MAGGIE: Models and Algorithms for Galaxy Groups, Interlopers and Environment

Manuel Duarte\textsuperscript{1}⋆ and Gary A. Mamon\textsuperscript{1}†

\textsuperscript{1} Institut d’Astrophysique de Paris, Paris, France (UMR 7095: CNRS & UPMC)

ABSTRACT

Redshift distortions make it notoriously difficult to recover real-space galaxy groups from redshift space data. We make use of the previously measured universal distribution of halo interlopers in projected phase space, which we combine with our knowledge of NFW haloes with realistic internal kinematics, to develop a prior-based, probabilistic, abundance matching (AM) grouping algorithm, which we call MAGGIE. We perform a suite of tests of MAGGIE-L and MAGGIE-M (in which the AM is performed on the group luminosities and stellar masses, respectively) as well as on our previously optimised Friends-of-Friends (FoF) group finder, all for groups of at least 3 galaxies extracted from two subsamples that are doubly complete in distance and luminosity within a mock, flux-limited, SDSS Legacy redshift survey, incorporating (for MAGGIE) reasonable observational errors on galaxy luminosities and stellar masses.

Our tests indicate that both MAGGIE-M and MAGGIE-L outperform the optimal FoF in all three considered bins of estimated group mass (12–13, 13–14, and 14–15 in $\log M/M_\odot$) in our various tests: groups extracted with MAGGIE are much less likely to be secondary fragments of true groups (typically 15% vs. >30% with FoF), and restricting to the primary fragments, the galaxy memberships (relative to the virial sphere of the real-space group) are much more complete and usually more reliable, the MAGGIE AM masses are much less biased and with less scatter (<0.25 dex vs. >0.3 dex for the FoF masses obtained by the virial theorem), as are the group luminosities and stellar masses (computed in MAGGIE using the membership probabilities as weights). Only for the reliability (purity) of the galaxy population in high-estimated-mass clusters does FoF outperform both implementations of MAGGIE. For the tests in common with other AM group finders, MAGGIE has lower mass scatter than Yang et al., higher galaxy completeness than Yang et al., Muñoz-Cuartas & Müller and Domínguez Romero et al., but cannot match the latter’s zero mean group fragmentation.

The excellent performance of both flavours of MAGGIE should lead to sharper trends of environmental effects on galaxies, as well as more precise mass/orbit modeling of groups of galaxies. MAGGIE should be adaptable (discarding AM) to optimal group/cluster detection and mass measurement in dark energy surveys.

Key words: galaxy groups — galaxy: environment.

1 INTRODUCTION

In the hierarchical growth of structure in the Universe, dominated by gravity (and dark energy), matter flows from low to high density regions. To first order, galaxies, which form in small dark matter haloes, follow this evolution and cluster into galaxy systems called clusters or groups, depending on their mass (clusters are often defined with masses within the virial radius greater than $10^{14} M_\odot$).

The properties of galaxies within these systems (hereafter denoted groups for simplicity), now attached to dark matter sub-

haloes, are likely to be modified by the peculiar environment of their parent groups. Many physical processes should indeed alter galaxy properties in groups: the high galaxy density in groups will lead to galaxy interactions and possibly mergers; the deeper gravitational potentials of the more massive groups will produce higher velocity dispersions for the galaxy population, favouring rapid flybys over mergers (e.g., [Mamon 1992]; the tides from the group potential will prevent outer gas from accreting onto galaxy disks [Larson et al. 1980]; the diffuse intra-group gas will exert ram pressure on the galaxy’s gas [Gunn & Gott 1972] and either compress it, enhancing star formation (e.g., [Kennicutt et al. 1984], or when the pressure gets very high it will expel the gas [Gunn & Gott], decreasing subsequent star formation.

⋆ E-mail: duarte@iap.fr
† E-mail: gam@iap.fr
Galaxy groups and clusters thus represent an ideal laboratory to test the environmental effects on galaxies in models of galaxy formation and evolution. Groups and clusters are also an important tool to probe the cosmological parameters, such as the dark energy parameter \cite{Wang2011}. Moreover, clusters have been recently used to test a major prediction of general relativity, with the recent discovery of weak but significant signs of gravitational redshifts \cite{Johansson2011}.

Since the early discovery of morphological segregation of galaxies in clusters \cite{Hubble1936, Shapley1926, van1960}, i.e. where inner regions of clusters preferentially contain elliptical galaxies, whose red colours are indicative of old stellar populations, it has been clear that the efficiency with which stars form within galaxies must depend on their environment. In other words, high density environments act to quench star formation in galaxies. More specifically, the specific star formation rate (SSFR) of galaxies (star formation rate divided by stellar mass) is likely to be a function of two separate environmental parameters: the global environment characterized by the total mass of their group, and the local environment that measures the position of the galaxy within its group.

\cite{Peng2010} studied the dependence of SSFR with stellar mass and environment, where they quantified the latter by the distance to the 5th nearest neighbour. They found that, at low stellar mass, SSFR varied more with the density of the environment, while at high mass, the environmental effects are small and the SSFR anti-correlates with stellar mass. Unfortunately, the use of an environment tracer such as the 5th nearest neighbour produces a mix between the global and local environments. In contrast to \cite{Peng2010, Weinmann2006} and \cite{von2010}, both considered indicators of both the global and local environments. \cite{Weinmann2006} found that the fraction of late-type satellite galaxies appears more anti-correlated with group mass than with stellar mass, while \cite{von2010} found that high-mass galaxies also show some moderate dependence of SSFR with the relative distance to the group/cluster centre (albeit limited to low cluster-centric radii).

