rate
Sersic_r0_R	The best fit to the variable r_0 in R band (arcsec)
Sersic_r50_R	50% light radius of best fit model in R band (arcsec)
Sersic_r90_R	90% light radius of best fit model in R band (arcsec)
C	Concentration Index: ratio between Sersic_r90_R and Sersic_r50_R
$U - z$	magnitude u minus magnitude z (u-z)
$J - H$	
$H - K$	

APPENDIX B: BOXPLOTS

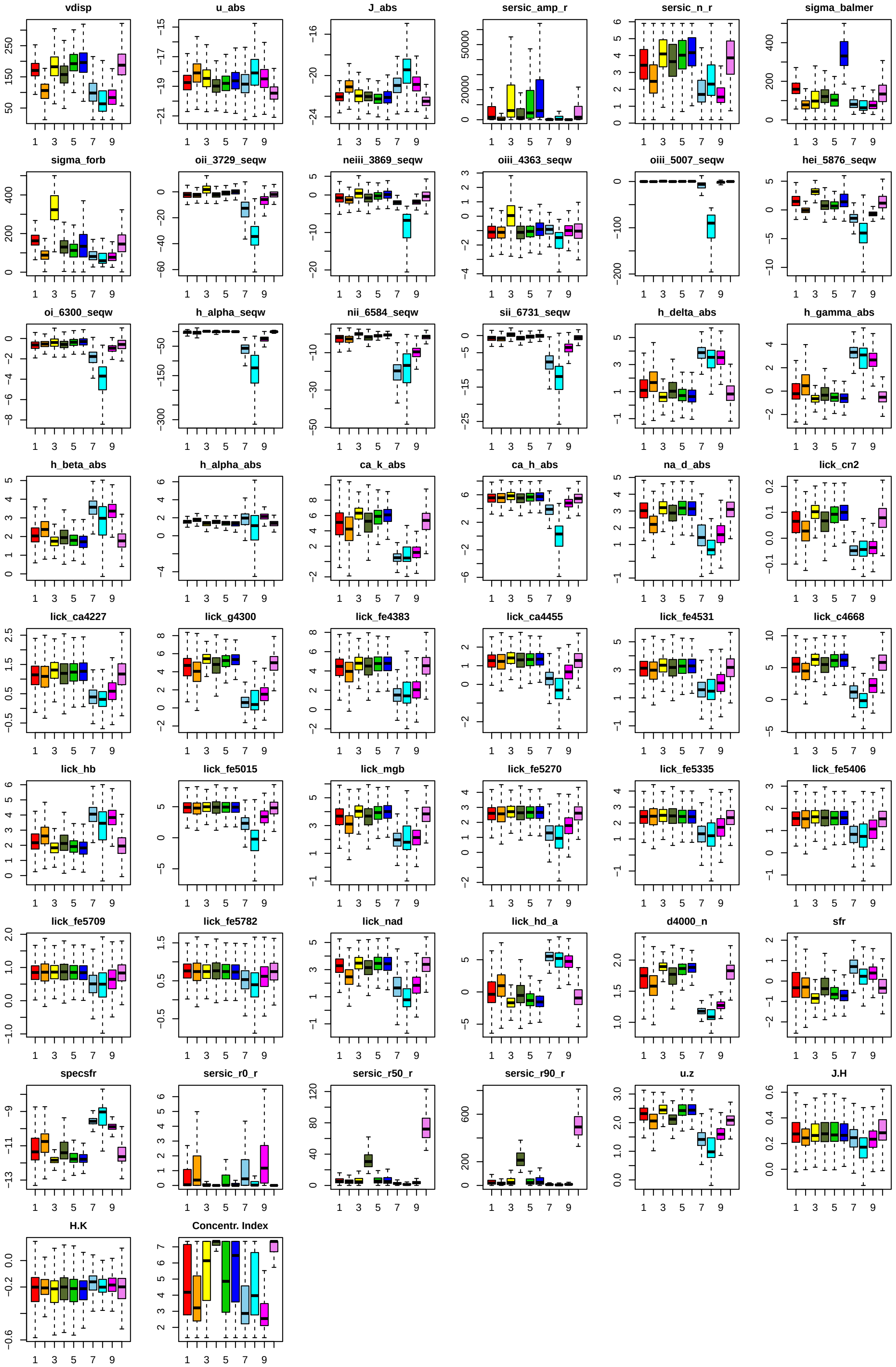

Figure B1. Boxplots for the parameters within the classified groups K1 to K10.

