Non-Gaussian velocity distributions – the effect on virial mass estimates of galaxy groups

A. L. B. Ribeiro,1⋆ P. A. A. Lopes2 and M. Trevisan3
1Laboratório de Astrofísica Teórica e Observacional, Universidade Estadual de Santa Cruz, 45650-000 Ilhéus-BA, Brazil
2Observatório do Valongo, Universidade Federal do Rio de Janeiro, 20080-090 Brazil
3Instituto Astronômico e Geofísico – USP, São Paulo, SP 05508-090, Brazil

Accepted 2011 February 22. Received 2011 February 6; in original form 2010 December 31

ABSTRACT

We present a study of nine galaxy groups with evidence for non-Gaussianity in their velocity distributions out to 4R200. This sample is taken from the 57 groups selected from the 2dF Percolation-Inferred Galaxy Groups (2PIGG) catalogue of galaxy groups. Statistical analysis indicates that the non-Gaussian groups have masses significantly higher than that of the Gaussian groups. We also have found that all non-Gaussian systems seem to be composed of multiple velocity modes. Besides, our results indicate that multimodal groups should be considered as a set of individual units with their own properties. In particular, we have found that the mass distributions of such units are similar to that of the Gaussian groups. Our results reinforce the idea of non-Gaussian systems as complex structures in the phase space, likely corresponding to secondary infall aggregations at a stage before virialization. The understanding of these objects is relevant for cosmological studies using groups and clusters through the mass function evolution.

Key words: galaxies: groups: general – galaxies: statistics.

1 INTRODUCTION

Groups of galaxies contain most of the galaxies in the Universe and are the link between individual galaxies and larger structures (e.g. Huchra & Geller 1982; Geller & Huchra 1983; Nolthenius & White 1987; Ramella, Geller & Huchra 1989). The dissipationless evolution of these systems is dominated by gravity. Interactions over a relaxation time tend to distribute the velocities of the galaxy members in a Gaussian distribution (e.g. Bird & Beers 1993). Thus, a way to access the dynamical stage of galaxy groups is to study their velocity distributions. Evolved systems are supposed to have Gaussian velocity distributions, while those with deviations from normality are understood as less evolved systems. Hou et al. (2009) have examined three goodness-of-fit tests [Anderson–Darling (AD), Kolmogorov and $\chi^2$ tests] to find which statistical tool is best able to distinguish between relaxed and non-relaxed galaxy groups. Using Monte Carlo simulations and a sample of groups selected from the CNOC2, they found that the AD test is far more reliable at detecting real departures from normality in small samples. Their results show that Gaussian and non-Gaussian groups present distinct velocity dispersion profiles, suggesting that discrimination of groups according to their velocity distributions may be a promising way to access the dynamics of galaxy systems.

Recently, Ribeiro, Lopes & Trevisan (2010) extended up this kind of analysis to the outermost edge of groups to probe regions where galaxy systems might not be in dynamical equilibrium. They found significant segregation effects after splitting up the sample in Gaussian and non-Gaussian systems. In the present work, we try to further understand the nature of non-Gaussian groups. In particular, we investigate the problem of mass estimation for this class of objects and their behaviour in phase space. The Letter is organized as follows: in Section 2 we present data and methodology; Section 3 contains a statistical analysis with sample tests and multimodality diagnostics for non-Gaussian groups; finally, in Section 4 we summarize and discuss our findings.

2 DATA AND METHODOLOGY

2.1 2PIGG sample

We use a subset of the 2PIGG catalogue corresponding to groups located in areas of at least 80 per cent redshift coverage in 2dF data out to 10 times the radius of the systems, roughly estimated from the projected harmonic mean (Eke et al. 2004). The idea of working with such large areas is to probe the effect of secondary infall on to groups. Members and interlopers were redefined after the identification of gaps in the redshift distribution according to the technique described by Lopes et al. (2009a). Finally, a virial analysis is performed to estimate the groups’ properties. See details in Lopes et al. (2009a), Ribeiro et al. (2009) and Ribeiro et al. (2010). We have classified the groups after applying the AD test to their galaxy velocity distributions (see Hou et al. 2009 for a good description

