MNRAS **000**, 1–6 (2016) Preprint 13 January 2016 Compiled using MNRAS LATEX style file v3.0

Insert clever, witty title here...

Ian D. Roberts,★ Laura C. Parker
Department of Physics and Astronomy, McMaster University, Hamilton ON L8S 4M1, Canada

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Key words: galaxies: clusters: general – galaxies: evolution – galaxies: groups: – galaxies: statistics

1 INTRODUCTION

2 DATA

2.1 Group sample

For this work we employ the group catalogue of Yang et al. (2007), which is constructed by applying the halo-based galaxy group finder from Yang et al. (2005, 2007) to the New York University Value-Added Galaxy Catalogue (NYU-VAGC; Blanton et al. 2005). The NYU-VAGC is a low redshift galaxy catalogue (primarily $z \lesssim 0.3$) consisting of 693319 galaxies derived from the Sloan Digital Sky Survey Data Release 7 (SDSS-DR7; Abazajian et al. 2009). We will briefly describe the halo-based group finding algorithm used to generate the Yang group catalogue, however for a more complete description the authors direct the reader to Yang et al. (2005) and Yang et al. (2007).

First, the centres of potential groups are identified. Galaxies are initially assigned to groups using a traditional "friends-of-friends" (FOF) algorithm (e.g. Huchra & Geller 1982) with very small linking lengths. The luminosity-weighted centres FOF groups with at least two members are then taken as the centres of potential groups and all galaxies not yet associated with a FOF group are treated as tentative centres for potential groups. A characteristic luminosity, $L_{19.5}$, defined as the combined luminosity of all group members with ${}^{0.1}M_r - 5\log h \leq -19.5$, is calculated for each tentative group and an initial halo mass is assigned using an assumption for the group mass-to-light ratio, $M_H/L_{19.5}$. Utilizing this tentative group halo mass, velocity dispersions and a virial radius are calculated for each group. Next, galaxies are assigned to groups under the assumption that the distribution of galaxies in phase space follows that of dark matter particles – the distribution of dark matter particles is taken to follow a spherical NFW profile (Navarro et al. 1997). Using the new group memberships, group centres are recalculated and the procedure is iterated until group memberships no longer change.

We take group halo masses, M_H, from the Yang catalogue calculated using a characteristic group stellar mass, $M_{\star,\mathrm{grp}}$, and assuming that there is a one-to-one relation between $M_{\star,\mathrm{grp}}$ and M_H. Yang et al. (2007) define $M_{\star,\mathrm{grp}}$ as

$$M_{\star,\mathrm{grp}} = \frac{1}{g(L_{19.5}, L_{\mathrm{lim}})} \sum_i \frac{M_{\star,i}}{C_i} \tag{1}$$

where $M_{\star,i}$ is the stellar mass of the ith member galaxy, C_i is the completeness of the survey at the position of that galaxy, and $g(L_{19.5}, L_{\mathrm{lim}})$ is a correction factor which accounts for galaxies missed due to the magnitude limit of the survey.

The Yang catalogue contains both haloes which would be broadly classified as groups ($10^{12} \lesssim M_H \lesssim 10^{14}\,\mathrm{M}_\odot$) as well as clusters ($M_H \gtrsim 10^{14}\,\mathrm{M}_\odot$), however for brevity we will refer to all haloes as groups regardless of halo mass unless otherwise specified.

We calculate group-centric radii for all group members within the sample using the redshift of the group angular separation of the galaxy from the luminosity-weighted centre of the host halo. Radii are all normalized by the virial radius, R_{200}, of the group using the definition given in Carlberg et al. (1997)

$$R_{200} = \frac{\sqrt{3}\sigma}{10H(z)}, \tag{2}$$

where the Hubble parameter, H(z), is defined as

$$H(z) = H_0\sqrt{\Omega_{\mathrm{M}}(1+z)^3 + \Omega_\Lambda}, \tag{3}$$

and the velocity dispersion, σ, is calculated using a fitting function given in Yang et al. (2007)

$$\sigma = 397.9\,\mathrm{km\,s^{-1}} \left(\frac{M_{\mathrm{halo}}}{10^{14}\,h^{-1}\,\mathrm{M}_\odot}\right)^{0.3214}. \tag{4}$$