These possible disagreements highlight the importance of properly measuring the global and local environments. Unfortunately, a clean characterization of the real space environment from the redshift space observed distribution of galaxies is difficult since the redshift distortions \cite{Jackson1972} caused by the velocity dispersion of the galaxy group distorts the group into elongated structures pointing towards the observer, i.e. Fingers-of-God \cite{Tully1978}. Moreover, because of redshift distortions, real space groups can be merged into single groups in redshift space. Conversely, grouping algorithms inevitably lead to some fragmentation of real-space groups, so that the secondary fragments do not represent bona-fide groups of their own (although some may represent subgroups of real-space groups). Finally, the even without group merging and fragmentation, a group finder may miss some of the real-space group galaxies, leading to incomplete galaxy membership; or conversely may include additionally galaxies that lie outside of the virial sphere of the real space group, producing unreliable galaxy membership. One then wonders to what extent the strength of environmental effects on galaxy properties may be washed out by the imperfect extraction of the global and local environments by the grouping algorithm (group finder).

Many galaxy group catalogues have already been published, usually following the first publications of data from galaxy surveys. First attempts were made with visual selections based on well-defined criteria \cite{Abell1958, Zwicky1961} for clusters and \cite{Rose1977, Hickson1982} for compact groups). The first automated (and simple) algorithm has been the percolation or Friends-of-Friends (FoF) method, first introduced by \cite{Turner1976} and \cite{Huchra1982}, in which groups are built by collections of galaxies linked, two-by-two, by their proximity. Redshift distortions are taken into account by the use of two different linking lengths, along the line-of-sight (LOS) and transverse directions. There is a fairly wide range of pairs of linking lengths used in the literature. In a previous study \cite{Duarte2014} hereafter Paper 1, we have analysed several mock SDSS samples of galaxies to optimize the pairs of linking lengths for minimal group fragmentation and merging, maximal galaxy completeness and reliability, and maximal group mass accuracy (see also \cite{Eke2003, Berlind2006} and \cite{Robotham2011}, and compared their optimal linking lengths with those of ten previous implementations of the FoF algorithm. Another fairly non-parametric grouping algorithm is to partition redshift space into Voronoi cells, constructed from Delaunay triangulation, providing local galaxy number densities that are inversely proportional to the volumes of the Voronoi cells \cite{Marinoni2002} see also \cite{Gerke2005}.

Building on our recently gained knowledge from cosmological \textit{N}-body simulations, grouping algorithms have begun to appear, where priors on galaxy group properties are incorporated to improve their extraction from galaxy redshift surveys. In pioneering studies, \cite{Yang2005, Yang2007} developed an iterative method to select galaxy groups based on a density contrast criterion in projected phase space (PPS, i.e. projected radius and LOS velocity dispersion), assuming a \cite{Navarro1996} hereafter NFW surface density profile and a Maxwellian LOS velocity distribution, both in reasonably good agreement with what is found in the group- and cluster-mass haloes of dissipationless cosmological simulations. In the Yang et al. group finder, the group masses, hence virial radii, are determined by \textit{abundance matching} (first introduced by \cite{Marinoni1999}, which assumes a one-to-one correspondence between group luminosity or stellar mass and its total (halo) mass to match the cumulative distribution functions (CDFs) of the cosmic halo mass function to the group luminosity or stellar mass function (measured, here, in the previous iteration of the algorithm). Abundance matching between groups and haloes has also been introduced by \cite{Muñoz2012} in their FoF algorithm that links haloes rather than galaxies: they consider halo virial radii in the transverse direction and maximum circular velocity in the LOS direction, combining the two links in an ellipsoidal fashion.

But galaxy surveys come with observational problems that are difficult to handle: surveys suffer from edge effects and from bright (saturated) stars masking regions, and those with photometric redshifts have large and sometimes catastrophic redshift errors. Probabilistic methods appear to be a promising way to deal with these aspects. For example, \cite{Li2008} designed a probabilistic FoF method for surveys with photometric redshifts. \cite{Romero2012} have recently adapted the Yang et al. group finder into a probabilistic algorithm: they initially assign haloes to single galaxies, and use abundance matching like Yang et al. to assign group masses and radii. But \cite{Romero2012} end their algorithm with a hard assignment of galaxies to their groups.

These studies can be improved in several respects:

(i) In their prediction of the density in PPS, \cite{Yang2005, Yang2007} and \cite{Romero2012} assume that the LOS velocity dispersion is independent of projected radius, while cosmological \textit{N}-body simulations (starting with \cite{Cole1996} indicate a convex profile in log-log. One can easily predict this LOS...
velocity dispersion profile (see Mamon & Lokas 2005 for a single integral expression) by solving the Jeans equation of local dynamical equilibrium, adopting the velocity anisotropy profile of the particles in the halos of CDM cosmological simulations (hereafter CDM haloes).

