Insert clever, witty title here...

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

ABSTRACT

Key words: galaxies: clusters: general – galaxies: evolution – galaxies: groups: – galaxies: statistics
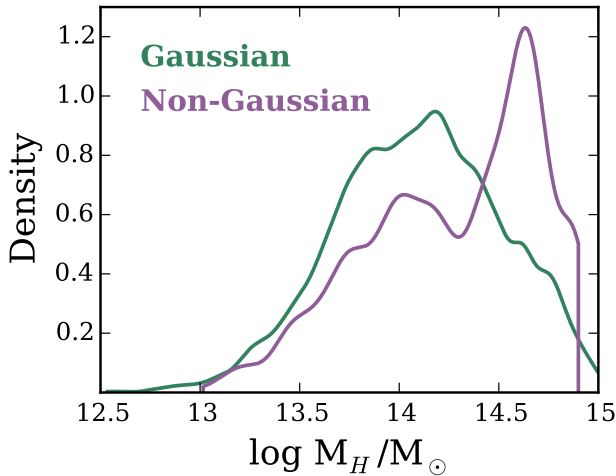
Figure 1. Smoothed host halo mass distributions for galaxies in the unmatched G and NG samples.

1 INTRODUCTION

2 DATA

2.1 Yang group catalogue

2.2 Field catalogue

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 mea-sured data sample and the CDF of a gaussian distribution. Under the assumption that the data is in fact normally dis-tributed, the AD test determines the probability (p) that the difference between the CDFs of the data and a normal dis-tribution equals or exceeds the observed difference. We ap-ply 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 cri-teria 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 un-der 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 $p_{\text{AD}} \geq 0.05$ *and* $p_{\text{Dip}} \geq 0.05$, whereas our NG data sample consists of all those groups with $p_{\text{AD}} < 0.05$ *or* $p_{\text{Dip}} < 0.05$.

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
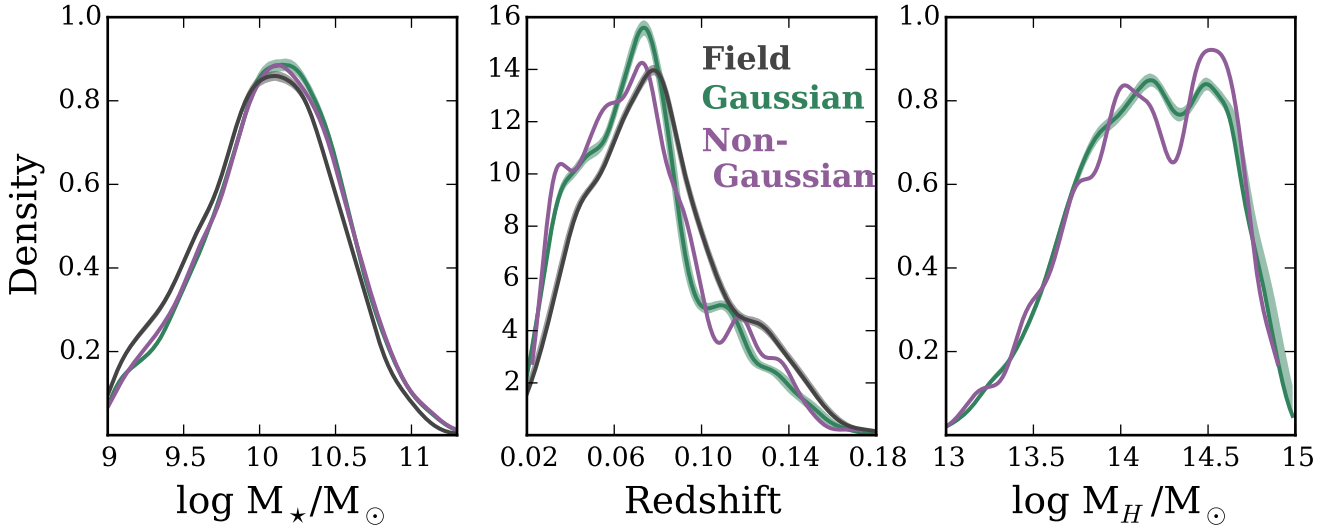
* E-mail: roberid@mcmaster.ca

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.