To study specific characteristics of galaxies within the group sample, we match various public SDSS galaxy catalogues to the group sample. We utilize galaxy stellar masses given in the NYU-VAGC, which are obtained through fits

★ E-mail: roberid@mcmaster.ca

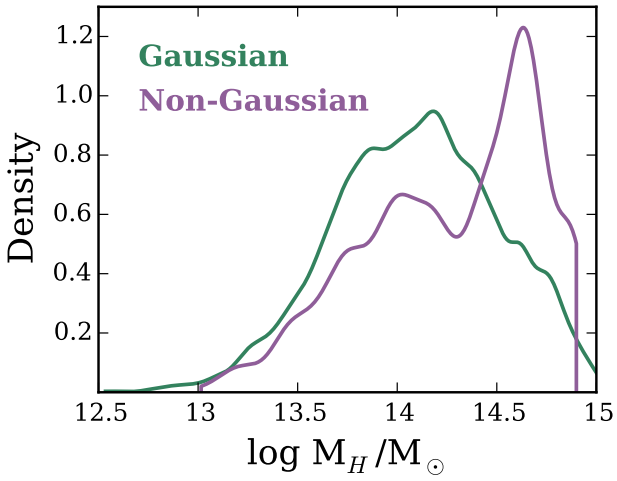

Figure 1. Smoothed host halo mass distributions for galaxies in the unmatched G and NG samples.

to galaxy spectra and broadband photometric measurements following the procedure of Blanton & Roweis (2007).

For our star formation indicator we use specific star formation rates ($SSFR = SFR/M_\star$) from Brinchmann et al. (2004). These SSFRs are primarily derived from emission lines, with an exception for galaxies with no clear emission lines or AGN contamination in which case SSFRs are based on the 4000 Å break. SSFRs for galaxies with S/N > 2 in Hα are determined used only the Hα line and SSFRs for galaxies with S/N > 3 in all four BPT lines are determined using a combination of emission lines.

For our morphology indicator we use a global Sérsic index taken from the single component Sérsic fits in Simard et al. (2011). We also weight all of the data by $1/V_{\rm max}$ as given in Simard et al. (2011) to account for the incompleteness of the sample.

2.2 Field sample

For this research we also define a sample of "field" galaxies. Like the group sample, the field sample is also derived from the NYU-VAGC. In order to construct the field sample we cross-match galaxies within the Yang group catalogue against all galaxies within the NYU-VAGC catalogue, and remove any galaxies which have been identified as being members of Yang groups. Furthermore, we apply an isolation criteria and only keep galaxies which are separated from their nearest-neighbour by a projected distance of at least 0.5 Mpc and by at least 500 km/s in line-of-sight velocity.

The same galaxy parameters which are matched to the group sample are also matched to the field sample.

2.3 Group dynamics

To classify the dynamical state of the haloes in the data set we use a combination of two statistical tests, the Anderson-Darling (AD) normality test (Anderson & Darling 1952; see Hou et al. 2009, 2013 for an astronomical application) and the Dip test (Hartigan & Hartigan 1985; see Ribeiro et al. 2013 for an astronomical application). The AD test is a non-parametric test of normality based upon the comparison between the cumulative distribution function (CDF) of a measured data sample and the CDF of a gaussian distribution. Under the assumption that the data is in fact normally distributed, the AD test determines the probability (p) that the difference between the CDFs of the data and a normal distribution equals or exceeds the observed difference. We apply the AD test to the velocity distributions of the member galaxies of each group in the data sample, thereby broadly classifying the dynamical state of each halo. Our first criteria in classifying a group as G is that the p-value given by the AD test be greater than or equal to 0.05. Our second criteria required for a group to be classified as G is that it be unimodal. To specifically gauge the modality of the velocity distribution of a given group we use the Dip test. Like the AD test, the Dip test is also a non-parametric CDF statistic. Where they differ is that the Dip test looks for a flattening of the CDF for the data which would correspond to a 'dip' in the distribution being tested. The Dip test operates under the null hypothesis that the data is unimodal, and we consider a group velocity distribution unimodal if the Dip test p-value is greater than or equal to 0.05. Therefore our G data sample consists of all those groups with $p_{\rm ad} \geq 0.05$ *and* $p_{\rm dip} \geq 0.05$, whereas our NG data sample consists of all those groups with $p_{\rm ad} < 0.05$ *or* $p_{\rm dip} < 0.05$.