⋆E-mail: albr@uesc.br

© 2011 The Authors
Monthly Notices of the Royal Astronomical Society © 2011 RAS

Downloaded from https://academic.oup.com/mnrasl/article-abstract/413/1/L81/1747626/1747626 on 28 July 2018
of the test). This is done for different distances, producing the following ratios of non-Gaussian groups: 6 per cent \((R \leq R_{200})\), 9 per cent \((R < 2R_{200})\) and 16 per cent \((R < 3R_{200} < R < 4R_{200})\). We assume this latter ratio (equivalent to nine systems) to be correct if one desires to extend the analysis to regions where galaxy groups might not be in dynamical equilibrium. Approximately 90 per cent of all galaxies in the sample have distances \(\leq 4R_{200}\). This is the natural cut-off in space we have made in this work. Some properties of galaxy groups are presented in Table 1, where non-Gaussian groups are identified with an asterisk. Cosmology is defined by \(\Omega_m = 0.3\), \(\Omega_{\Lambda} = 0.7\) and \(H_0 = 100\ h\ \text{km}\ \text{s}^{-1}\ \text{Mpc}^{-1}\). Distance-dependent quantities are calculated using \(h = 0.7\).

3 NON-GAUSSIANITY AND MASS

3.1 Mass bias

We consider the two-sample statistical problem for Gaussian (G) and non-Gaussian (NG) subsamples. We choose the mass resulting from virial analysis as the property to illustrate the comparison between the subsamples. Both Kolmogorov–Smirnov (KS) and Cramer–von Mises (CvM) tests reject the hypothesis that the NG subsample is distributed as the G subsample, with \(p\)-values 0.00001 and 0.00029, respectively. For these tests, we have used 1000 bootstrap replicates of each subsample to alleviate the small-sample effect.

The result indicates an inconsistency between the mass distributions of the G and NG groups. This could represent a real physical difference or, more probably, an indication of a significant bias to higher masses in NG groups. The median mass for this subsample is \(57\ \text{M}_\odot\) for the G subsample; thus the mass of the G and NG groups is larger by a factor of \(\sim 2.9\). In the following, we investigate this mass bias looking for features in the velocity distributions of galaxy groups.

3.2 Exploring non-Gaussianity

The shape of the velocity distributions may reveal a signature of the dynamical stage of the galaxy groups. For instance, systems with heavier tails than that predicted by a normal parent distribution may be contaminated by interlopers. Otherwise, systems with lighter tails than a normal may be multimodal, consisting of overlapping distinct populations (e.g. Bird & Beers 1993). We now try to understand the mass bias in the NG subsample by studying non-Gaussianity in the velocity distributions. Since we have carefully removed interlopers from each field (see Section 2.1; and Lopes et al. 2009a), we consider here that the most probable cause of normality deviations in our sample is due to a superposition of modes in the phase space (e.g. Diemand & Kuhlen 2008). Visual inspection of radial velocity histograms of NG systems suggests that multipeaks really happen in most cases (see Fig. 1).

We statistically check multimodality by assuming the velocity distributions as Gaussian mixtures with unknown number of components. We use the Dirichlet process mixture (DPM) model to study the velocity distributions. The DPM model is a Bayesian non-parametric methodology that relies on Markov Chain Monte Carlo (MCMC) simulations for exploring mixture models with an unknown number of components (Diebolt & Robert 1994). It was first formalized in Ferguson (1973) for the general Bayesian statistical modelling. The DPM is a distribution over \(k\)-dimensional discrete distributions, so each draw from a Dirichlet process is itself a distribution. Here, we assume that a galaxy group is a set of \(k\) components, \(\sum_{i=1}^{k} \pi_i, f(y|\theta_i)\), with galaxy velocities distributed according to Gaussian distributions with mean and variance unknown. In this framework, the numbers \(\pi_i\) are the mixture coefficients that are drawn from a Dirichlet distribution. In the DPM model, the actual number of components \(k\) used to model the data is not fixed, and can be automatically inferred from the data employing the usual
Bayesian posterior inference framework. See Neal (2000) for a survey of MCMC inference procedures for DPM models.