(ii) Yang et al. (2005, 2007) and Domínguez Romero et al. (2012) assume that the LOS velocity distribution is Maxwellian, whereas the velocity anisotropy alters this Gaussianity (Merritt 1987), hence one can do better and predict its precise shape from the three-dimensional velocity distribution (Mamon, Biviano, & Boué 2013). The threshold in the PPS density used by Yang et al. (2005, 2007) and Domínguez Romero et al. (2012) is ad hoc. Instead, one can take advantage of our knowledge of the distribution of galaxies in PPS for two terms: (a) the galaxies within the virial sphere of the parent real-space group (hereafter the halo term); (b) the galaxies that are in the virial cone but outside the virial sphere (hereafter the interloper term). The interloper PPS density was quantified by Mamon, Biviano, & Murante (2010) using a cosmological simulation, and it turns out to be fairly independent of halo mass (for cluster-mass haloes).

(iv) Comparing the PPS densities from the halo and interloper terms yields a probability of membership. There is no need to perform a hard assignment of galaxies to groups in the end as was done by Domínguez Romero et al. (2012). Group properties are easily obtained using the membership probabilities as weights.

(v) Yang et al. (2007) employ a complicated and imprecise scheme (see their fig. 4) to estimate how the luminosity incompleteness varies with redshift in their flux-limited sample. Errors in the luminosity incompleteness will propagate, among other places, to the abundance matching technique they use to infer group masses. However, the issue can be entirely avoided by restricting the group finder to subsamples that are doubly complete in both distance and luminosity. Admittedly, such samples are, at best, less than one-third the size of the parent flux-limited samples (see Tempel et al. 2014 for the SDSS). However, the very large sizes of the samples from recent or ongoing galaxy spectroscopic surveys (250 000 for the Two Degree Field Galaxy Redshift Survey [2dFGRS, Colless et al. 2001], 125 000 for the Six Degree Field Galaxy Survey [6dFGRS, Jones et al. 2009], 700 000 for the primary spectroscopic sample of the Sloan Digital Sky Survey [SDSS, Abazajian et al. 2009], 300 000 for the ongoing Galaxy and Mass Assembly survey [GAMA, Hopkins et al. 2013]) lead to substantial sizes for doubly complete subsamples, which can be used for studies of environmental effects on galaxies. Moreover, it is wiser to study environmental effects on a group catalogue derived from a doubly-complete galaxy subsample, rather than start with a group catalogue derived from a flux-limited subsample and then cut it into a doubly complete subsample of groups to study environmental effects. For example, Tempel et al. (2014) have recently produced publicly available FoF group catalogues that they had run on doubly-complete SDSS galaxy subsamples.

(vi) Yang et al. (2007), Muñoz-Cuartas & Müller (2012) and Domínguez Romero et al. (2012) wisely test their grouping algorithms using mocks. However, their adopted definitions for purity and contamination take values above (and below) unity, while we prefer a measure of the reliability that is restricted to values between zero and unity (see Sect. 3.3 below). Moreover, these mocks should include observational errors (on galaxy luminosities and stellar masses), and while this is briefly mentioned by Yang et al. (2005), it is not clear what level of errors were considered by them and Yang et al. (2007), while observational errors were not mentioned by Domínguez Romero et al. (2012).

In this work, we present a new probabilistic grouping algorithm, Models and Algorithms for Galaxy Groups, Interlopers and Environment, a.k.a. MAGGIE. The galaxy membership of MAGGIE groups is determined probabilistically, combining the distribution of interlopers in PPS measured by Mamon et al. (2010) with a realistic model for the distribution in PPS of halo members, while the group masses are determined by abundance matching in an iterative fashion, as in Yang et al. (2005, 2007).

We present MAGGIE in Sect. 3 and our mocks and testing procedure are described in Sect. 4. We compare, in Sect. 5, the optimal FoF group finder with two implementations of MAGGIE, on their ability to recover physical properties and galaxy membership of real-space groups, we discuss our findings in Sect. 5 and summarize our results in Sect. 6.

2 MAGGIE

We present here a complete description of the different steps of MAGGIE. We start with a basic description of the algorithm, and then we explain how we take into account the edges of the galaxy sample.

2.1 Basic Grouping Algorithm

We assume that we have a galaxy sample that is doubly complete in distance and luminosity, with positions on the sky (right ascension and declination), redshifts, as well as apparent magnitudes in a given waveband and/or stellar masses. This is the minimum required data set.

MAGGIE groups are built around the most luminous galaxy (MAGGIE-L) or the most massive in stars (MAGGIE-M). This galaxy is assumed to be the central galaxy and at rest relative to the group. Although the most massive group galaxies can be offset and not at rest with the group (e.g., Skibba et al. 2011), we prefer this definition to the barycentre, since the galaxy number density profiles in clusters are known to be less cuspy when clusters are centered on their barycentres (Beers & Tonry 1986), and indeed most analyses adopt the central galaxy as the position of the group centre.

MAGGIE then builds groups with the following iterative method. Steps 2 to 4 are similar to those of Yang et al. (2005, 2007).