After applying the above criteria we find a G sample consisting of 42655 galaxies within 2447 groups and a NG sample consisting of 5306 galaxies within 215 groups. The authors note that simply applying these normality criteria in this fashion can lead to the NG sample being biased toward rich, high halo mass groups (see Fig. 1). To address this we match of G and NG samples by halo mass (as well as stellar mass and redshift), this matching procedure is laid out in the next section.

2.4 Matched data set

To ensure a fair comparison between galaxies in different environments (ie. field galaxies, galaxies in G groups, and galaxies in NG groups) we match our sample of G group galaxies and NG group galaxies by stellar mass, redshift, and halo mass. Additionally, we then match our sample of field galaxies by stellar mass and redshift ensuring that all of our galaxy samples are matched according to important galaxy properties. This is especially important when trying to elucidate information on the effect of group dynamics on galaxy SF and morphological properties for two main reasons:

First, stellar mass, redshift, and halo mass have all been shown to influence galaxy SF and morphology (e.g. Brinchmann et al. 2004; Feulner et al. 2005; Zheng et al. 2007; Cucciati et al. 2012; Wetzel et al. 2012; Lackner & Gunn 2013; Tasca et al. 2014); whereas the impact of group dynamics is less clear (Hou et al. 2013; Ribeiro et al. 2013) (REF) which is perhaps suggestive of a more modest role. Therefore, if one hopes to identify trends in galaxy SF and morphology with group dynamics it is crucial to properly control for these other effects.

Second, standard statistical normality tests, such as the AD test, are inherently biased in identifying non-Gaussian distributions when sample size is large. This is a result of

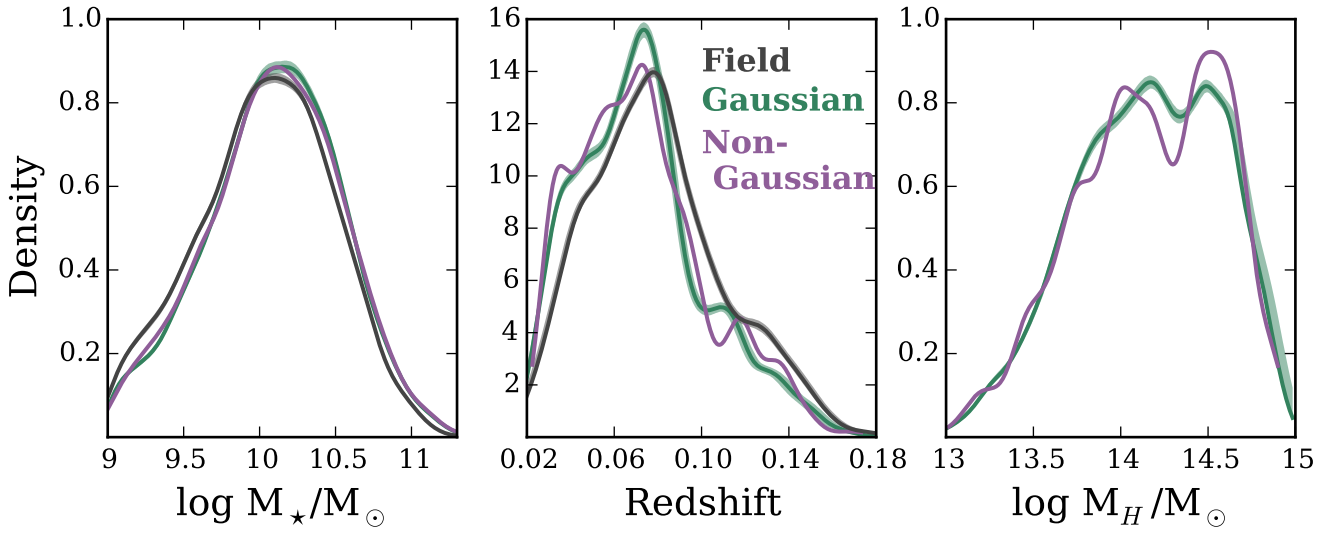

Figure 2. Smoothed distributions for stellar mass, redshift, and host halo mass for galaxies in the matched G, NG and field (where applicable) samples. Shaded regions around the G and field lines are 90 per cent confidence intervals corresponding to the stochastic nature of our matching procedure.