In this work, we find \( k \) using the \( \mathbb{R} \) language and environment (\textit{R Development Core Team}) under the dpmixsim library (da Silva 2009). The code implements mixture models with normal structure (conjugate normal–normal DPM model). First, it finds the coefficients \( \pi_i \), and then separates the components of the mixture, according to the most probable values of \( \pi_i \), in the distributional space, leading to a partition of this space into regions (da Silva 2009). The results can be visually analysed by plotting the estimated kernel densities for the MCMC simulations. In Fig. 2, we show the DPM diagnostics for each group, that is the deblended modes in the velocity distributions. We have found the following number of modes per group: 4/102, 3/138, 3/177, 2/179, 5/352, 3/447, 2/475, 3/505 and 3/507. Therefore, all non-Gaussian groups in our sample are multimodal (reaching a total of 28 modes) according to the DPM analysis. Unfortunately, we cannot compute the physical properties of 13 modes, due to intrinsic scattering in velocity data (and/or the smallness of the modes – those with less than four members). The properties of the other 15 modes are presented in Table 2.

### 3.3 Mass bias revisited

Now, we perform again the statistical tests, comparing the distribution of mass in NG and G with the new sample (see Fig. 3.) First, we compare the NG subsample and the sample of modes (M). Both KS and CvM tests reject the hypothesis that M is distributed as NG, with \( p = 0.0211 \) and 0.0189, respectively. Then, we compare G and M. Now, the tests accept the hypothesis that M is distributed as G, with \( p = 0.4875 \) and 0.4695. Hence, the distributions of the

![Figure 1. Histograms of radial velocities for all non-Gaussian groups.](image)

![Figure 2. DPM density probability decomposition. Velocities are rescaled to the [0, 1] interval.](image)

![Figure 3. Histograms for Gaussian (dark grey) and non-Gaussian (white) groups. Histogram of modes is plotted in light grey. Density probability comparison among systems. Intersection areas are in white.](image)

| Mode | \( R_{200} \) (Mpc) | \( M_{200} \) (10^{12} M_\odot) | \( \sigma \) (km s\(^{-1}\)) | \( N_{200} \) |
|------|-----------------|------------------|----------------|---------|
| 102a | 1.077           | 1.549            | 337.482        | 21      |
| 102b | 0.728           | 0.477            | 153.372        | 4       |
| 138a | 1.550           | 4.677            | 498.927        | 35      |
| 177a | 1.249           | 2.481            | 369.569        | 11      |
| 177b | 1.212           | 2.289            | 362.462        | 14      |
| 179a | 1.326           | 3.001            | 336.265        | 8       |
| 352a | 0.487           | 0.142            | 158.571        | 6       |
| 352b | 0.941           | 1.033            | 218.210        | 6       |
| 352c | 0.697           | 0.416            | 208.670        | 3       |
| 352d | 0.558           | 0.213            | 181.206        | 5       |
| 447a | 1.016           | 1.330            | 240.437        | 3       |
| 447b | 1.192           | 2.155            | 441.382        | 5       |
| 475a | 1.500           | 4.308            | 470.516        | 19      |
| 475b | 2.345           | 16.347           | 829.977        | 10      |
| 505a | 0.928           | 1.040            | 228.411        | 3       |
|      |                 |                  |                |         |

The letters a, b, c, d indicate the individual velocity modes in each NG system.
modes deblended from non-Gaussian groups are themselves mass distributed as Gaussian groups. Also, the median mass of the M sample is $\langle M_{200}^{\text{NG}} \rangle = 1.33 \times 10^{14} M_\odot$, a value larger than $\langle M_{200}^{\text{G}} \rangle$ only by a factor of $\sim 1.5$ (before deblending the groups the factor was $\sim 2.9$). This consistency between G and M objects indicates that the non-Gaussian groups are a set of smaller systems, probably forming an aggregate out of equilibrium. Indeed, groups are the intermediate level between galaxies and larger galaxy systems. Thus, our results suggest that we are witnessing secondary infall (the secondary modes) on to a previously formed (or still forming) galaxy system (the principal mode). Naturally, we cannot discard the possibility of the NG groups being unbound systems of smaller groups seen in projection, although the properties of the independent modes are quite similar to those found in physically bound groups.