1. Sort galaxies by decreasing stellar mass and loop over potential groups

We loop over the potential group central galaxies, sorted by decreasing galaxy stellar mass (MAGGIE-M) or luminosity (MAGGIE-L), performing the following steps:

2. Group total masses

2a. Initial group total masses On first pass, we determine the mass of each group, either by adopting group masses $M = 300 L_r$ (MAGGIE-L) or using the relation between halo mass and central galaxy stellar mass (MAGGIE-M) that Behroozi, Conroy, & Wechsler (2010) derived from abundance matching (basically matching the halo mass and central galaxy stellar mass CDFs). This initial choice has no effect on the final outcome (see Sect. 5.3.1).

2b. Group total masses on subsequent iterations On subsequent passes, we determine the group mass by performing our own abundance matching between our group luminosity (MAGGIE-L) or group stellar mass (MAGGIE-M) function (determined in the
previous pass) and a chosen halo mass function:

\[
N(> M_{\text{group}}) = N(> M) \quad \text{(MAGGIE-L)}
\]

\[
N(> m_{\text{group}}) = N(> M) \quad \text{(MAGGIE-M)}
\]

where \( M \) is the group (halo) total mass, \( L_{\text{group}} \), and \( m_{\text{group}} \), and represent the group luminosity and stellar mass, respectively.\(^{[1]}\) (In this abundance matching, we must assume that the group has the same central galaxy as in the previous iteration, which is true for the great majority of groups.) The cumulative mass functions are considered for the comoving volume of the subsample, i.e.

\[
N(> M) = \int_{z_{\text{min}}}^{z_{\text{max}}} \left( \frac{dV}{dz} \right) \int_{M}^{\infty} f(M', z) \, dM',
\]

where \( f(M) \) is the differential halo mass function. Numerically solving equations \([1]\) or \([2]\) together with equation \([3]\) provides the group total mass as a function of the group luminosity or stellar mass.\(^{[2]}\)

\([3]\) **Group radii**

We estimate the group radius from the group mass, using

\[
r_{200} = \left( \frac{GM_{200}}{100 H^2(z)} \right)^{1/3},
\]

where \( r_{200} \) is our proxy for the virial radius and is the radius of the virial sphere centered on the position of the central galaxy and whose mean density is 200 times the critical density of the Universe, \( \rho_{\text{crit}} = 3H^2(z)/(8\pi G) \), while \( M_{200} \) is the mass within the virial sphere. The Hubble constant, in equation \([4]\), for a flat Universe, is

\[
H(z) = H_0 \sqrt{\Omega_m(1+z)^3 + 1 - \Omega_m}.
\]

The factor 100 in equation \([4]\) is really \( \Delta/2 \) for the overdensity relative to critical of \( \Delta = 200 \).\(^{[3]}\)

\([4]\) **Coordinates in projected phase space**

The projected separation \( R \) (hereafter, **projected radius**) and LOS velocity \( v \) of a galaxy relative to a central group galaxy (assumed at rest in the group) are written with the standard cosmological formulae:

\[
R = \theta \, d_{\text{ang}}(z_{\text{group}}),
\]

\[
v = c \frac{(z - z_{\text{group}})}{1 + z_{\text{group}}},
\]

where \( \theta \) is the angular separation, \( c \) is the speed of light,

1 We denote stellar masses as \( m \) and total (group/halo) masses (including dark matter and gas) as \( M \).

2 In this work, the cumulative halo mass function \( N(> M) \) is measured on the simulation from which our mocks were built, but a theoretical halo mass function can be chosen when working on real data (see Sect. 5.3.2). Also, while it would be preferable to fit an analytical form to the total group luminosity function or stellar mass function, to avoid shot noise and cosmic variance, such a fit is difficult with a single or double Schechter 1976 function. We therefore use the raw list of luminosities or stellar masses for the abundance matching. We solve equation \([1]\) or equation \([2]\) by performing linear interpolation (in log-space) of the cumulative halo mass function.

3 Note that Yang et al.\(^{[2]}\) use virial radii corresponding to mean densities equal to 180 times the mass density of the Universe, which for their assumed \( \Omega_m = 0.238 \) corresponds to 43 times the critical density of the Universe. In other words, for typical NFW density profiles, the virial radius used by Yang et al. is roughly double the radius \( r_{200} \) (critical) used here.

\[
d_{\text{ang}}(z) = \frac{c}{1+z} \int \frac{dz'}{H(z')},
\]

is the cosmological angular distance (for a flat Universe), and \( z_{\text{group}} \) is the redshift of the central group galaxy.\(^{[4]}\)

\([5]\) **Membership probability**

In MAGGIE, galaxies are not assigned to groups, but are provided with probabilities that they belong to a given group, i.e. to the virial sphere of the real-space group. The probability that a galaxy lies within the virial sphere of the real space group is necessarily zero if the galaxy is outside the virial cone (circumscribing the virial sphere). Inside, the virial cone, the probability is obtained by comparing the predicted densities in PPS of the halo members (galaxies within the virial sphere) and the interlopers (galaxies within the virial cone, but outside the virial sphere). This can be written

\[
p(R, v) = \left\{ \begin{array}{ll}
g_h(R, v) & R \leq r_{200} \\
g_i(R, v) & R > r_{200}
\end{array} \right.
\]

where \( g_h \) and \( g_i \) are the densities in PPS of the halo members and interlopers, respectively. In practice, since the computation of the first expression of equation \([8]\) is limited to galaxies within the virial cone, there are few galaxy distances to compute around each group centre.