the statistical power of the test increasing with sample size which subsequently allows the detection of more and more subtle departures from normality. These subtle departures from normality may not be physically relevant (in principle, no group is truly Gaussian anyways) and what really matters is whether galaxies in groups which show large departures from normality have different properties than galaxies in groups which show smaller departures from normality. Since group richness generally scales with halo mass, in the absence of any matching procedure, a sample of NG groups will be biased towards large halo masses compared to a similar sample of G groups – even though many high halo mass NG groups may have been identified on the basis of very small departures from normality. Ensuring that our G and NG samples have very similar halo mass distributions allows us to make a fairer comparison between the two samples.

Our algorithm for matching the G and NG samples is as follows:

1. Our list of galaxies found in NG groups is iterated through, for each galaxy one 'matching' galaxy from the G sample is found. To be considered matching the two galaxies must have stellar masses within 0.1 dex, redshifts within 0.01, and halo masses within 0.1 dex.
2. Step 1 is repeated continually until no more matches are found, the end result is a list of galaxies from the NG sample each of which will have one or more matching galaxies from the G sample assigned to them
3. The matched G sample is generated by including two galaxies from the G sample for every one matching galaxy from the NG sample. By definition this excludes any galaxies in the NG sample which only have one identified match. However, 85 per cent of galaxies in the NG sample have two or more matches so although we reduce the NG sample size by 15 per cent it allows us to increase the matched G sample size twofold.
4. In the case where a given galaxy in the NG sample has more than two identified matches, the two matching galaxies from the G sample are chosen randomly. This introduces a stochastic nature to our analysis as each generation of the matched G sample will not contain exactly the same galaxies (although in each generation the G and NG samples will indeed be matched). To account for this, any quantities calculated using the matched G sample are done so in a Monte Carlo sense where the median of 1000 stochastic generations is quoted along with 90 per cent confidence intervals.

The field sample is subsequently matched to the NG sample following the same procedure, the same method is used to account for the stochastic nature of the matching procedure. Fig. 2 shows smoothed density distributions of stellar mass, redshift, and halo mass for the matched G, NG, and field samples. Please note that for the remainder of the paper all analysis is done using the matched samples, therefore from this point forward any reference to the G, NG, or field samples is implicitly referencing the matched samples.

3 IDENTIFYING INFALLING AND VIRIALIZED GALAXIES

Galaxies within group haloes can be broadly classified into three main subclasses: galaxies infalling to the group at large radii, galaxies virialized within the inner regions of the halo, and galaxies backsplashing beyond the virial radius after making a passage through the group centre. In order to understand the radial dependence of galaxy properties within groups **(REF)** it is crucial to be able to identify these different galaxy populations (Gill et al. 2005; Mahajan et al. 2011; Pimbblet 2011).

The main focus of this paper is the comparison between the three galaxy samples (field, G, NG) and to elucidate how the relationship between these three samples evolves with the infall state of member galaxies. In particular we

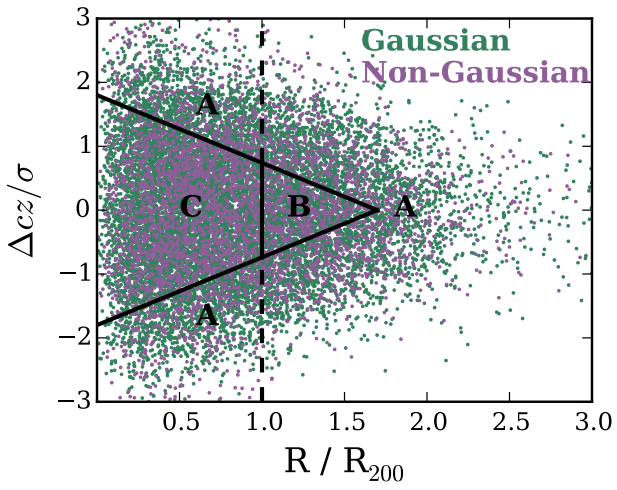

Figure 3. Projected radial phase space for galaxies within the G and NG samples. Virialized (C), backsplash (B), and infalling (A) regions from Mahajan et al. (2011) are shown.

will compare star formation and morphological properties for galaxies which are infalling into G and NG groups to the same properties for galaxies which are virialized within G and NG groups.