4 DISCUSSION

We have classified galaxy systems after applying the AD normality test to their velocity distributions up to the outermost edge of the groups. The purpose was to investigate regions where galaxy systems might not be in dynamical equilibrium. We have studied 57 galaxy groups selected from the 2PIGG catalogue (Eke et al. 2004) using 2dF data out to 4$R_{200}$. This means that we probe galaxy distributions near to the turnaround radius, thus probably taking into account all members in the infall pattern around the groups (e.g. Rines & Diaferio 2006; Cupani, Mezzetti & Mardirossian 2008).

The corresponding velocity fields depend on the local density of matter. High-density regions should drive the formation of virialized objects, whereas low-density environments are more likely to present streaming motions, i.e. galaxies falling towards larger potential wells constantly increasing the amplitude of their clustering strength (e.g. Diaferio & Geller 1997).

We have found that 84 per cent of the sample is composed of systems with Gaussian velocity distributions. These systems could result from the collapse and virialization of high-density regions with no significant secondary infall. They could be groups surrounded by well-organized infalling motions, possibly reaching virialization at larger radii. Theoretically, in regions outside $R_{200}$, we should apply the non-stationary Jeans formalism, leading us to the virial theorem with some correction terms. These are due to the infall velocity gradient along the radial coordinate and to the acceleration of the mass accretion process (see Cupani 2008). Their contribution is likely to be generally negligible in the halo core where the matter is set to virial equilibrium and it becomes significant in the halo outskirts where the matter is still accreting (see e.g. Cupani et al. 2008). However, a slow and well-organized infalling motion to the centre of the groups could diminish the importance of the correction terms, i.e. the systems could be in a quasi-stationary state outside $R_{200}$. The high fraction of groups immersed in such surroundings suggests that ordered infalling motion around galaxy groups might happen quite frequently in the Universe.

In any case, Gaussian groups can be considered as dynamically more evolved systems (see Ribeiro et al. 2010). The remaining 16 per cent of the sample is composed of non-Gaussian groups. We found that these systems have masses significantly larger than that of Gaussian groups. This biasing effect in virial masses is basically due to the higher velocity dispersions in NG groups. Ribeiro et al. (2010) found that NG groups have rising velocity dispersion profiles. A similar result was found by Hou et al. (2009). Rising velocity dispersion profiles could be related to the higher fraction of blue galaxies in the outskirts of some galaxy systems (see Popesso et al. 2007; Ribeiro et al. 2010).

At the same time, the NG subsample is composed of multipeak objects, identified by the DPM model analysis applied to the velocity space. These results indicate that secondary infall might be biasing the mass estimates of these groups. Thus, the NG systems could result from the collapse of less dense regions with significant secondary infall input. Contrary to Gaussian groups, the surroundings of the NG systems do not seem to be in a quasi-stationary state. They are dynamically complex. Actually, we have found that these groups can be modelled as assemblies of smaller units. After deblending the groups into a number of individual modes, we have verified that the mass distribution of these objects is consistent with that of Gaussian groups, suggesting that each unit is probably a galaxy group itself, a system formed during the streaming motion towards the potential wells of the field.

Our results reinforce the idea of NG systems as complex structures in the phase space out to $4R_{200}$. This scenario is illustrated in Fig. 4, where we show the stacked G and NG groups, and the stacked modes. Galaxies in these composite groups have distances and line-of-sight velocities with respect to the centres normalized by $R_{200}$ and $\sigma$, respectively. In the upper box, we present the G stacked group. Note that galaxies are extensively concentrated in the phase-space diagram, with a single density peak near to $0.5R_{200}$, revealing a well-organized system around this point. Also, note that the density peak has $V_{\text{los}}/\sigma \approx 0$. Contour density lines suggest that ordered shells of matter are moving towards the centre. A different result is found for the NG stacked group. In the middle box of Fig. 4, we see a less concentrated galaxy distribution, with less tight density contour levels, presenting a density peak slightly larger than $R_{200}$, and two additional peaks, around $2R_{200}$, possibly interacting with the central mode. The additional peaks are not aligned at $V_{\text{los}}/\sigma \approx 0$, suggesting a less symmetrical galaxy distribution in the phase space. These features suggest that non-Gaussian systems are distinct, and dynamically younger than Gaussian groups, which