\([5a]\) **Halo density in projected phase space**

Given a galaxy number density profile \( \nu(r) \), the density of halo particles in PPS is (following Mamon et al.\(^{[3]}\) replacing infinities by the virial radius):

\[
g_h(R, v) = \Sigma_{\text{sph}}(R) \langle h(v|R, r) \rangle_{\text{LOS-sph}},
\]

where \( \Sigma_{\text{sph}} \) is the surface density of the galaxies limited to the virial sphere:

\[
\Sigma_{\text{sph}}(R) = 2 \int_R^{r_{200}} \nu(r) \frac{r \, dr}{\sqrt{r^2 - R^2}},
\]

while \( h(v|R, r) \) is the probability of having a LOS velocity \( v \) at the position in space given by \( (R, r) \), or when taking a LOS coordinate whose origin is at the group centre, at position \( (R_c, z_c) = \sqrt{r^2 - R^2} \). Combining equations \([9]\) and \([10]\) one obtains

\[
g_h(R, v) = 2 \int_R^{r_{200}} \nu(r) h(v|R, r) \frac{r \, dr}{\sqrt{r^2 - R^2}}.
\]

Assuming Gaussian (Maxwellian) three-dimensional velocities, Mamon et al.\(^{[3]}\) have shown that the LOS velocity distribution at position \( (R, r) \) is also a Gaussian:

\[
h(v|R, r) = \frac{1}{\sqrt{2\pi \sigma^2_v(R, r)}} \exp \left[ -\frac{v^2}{2 \sigma^2_v(R, r)} \right],
\]

with

\[
\sigma^2_v(R, r) = \left( 1 - \beta(r) \frac{R^2}{r^2} \right) \sigma^2_v(R, r),
\]

4 In this article, all instances of the symbol \( v \) represent line-of-sight velocities of galaxies relative to the group.

5 It is easy to improve this model using the joint q-Gaussian (Tsallis) velocity dispersion that Bernal et al.\(^{[3]}\) found to represent better the 3D velocity distribution in \( \Lambda \)CDM haloes.
where \( \beta = 1 - \sigma_b^2/\sigma_v^2 \) is the velocity anisotropy (for radial velocity dispersion \( \sigma_r \), and one component of the tangential velocity dispersion \( \sigma_\theta \)). In the presence of measurement errors of the LOS velocity, assumed Gaussian with zero bias and standard deviation \( \epsilon(v) \), the new distribution of LOS velocities is the convolution of the zero-error \( h(v|R,r) \) of equation (12) by a Gaussian of standard deviation \( \epsilon(v) \). Then, in the expression of \( h(v|R,r) \) (eq. (12)), the variance \( \sigma_v^2 \) in equation (13) is replaced by \( \sigma_v^2(R,r) + \epsilon^2(v) \). The radial velocity variance \( \sigma_v^2 \) in equation (13) is obtained (eq. (A1) in Appendix A) from the stationary spherical Jeans equation of local dynamical equilibrium

\[
\frac{d(\sigma_v^2)}{dr} + 2\beta(r) \frac{\nu \sigma_v^2}{r} = -\nu \frac{GM(r)}{r^2},
\]

where \( M(r) \) is our chosen total mass profile.

We assume that the galaxy distribution follows the mass distribution, and assume an NFW model for these two quantities. Denoting \( a \) the scale radius of the NFW density profile

\[ \nu_{\text{NFW}}(r) \propto \frac{1}{r(r+a)^2}, \]

(in the NFW model, \( a \) happens to be equal to the radius where the logarithmic slope of the density profile is equal to \(-2\)), we define the concentration parameter \( c_{200} = r_{200}/a \). We adopt the scaling between \( r_{200} \) and \( M_{200} = M(r_{200}) \) from the measurements on \( \Lambda \)CDM haloes at \( z = 0 \) by Macciò et al. (2008).

The NFW density profile can then be written

\[ \nu(r) = \frac{N_{200}}{4\pi r_{200}^3} \left( \frac{r}{r_{200}} \right), \]

\[ \hat{\nu}(x) = \frac{1}{\ln(c_{200} + 1) - c_{200}/(c_{200} + 1)} \frac{x^{-1}}{(x+1)^2}, \]

where \( N_{200} \) is the number of predicted galaxies (above some minimum luminosity or stellar mass) within the virial sphere. We shall see, below, that the normalization \( N_{200} \) cancels from equation (8). The mass profile of the groups is

\[ M(r) = M_{200} \ln(x + 1) - x/(x+1) \ln(c_{200} + 1) - c_{200}/(c_{200} + 1), \]

where \( x = r/r_{200} \).

Finally, we adopt the velocity anisotropy profile that Mamon & Łokas (2005) found to represent well the particles in clusters of \( \Lambda \)CDM haloes

\[ \beta(r) = \frac{1}{2} \frac{r}{r_\beta}, \]

with \( r_\beta = r_{200}/c \) (Mamon et al. 2010).

For our choice of NFW mass model and Mamon & Łokas anisotropy model, the radial velocity variance is given in equation (A2) of Appendix A.