One method used to distinguish between infalling, virialized, and backsplash populations is to look for distinct populations in radial phase space. In particular Mahajan et al. (2011) follow Sanchis et al. (2004) and identify galaxies within one virial radius as virialized and use the cut

$$\frac{v_r}{V_v} = -1.8 + 1.06\left(\frac{r}{R_v}\right) \quad (5)$$

to distinguish between infalling and backsplash galaxies. Where v_r is the radial velocity of the galaxy, V_v is the velocity dispersion of the group, r is the group-centric radius, and R_v is the virial radius of the group.

The cuts described above were determined using full 6-d phase space information, however observationally we are limited to line-of-sight velocities and projected radii. Although working in projection removes much of the distinct phase space structure (Oman et al. 2013), density contours for the virialized, backsplash, and infalling populations still occupy similar regions in projected phase space with the contamination between different populations being more substantial (Mahajan et al. 2011). While the divisions between populations is certainly less clear in projection, equation 5 can still be used to obtain an approximate division between infalling, virialized, and backsplash galaxies – this approximation is preferred over the incorrect assumption that all galaxies beyond the virial radius are infalling for the first time.

To make the transformation to observational quantities we replace v_r/V_v with $\Delta cz/\sigma$ and r/R_v with R/R_{200} in equation 5. We also symmetrize the phase space cuts to account for projection by using the mirror of equation 5. After implementing these observational adjustments, and utilizing the best-fitting scheme from Mahajan et al. (2011), we are able to plot in Fig. 3 the phase space distribution of galaxies within the G and NG samples divided into the infalling (Regions A), backsplash (Region B), and virialized (Region C) populations.

4 GALAXY PROPERTIES IN THE INFALL REGION

We first consider the star forming and morphological properties of galaxies infalling (Regions A in Fig. 3) into both G and NG groups, as well as galaxies within the field sample. In Fig. 4 we plot star forming and disc fraction versus stellar mass for the three different galaxy samples. We define star forming galaxies to be all galaxies with $\log SSFR \geq -11$, Wetzel et al. (2012) show that in the local Universe the division between the red sequence and the blue cloud is consistently found at $\log SSFR \simeq -11$ across a wide range of halo masses. Likewise we define disc galaxies to be all galaxies having a global Sérsic index of $n \leq 1.5$. While the distribution of Sérsic index is not as clearly bimodal as the SSFR distribution, we find that our observed trends are insensitive to our choice of dividing Sérsic index.

Inspection of Fig. 4 reveals certain interesting features. First off, there is no observed difference between the star forming or disc fractions for galaxies infalling into Gaussian groups compared to galaxies infalling into non-Gaussian groups. This suggests that any influence that the dynamical state of the group has on star forming or morphological properties is not in place during a galaxies first infall toward the virialized region of the halo.

We also see a small excess of star forming in the field compared to galaxies falling into haloes (although this excess is interestingly absent in the lowest mass bin). When looking as disc fraction, this "field excess" is very marginal and arguably not present at all. Previous studies (Lewis et al. 2002; Gray et al. 2004; Rines et al. 2005; Verdugo et al. 2008) have found that star formation of galaxies within infall regions remains suppressed compared to the field out to radii on the order of $2-3\,R_{200}$. This suppression is generally attributed to backsplash galaxies which have already made a passage through the halo centre, the pre-processing of galaxies in small groups prior to infall, or some combination of the two. Using the cuts described in Section 4 we have attempted to "clean" our infall sample of backsplashing galaxies, although there is still almost certainly some level of contamination.