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 (REF); whereas the impact of group dynamics has been more difficult to pin down (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 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 tends 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 $0.1\,\mathrm{dex}$, redshifts within 0.01, and halo masses within $0.1\,\mathrm{dex}$.

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 or field samples is implicitly referencing the matched samples.
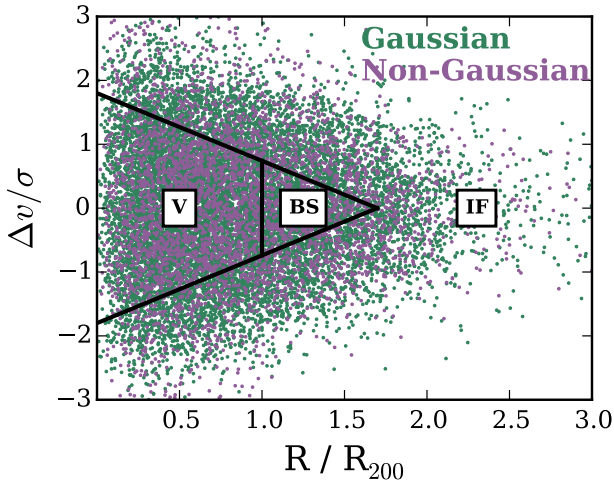
Figure 3.

3 GALAXY PROPERTIES IN THE INFALL REGION

4 GALAXY PROPERTIES IN THE VIRIALIZED REGION

5 DISCUSSION

6 SUMMARY & CONCLUSIONS

REFERENCES

Anderson T. W., Darling D. A., 1952, The Annals of Mathematical Statistics, 23, 193

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

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

This paper has been typeset from a TeX/LaTeX file prepared by the author.