Figure 4. Phase-space diagrams for a typical Gaussian (upper box), non-Gaussian (middle box) and mode (lower box) systems. Density contours indicate galaxy concentration across the diagrams. Distances to the centre of the systems are normalized by $R_{200}$. Radial velocities are subtracted from the median velocity and divided by the velocity dispersion of the groups.
agrees well with the results of Ribeiro et al. (2010). Also in Fig. 4, we present the mode stacked system in the lower box. Similar to the Gaussian case, we have a single peak in the phase space, near to $1.5R_{200}$. Galaxy distribution however is still less symmetrical than in Gaussian groups, with the peak not aligned at $V_{\text{los}}/\sigma \approx 0$. This suggests that, although more organized than NG systems, modes are probably dynamically distinct and younger than Gaussian groups as well.

Our work points out the importance of studying NG systems both to possibly correct their mass estimates and multiplicity functions and to better understand galaxy clustering at group scale. The understanding of these objects is also relevant for cosmological studies using groups and clusters through the evolution of the mass function (Voit 2005). Using systems with overestimated properties may lead to a larger scatter in the mass calibration (Lopes et al. 2009b) and could also affect the mass function estimate (Voit 2005).

ACKNOWLEDGMENTS

We thank the referee for raising interesting points. We also thank A. C. Schilling and S. Rembold for helpful discussions. ALBR thanks the support of CNPq, grants 306870/2010-0 and 478753/2010-1. PAAL thanks the support of FAPERJ, process 110.237/2010. MT thanks the support of FAPESP, process 2008/50198-3.

REFERENCES

Bird C., Beers T., 1993, AJ, 105, 1596
Cupani G., 2008, PhD thesis, Univ. Trieste
Cupani G., Mezzetti M., Mardirossian F., 2008, MNRAS, 390, 645
da Silva A. F., 2009, Comput. Methods Programs Biomedicine, 94, 1
Diaferio A., Geller M., 1997, ApJ, 481, 633
Diebolt J., Robert C. P., 1994, J. R. Statistical Soc. Ser. B, 56, 363
Diemand J., Kuhlen M., 2008, ApJ, 680, L25
Eke V. R. et al., 2004, MNRAS, 348, 866
Ferguson T. S., 1973, Ann. Statistics, 1, 209
Geller M. J., Huchra J. P., 1983, ApJS, 52, 61
Hou A., Parker L., Harris W., Wilman D. J., 2009, ApJ, 702, 1199
Huchra J. P., Geller M. J., 1982, ApJ, 257, 423
Lopes P. A. A., de Carvalho R. R., Kohl-Moreira J. L., Jones C., 2009a, MNRAS, 392, 135
Lopes P. A. A., de Carvalho R. R., Kohl-Moreira J. L., Jones C., 2009b, MNRAS, 399, 2201
Neal R. M., 2000, J. Comput. Graphical Statistics, 9, 249
Nolthenius R., White S. D. M., 1987, MNRAS, 225, 505
Popesso P., Biviano A., Böringer H., Romaniello M., 2007, A&A, 461, 397
Ramella M., Geller M., Huchra J. P., 1989, ApJ, 344, 57
Ribeiro A. L. B., Trevisan M., Lopes P. A. A., Schilling A. C., 2009, A&A, 505, 521
Ribeiro A. L. B., Lopes P. A. A., Trevisan M., 2010, MNRAS, 409, L124
Rines K., Diaferio A., 2006, AJ, 132, 1275
Voit G. M., 2005, Rev. Modern Phys., 77, 207

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