It is expected that pre-processing should play a larger role in large clusters compared to smaller groups, as a larger fraction of galaxies infalling into clusters will have been a part of a group at some time prior to infall. In particular, Hou et al. (2014) find that for groups with halo masses $M_H < 10^{14}\,M_\odot$ less than 5 per cent of infalling galaxies are a part of a subhalo, for clusters with halo masses in the range $10^{14} < M_H < 10^{14.5}\,M_\odot$ approximately 10 per cent of infalling galaxies are within subhaloes, and for the largest clusters ($M_H > 10^{14.5}\,M_\odot$) the fraction of infalling galaxies in subhaloes is $\sim$ 25 per cent. Using semi-analytic models, McGee et al. (2009) find that the fraction of galaxies which were accreted into the final cluster as a member of a halo with $M_H > 10^{13}\,h^{-1}\,M_\odot$ depends strongly on the mass of the ultimate cluster, ranging from ~ 0.1 for $10^{13.5}\,h^{-1}\,M_\odot$ haloes to ~ 0.45 for haloes with masses of $10^{15}\,h^{-1}\,M_\odot$.

We look for evidence of pre-processing by examining the "field excess", which we define as the difference in star

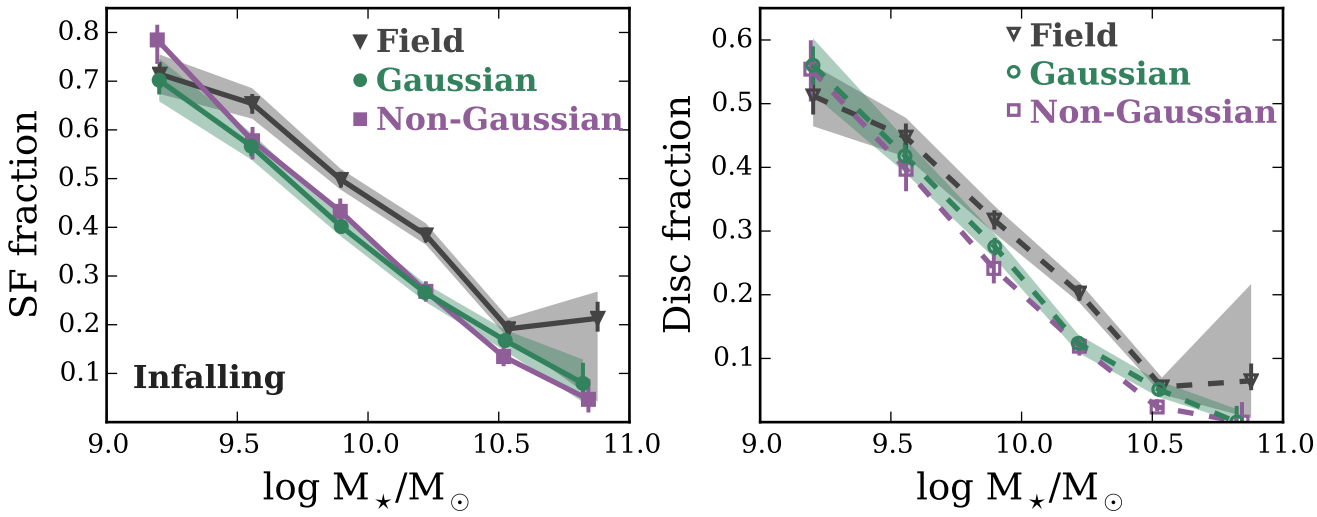

Figure 4. Star forming and disc fraction versus stellar mass for field galaxies as well as infalling galaxies in the G and NG samples. Error bars correspond to 1σ binomial confidence intervals as given in Cameron (2011), shaded regions are 90 per cent Monte Carlo confidence intervals derived from the stochastic nature of the matching procedure.