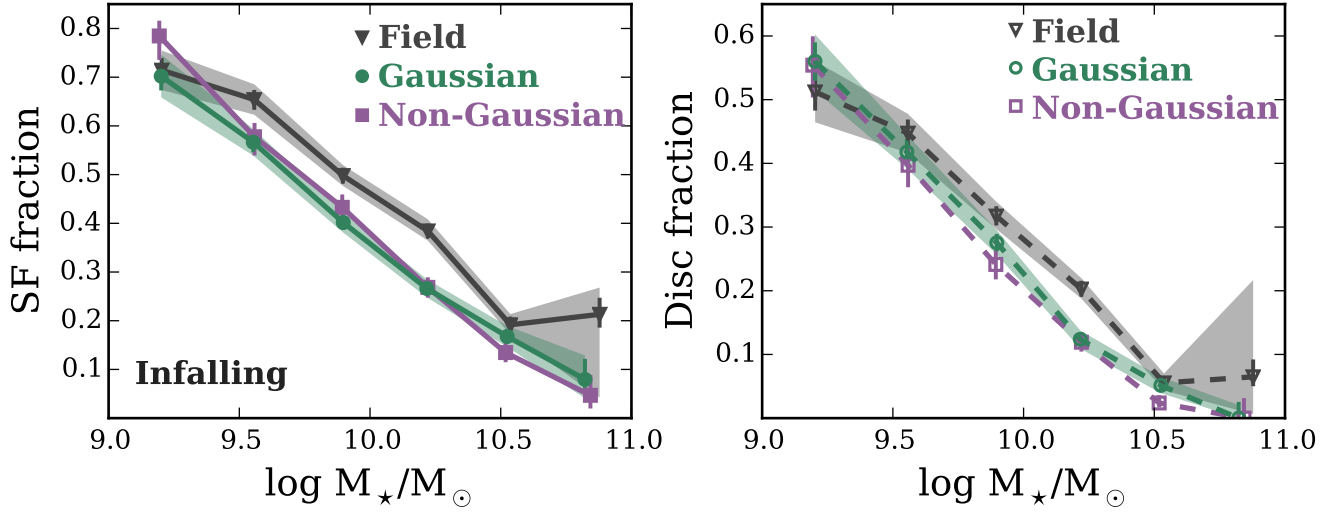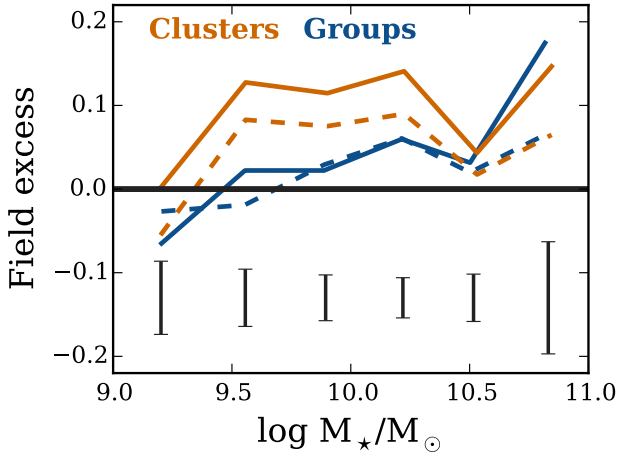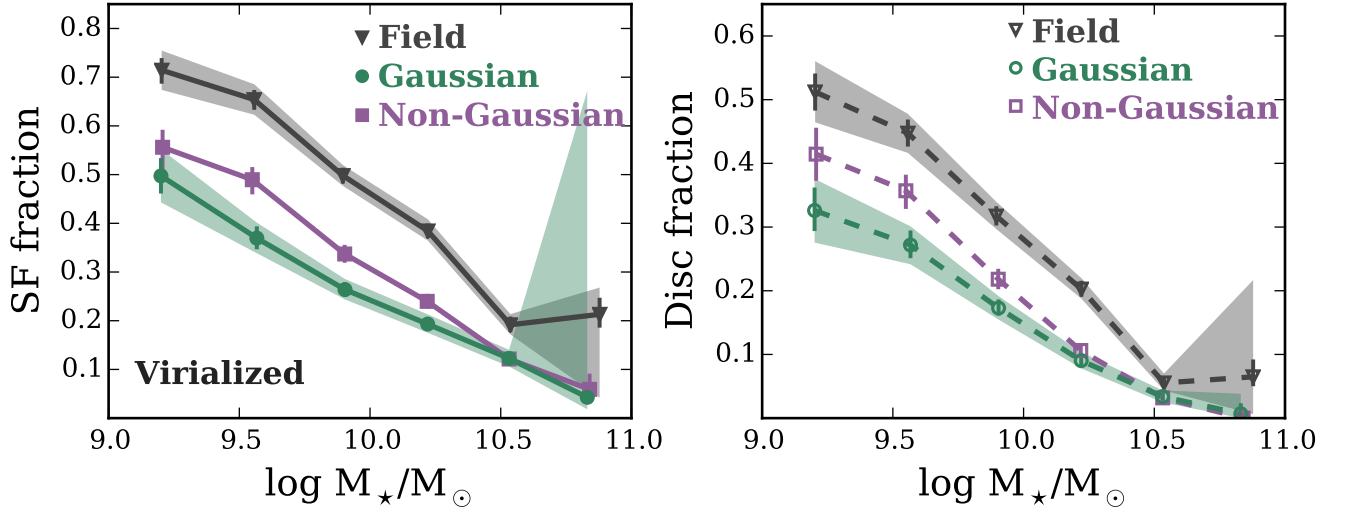**Figure 4.**

Figure 5.

Figure 6.