(5b) Interloper surface density in projected phase space

Analyzing the distribution of dark matter particles within a hydrodynamical cosmological \( N \)-body simulation, Mamon et al. (2010) have found that the distribution of interlopers in PPS can be written as a Gaussian of the LOS velocity plus a constant term, where the coefficients of the Gaussian depend on projected radius:

\[ g_i(R, v) = \frac{N_{200}}{r_{200}^2 v_{200}} \hat{g}_i \left( \frac{R}{r_{200}}, \frac{v}{v_{200}} \right), \]

\[ \hat{g}_i(X, u) = A(X)\exp \left[ -\frac{1}{2} \frac{u^2}{\sigma_i^2(X)} \right] + B, \]

where

\[ A(X) = \text{dex} \left( -1.061 + 0.364 X^2 - 0.580 X^4 + 0.533 X^6 \right), \]

\[ \sigma_i(X) = 0.612 - 0.0653 X^2, \]

\[ B = 0.0075, \]

where cosmic variance fluctuations are 0.11, 0.23 and 0.40 dex for \( A(X), \sigma_i(X), \) and \( B \), respectively (Mamon et al. 2010).

The velocity \( v_{200} \) is the circular velocity at \( r_{200}, \) i.e. \( v_{200} = 10 H(z) r_{200} \). In the presence of velocity measurement errors of dispersion \( \epsilon(v) \), one should replace \( \sigma_i^2 \) by \( \sigma_i^2 + \epsilon^2(v)/v_{200}^2 \). Equations (19) - (23) depend little on halo mass in the cluster mass regime (Mamon et al. 2010), and we assume here that these equations extend to group masses. Moreover, we assume that the radius \( r_{200} \) in these formulae is the radius where the mean overdensity is 200 times the critical density at the group redshift.

We note that the normalization \( N_{200} \) appears in both \( g_i \) and \( s_i \), so it cancels out of the probability \( p(R, v) \) of equation (6). In our scheme, central galaxies have \( R = 0 \) and \( v = 0 \), by definition, and we set to unity their probability of membership (since the NFW central surface density diverges). To avoid too much group fragmentation, we do not assign a galaxy as a potential central group galaxy if it has a probability \( p > p_{\text{cen}} \) of belonging to another group of greater central galaxy stellar mass (since we proceed with groups of decreasing central galaxy stellar masses). Here, \( p_{\text{cen}} \) is a free parameter of MAGGIE. If \( p_{\text{cen}} = 1 \), all galaxies can be group centres (case of maximum group fragmentation and no group merging). If \( p_{\text{cen}} = 0 \), no satellite galaxy of a massive group can be the centre of another one (no group fragmentation, but maximal group merging). In other words, with \( p_{\text{cen}} = 0 \), galaxies lying in the virial cone of a massive central galaxy, but far in the foreground/background, will be assigned membership probabilities to the group around this first galaxy, but will not be assigned membership probabilities to potential groups around potential central galaxies lying in the same virial cone. However, if the central galaxy of the first group was wrongly determined, then one can effectively have group fragmentation, even with \( p_{\text{cen}} = 0 \) (but this occurs very rarely). Our initial tests showed that the performance of MAGGIE was optimal for \( p_{\text{cen}} < 0.01 \), and we adopt \( p_{\text{cen}} = 0.001 \).

(6) Group global properties

The group global properties are obtained by using the galaxy membership probabilities as weights, i.e. group luminosities \( L_{\text{group}} \) and stellar masses \( m_{\text{group}} \) are obtained with

\[ L_{\text{group}} = \sum_i p(R_i, v_i) L_i, \]

\[ m_{\text{group}} = \sum_i p(R_i, v_i) m_i, \]

over all galaxies with \( p(R, v) \geq p_{\text{mem}} \), where \( p_{\text{mem}} \) is another free parameter of MAGGIE. If \( p_{\text{mem}} = 1 \), the total group luminosities and stellar masses will correspond to the values of the central group.
galaxies, while if \( p_{\text{mem}} = 0 \), all galaxies within the virial cone will be considered when computing the luminosities and stellar masses, even those that contribute a tiny probability. Clearly, there should be little difference between setting \( p_{\text{mem}} = 0.001 \) or \( p_{\text{mem}} = 0 \). But physically, galaxies with extremely low \( p_{\text{mem}} \) typically correspond to interlopers that are many group standard deviations in the foreground or background in projection and it makes little sense to keep them in the group. We thus choose to set \( p_{\text{mem}} = 0.001 \).

(7) **Loop convergence**

We return to step 1 waiting for convergence when the number of groups found on the current pass matches the numbers found in the previous 3 passes. While the number of groups evolves towards a fixed value, it does not converge after 20 passes, hence we stop the iteration after the twentieth pass.

Note that the central galaxy is in general the most luminous (**MAGGIE-L**) or the most massive in stars (**MAGGIE-M**). However, there are rare exceptions where a group may contain a galaxy that is more luminous or massive than its central, and yet is not the central of another previously found group, i.e. a group whose central is more luminous or massive.

### 2.2 Edge effects

Aside from all-sky surveys, galaxy surveys have edges on the sky. Moreover, all volume-limited subsamples of galaxy surveys (including all-sky) will have edges in redshift space. Galaxy groups lying too close to an edge may be truncated. The grouping algorithm may detect the truncated group without knowing how much of the group lies beyond the survey edge. There is therefore no simple recipe to handle survey edges.