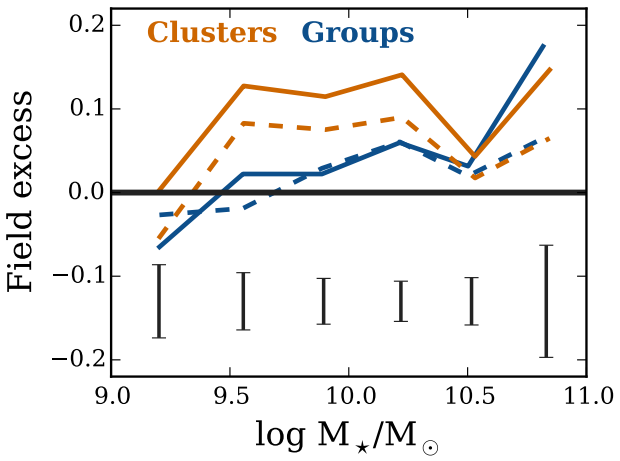

Figure 5. Field excess versus stellar mass for galaxies within groups ($10^{12} < M_H < 10^{14}\,\mathrm{M}_\odot$; blue) and clusters ($M_H \geq 10^{14}\,\mathrm{M}_\odot$; orange). Solid lines correspond to a field excess in star forming fraction and dashed lines correspond to a field excess in disc fraction. Mean 1σ error bars are shown for each bin.

forming or disc fraction between field galaxies and galaxies infalling into haloes, for groups and clusters separately. We use a loose definition of groups and clusters based on a halo mass cut, where groups are all haloes with $M_H < 10^{14}\,\mathrm{M}_\odot$ and conversly clusters are all haloes with $M_H \geq 10^{14}\,\mathrm{M}_\odot$. In Fig. 5 we plot the field excess, in star forming fraction (solid) and disc fraction (dashed), versus stellar mass for both groups (blue) and clusters (orange). The field excess is larger for clusters than for groups and similar trends are observed whether one considers star forming or disc fraction. This is consistent with pre-processing, as the expected relatively high degree of pre-processing of galaxies falling into clusters would drive a larger deviation from the field population, and similarly the lower expected levels of pre-processing for smaller groups would lead to little difference between galaxies infalling into these groups and galaxies in the field.

5 GALAXY PROPERTIES IN THE VIRIALIZED REGION

6 DISCUSSION

7 SUMMARY & CONCLUSIONS

ACKNOWLEDGMENTS

IDR thanks the Ontario Graduate Scholarship program for funding. LCP thanks the National Science and Engineering Research Council of Canada for funding. The authors thank F. Evans for matching together the various SDSS catalogues used in this research. We thank X. Yang et al. for making their SDSS DR7 group catalogue publicly available, L. Simard et al. for the publication of their SDSS DR7 morphology catalogue, J. Brinchmann et al. for publication of their SDSS SFRs, and the NYU-VAGC team for the publication of their SDSS DR7 catalogue. This research would not have been possible without access to these public catalogues.

Funding for the SDSS has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is http://www.sdss.org/.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab,