APPENDIX C: TABLES OF COEFFICIENTS FOR THE STEP WISE REGRESSION OF SFR AND SPECSFR

Table C1. Coefficients of the step wise regression of SFR. Boldface fonts indicate absolute values higher than 0.05.

	All	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10
U	**-0.16**	**-0.20**	**-0.22**	**-0.24**	**-0.16**	**-0.19**	**-0.14**	**-0.22**	**-0.27**	**-0.30**	**-0.27**
J	**-0.27**	**-0.20**	**-0.23**	**-0.10**	**-0.24**	**-0.19**	**-0.18**	**-0.15**	**-0.11**	**-0.11**	**-0.10**
EW(Hα)	-0.00	-0.05	-0.01	-0.05	-0.02	-0.03	**-0.07**	-0.00	0.00	-0.01	-0.02
EW(Hα_{abs})	0.02	**0.06**	0.01	**0.07**	**0.06**	**0.06**	**0.11**	-0.00	0.00	-0.01	**0.07**
Ca_K_abs	-0.01	-0.05	-0.01	-0.01	-0.02	-0.03	-0.02	0.00	-0.01	-0.01	-0.02
NaD_abs	**0.08**	0.02	**0.06**	0.00	**0.11**	0.03	0.00	**0.13**	0.03	**0.06**	**0.07**
Lick_CN2	-0.02	-0.03	-0.01	0.00	**-0.06**	0.00	0.00	0.04	0.00	0.00	0.00
D4000_n	**-1.59**	**-0.68**	**-1.44**	**-0.98**	**-1.61**	**-1.37**	**-0.97**	0.00	**-1.09**	0.00	**-1.54**
C	**-0.06**	0.05	**0.06**	-0.03	**0.79**	0.00	-0.05	0.00	0.00	-0.04	0.00
$U - z$	0.00	0.05	0.01	**0.17**	**-0.10**	**0.11**	**0.08**	**0.18**	**0.20**	**0.26**	**0.09**
$J - H$	**0.23**	**0.19**	**0.19**	**0.13**	**0.21**	**0.18**	**0.18**	**0.09**	**0.13**	**0.09**	**0.09**
$H - K$	**0.12**	**0.11**	**0.08**	**0.07**	**0.10**	**0.11**	**0.08**	**0.10**	**0.08**	**0.07**	0.05
EW(OIII_4363)	**-0.06**	-0.04	-0.04	-0.01	-0.04	-0.04	-0.02	-0.02	-0.03	-0.01	-0.02
EW(HeI_5876)	-0.04	-0.03	-0.04	0.00	-0.06	-0.02	0.02	-0.03	-0.04	0.01	-0.04

Table C2. Coefficients of the step wise regression of specSFR. Boldface fonts indicate absolute values higher than 0.05.

	All	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10
U	0.00	0.05	**-0.08**	**0.06**	**0.08**	**0.06**	**0.14**	**-0.09**	**-0.13**	**-0.15**	**0.07**
J	-0.03	**-0.06**	0.02	0.00	**-0.09**	-0.04	**-0.06**	**0.09**	**0.15**	**0.12**	-0.04
EW(Hα)	-0.00	-0.05	-0.02	-0.05	-0.02	-0.03	**-0.06**	-0.00	-0.00	-0.01	-0.03
EW(Hα_{abs})	0.02	**0.06**	0.02	**0.07**	0.05	**0.06**	**0.10**	-0.00	0.00	-0.01	**0.07**
Ca_K_abs	-0.01	-0.05	-0.01	-0.01	-0.02	-0.03	-0.02	-0.00	-0.02	-0.01	-0.02
NaD_abs	0.02	-0.03	0.00	-0.05	0.03	-0.01	-0.04	**0.07**	0.00	0.02	0.01
D4000_n	**-1.60**	**-0.69**	**-1.49**	**-1.04**	**-1.61**	**-1.40**	**-1.03**	**-0.57**	**-1.56**	**-0.31**	**-1.54**
C	0.01	0.00	0.04	0.00	**0.48**	0.00	0.00	0.05	0.00	0.00	0.00
$U - z$	**-0.20**	**-0.22**	**-0.18**	**-0.14**	**-0.39**	**-0.17**	**-0.22**	-0.03	0.00	-0.02	**-0.28**
$J - H$	**0.08**	**0.09**	0.03	**0.06**	**0.12**	**0.09**	**0.09**	-0.02	0.00	-0.03	0.05
$H - K$	**0.06**	**0.07**	0.02	0.04	**0.07**	**0.07**	0.04	**0.06**	0.00	0.02	0.04

This paper has been typeset from a TEX/LATEX file prepared by the author.