For groups lying near a survey edge, following Yang et al. (2007) we generate 700 galaxies (Yang et al. use 200) following the NFW profile, using the halo concentration estimated by the halo mass from Macciò et al. (2008). Then, we project galaxies on the celestial sphere and we estimate the number of galaxies weighted by galaxy probabilities and then the total stellar mass and luminosity of the group by dividing by this fraction (see Yang et al. 2007). Admittedly, if a large group is centered just beyond the survey edge, only a small fraction of this group will intersect our survey mask, so we will underestimate its virial radius and mass.

### 3 TESTS OF MAGGIE ON MOCK CATALOGUES

We test MAGGIE using realistic mock, doubly complete in distance and luminosity, galaxy redshift catalogues, which we had previously used in Paper I to optimize the FoF linking lengths. The construction of the mock catalogue and the description of the tests are discussed in detail in Paper I, and are briefly recalled below.

---

**Table 1.** Doubly complete subsamples of the mock SDSS/Legacy survey.

| Subsample | \( z_{\text{min}} \) | \( z_{\text{max}} \) | \( M_{p}^{\text{min}} \) | \( L_{\text{min}} / L^{*} \) | Galaxies |
|-----------|-----------------|----------------|-----------------|------------------|---------|
| 2 Nearby  | 0.01            | 0.053          | \(-19.0\)        | 0.14             | 72 510  |
| 5 Distant | 0.01            | 0.102          | \(-20.5\)        | 0.56             | 213 546 |

---

3.1 **Mock galaxy sample**

We have constructed a mock galaxy catalogue corresponding to the extent on the sky and depth of the largest contiguous (2.2 sr) region of the primary (Legacy) spectroscopic sample of the SDSS. For this, we replicated the galaxy outputs at \( z = 0 \) generated from the Guo et al. (2011) semi-analytical model (SAM) of galaxy formation and evolution, which was run on the halo merger trees extracted from the Millennium-II N-body simulation, itself which had been run in a box of comoving size \( L_{\text{box}} = 100 \, h^{-1} \text{Mpc} \), with cosmological parameters \( \Omega_{m} = 0.25 \), \( \Omega_{\Lambda} = 0.75 \), \( H_{0} = 73 \) and \( \sigma_{8} = 0.9 \), and particle mass \( 1.1 \times 10^{7} M_{\odot} \). Haloes were identified by applying the Friends-of-Friends (FoF) technique to the real space particle data.

In the output of the Guo et al. (2011) SAM, each galaxy is associated to a halo, making it easy to compare the groups extracted from our algorithm to the real space groups. Guo et al. found that the \( z = 0 \) galaxy luminosity and stellar mass functions agree well with the corresponding observed functions, making their galaxy catalogue realistic and useful to test our algorithm on data similar to observations.

The maximal redshift spanned by the simulation box is approximately \( H_{0} L_{\text{max}} / c \approx 0.025 \). Simulating the SDSS survey requires a deeper sample (see Table 1). For this, we have juxtaposed several boxes of the galaxy catalogue, applying random translations and rotations in galaxies coordinates to avoid perspective effects. This produced a larger superbox composed of the replicas of the galaxies in the computation box. We placed the observer at the middle of one of the sides of the superbox (see Fig. 1 of Paper I). Redshifts of the galaxies were computed using velocities given in the galaxy catalogue and adding the Hubble flow to it (see Paper I).

Our mock survey had no holes caused by saturated stars or bad data. Nevertheless, we allowed for observational errors on galaxy luminosities and stellar masses. According to Appendix B1 the errors on galaxy stellar masses, determined by comparison of different stellar mass algorithms on hundreds of thousands of SDSS galaxies, are roughly 0.2 dex. This value is much more conservative than the value of 0.10 dex (Taylor et al. 2011), and 0.15 dex (Mendel et al. 2014), but consistent with the 95% confidence errors of 0.30 and 0.35 dex for blue and red galaxies, as deduced by Conroy et al. (2009). In Appendix B2 we estimate the errors on galaxy luminosities, taking into account errors in photometry and redshift, uncertainties on extinction corrections and k-corrections, and neglect of peculiar velocities. We find that the errors on galaxy luminosities are of order of 0.08 dex at our minimum redshift of \( z = 0.01 \) decreasing to 0.06 dex at our maximum redshift of \( z \approx 0.1 \). In our analysis, we have therefore generated Gaussian errors without bias and with dispersion of 0.2 dex for log stellar masses and 0.08 dex for log luminosities.

From our flux-limited mock galaxy survey, we constructed several subsamples that are doubly complete in distance and luminosity. We focus our results on the two subsamples shown in Table 1.

---

\[7\] Groups lying very close to the virial cone also have very low membership probabilities.

\[8\] The number of groups oscillates around a value, but in an aperiodic fashion.
3.2 Flags

We flagged all galaxies belonging to real-space FoF groups containing at least one member that was on the other side of the periodic box (their groups would thus be split by the transformations of the box).

We also flagged the galaxies in the extracted groups that lie close to the redshift space edges of the doubly complete subsamples: to be very conservative, we flagged all extracted groups lying closer (roughly 2.5 Mpc) to the angular edges than would be the virial radius of a massive ($\log_{10} M = 15.2$) cluster, and all groups lying closer to the redshift limits than 13 times this distance (see Mamon et al. 2010) to account for redshift distortions.