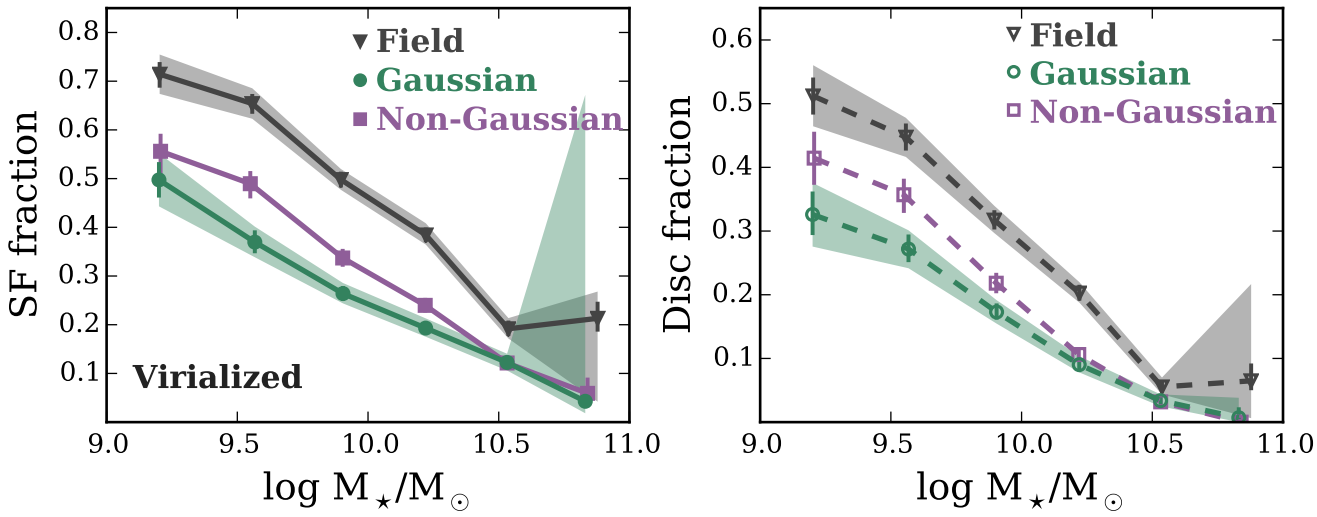

Figure 6. Star forming and disc fraction versus stellar mass for field galaxies as well as virialized galaxies in the G and NG samples. Error bars correspond to 1σ binomial confidence intervals as given in Cameron (2011), shaded regions are 90 per cent Monte Carlo confidence intervals derived from the stochastic nature of the matching procedure.

the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

REFERENCES

Abazajian K. N., et al., 2009, ApJS, 182, 543
Anderson T. W., Darling D. A., 1952, The Annals of Mathematical Statistics, 23, 193
Blanton M. R., Roweis S., 2007, AJ, 133, 734
Blanton M. R., et al., 2005, AJ, 129, 2562
Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, MNRAS, 351, 1151
Cameron E., 2011, PASA, 28, 128
Carlberg R. G., et al., 1997, ApJ, 485, L13
Cucciati O., et al., 2012, A&A, 539, A31
Feulner G., Gabasch A., Salvato M., Drory N., Hopp U., Bender R., 2005, ApJ, 633, L9
Gill S. P. D., Knebe A., Gibson B. K., 2005, MNRAS, 356, 1327
Gray M. E., Wolf C., Meisenheimer K., Taylor A., Dye S., Borch A., Kleinheinrich M., 2004, MNRAS, 347, L73
Hartigan J. A., Hartigan P. M., 1985, The Annals of Statistics, 13, 70
Hou A., Parker L. C., Harris W. E., Wilman D. J., 2009, ApJ, 702, 1199
Hou A., et al., 2013, MNRAS, 435, 1715
Hou A., Parker L. C., Harris W. E., 2014, MNRAS, 442, 406
Huchra J. P., Geller M. J., 1982, ApJ, 257, 423
Lackner C. N., Gunn J. E., 2013, MNRAS, 428, 2141
Lewis I., et al., 2002, MNRAS, 334, 673
Mahajan S., Mamon G. A., Raychaudhury S., 2011, MNRAS, 416, 2882
McGee S. L., Balogh M. L., Bower R. G., Font A. S., McCarthy I. G., 2009, MNRAS, 400, 937
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Oman K. A., Hudson M. J., Behroozi P. S., 2013, MNRAS, 431, 2307
Pimbblet K. A., 2011, MNRAS, 411, 2637
Ribeiro A. L. B., de Carvalho R. R., Trevisan M., Capelato H. V., La Barbera F., Lopes P. A. A., Schilling A. C., 2013, MNRAS, 434, 784
Rines K., Geller M. J., Kurtz M. J., Diaferio A., 2005, AJ, 130, 1482
Sanchis T., Łokas E. L., Mamon G. A., 2004, MNRAS, 347, 1198
Simard L., Mendel J. T., Patton D. R., Ellison S. L., McConnachie A. W., 2011, ApJS, 196, 11
Tasca L. A. M., et al., 2014, A&A, 564, L12
Verdugo M., Ziegler B. L., Gerken B., 2008, A&A, 486, 9
Wetzel A. R., Tinker J. L., Conroy C., 2012, MNRAS, 424, 232
Yang X., Mo H. J., van den Bosch F. C., Jing Y. P., 2005, MNRAS, 356, 1293
Yang X., Mo H. J., van den Bosch F. C., Pasquali A., Li C., Barden M., 2007, ApJ, 671, 153
Zheng X. Z., Bell E. F., Papovich C., Wolf C., Meisenheimer K., Rix H.-W., Rieke G. H., Somerville R., 2007, ApJ, 661, L41

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