We run MAGGIE on all galaxies of the mock (flagged or unflagged), and subsequently flag the groups that contain at least one flagged galaxy with $p > p_{\text{mem}}$.

3.3 Testing procedures

Following Paper I, we applied a suite of tests to groups containing at least one member that was on the other side of the periodic box (their groups would thus be split by the transformations of the box). We flagged all galaxies belonging to real-space FoF groups containing at least one member that was on the other side of the periodic box.

We also flagged the galaxies in the extracted groups that lie close to the redshift space edges of the doubly complete subsamples: to be very conservative, we flagged all extracted groups lying closer (roughly 2.5 Mpc) to the angular edges than would be the virial radius of a massive ($\log_{10} M = 15.2$) cluster, and all groups lying closer to the redshift limits than 13 times this distance (see Mamon et al. 2010) to account for redshift distortions.

We refer the reader to fig. 3 of Paper I for more details.

In an optimal grouping algorithm, the TGs minimally suffer from fragmentation into several EGs. A fragmented TG contains the central galaxy (see beginning of Sect. 2.1) of several EGs (with different membership probabilities in each, all with $p > p_{\text{mem}}$). Following Yang et al. (2007) and Paper I, the EGs and TGs are linked by their respective central galaxies. When fragmentation occurs, the primary EG is that containing the central galaxy of the parent TG. When merging occurs, the primary TG is that containing the central galaxy of the EG. We refer the reader to fig. 3 of Paper I for illustrations of group fragmentation and merging.

Also, in the optimal grouping algorithm, the galaxies of the EG should represent a maximally complete sample of the parent TG galaxies, and a maximally reliable (pure) sample, i.e. with as high as possible fraction of galaxies that belong to the parent TG (recall that the TG is the set of galaxies within the virial sphere).

Finally, the optimal grouping algorithm will produce EG luminosities, stellar masses and total masses as close as possible to those of the parent TG, i.e. with minimal bias and scatter. While bias can be corrected for, a measurement with strong scatter will be inefficient.

When TGs are fragmented, it makes little sense to measure the reliability of the galaxy membership of the secondary EGs (secondary fragments), and when TGs are merged, it would similarly not be useful to measure the completeness of the galaxy membership of a secondary TG. And it only makes sense to compare EG properties with the corresponding TG ones for primary fragments or relative to primary parent TGs. So all measures of completeness, reliability, as well as bias and scatter of group luminosity, stellar and total masses are limited to the primary EGs. The reader is referred to Paper I for more details.

Since the galaxy membership of MAGGIE groups is probabilistic, some of the statistical tests must be modified. In Paper I, we defined the galaxy reliability as

$$R = \frac{\#\text{EG} \cap \text{TG}}{\#\text{EG}},$$

where we adopt the notation $N_{i \in E}$ to represent the number of elements in space $E$. For our probabilistic MAGGIE group finder, we modify equation (26) to

$$R = \frac{\#\text{EG} \cap \text{TG}}{\#\text{TG}} = \frac{\sum_{i \in \text{EG} \cap \text{TG}} p_i}{\sum_{i \in \text{TG}} p_i},$$

where $p_i \equiv p(R, v_i)$ is the probability of membership of galaxy $i$ (eq. 8). The equivalent of equation (27) for the completeness would be

$$C = \frac{\#\text{EG} \cap \text{TG}}{\#\text{TG}} = \frac{\sum_{i \in \text{EG} \cap \text{TG}} p_i}{N_{i \in \text{TG}}}.$$

However, it is inconsistent to consider probabilities in the numerator of equation (28) and not in its denominator. We therefore adopt instead a definition based on hard assignments:

$$C = \frac{\#\text{EG} \cap \text{TG}}{\#\text{TG}} = \frac{N_{i \in \text{TG} \cap \text{EG}} \cap p_i > p_{\text{mem}}}{N_{i \in \text{TG}}}.$$

Since our chosen value of $p_{\text{mem}}$ is very small, the definition of completeness in equation (29) is very close to the definition of paper I for group total luminosities and stellar masses, we use the probabilities as in equations (23) and (25).

Finally, we did not measure group merging in this work. The total luminosities and stellar masses of the EGs are expected to follow the same trends as the TGs, with a possible enhancement at some point. However, the number of groups and the number of simulations are limited, making it difficult to draw firm conclusions.

4 RESULTS

We present the results of our tests on group fragmentation, galaxy completeness and reliability, accuracies of group total masses, total luminosities and total stellar masses. We ran these tests on both MAGGIE-L and MAGGIE-M, as well as on the FoF algorithm with the dimensionless linking lengths of $b_\perp = 0.07$ and $b_\parallel = 1.1$, which, in Paper I, we had determined to be optimal for studies of environmental effects on galaxies. These linking lengths are close to the values $b_\perp = 0.06, b_\parallel = 1.08$ optimized by Robotham et al. (2011).

4.1 Fragmentation

Figure 1 displays the fraction of extracted groups (EGs) that are secondary fragments as a function of estimated group mass. While

Note that our reliability (or ‘purity’, which can take values in the range $[0,1]$) is different from the purity used by Yang et al. (2007) and elsewhere, and also different to one minus their contamination. We feel that their purity, defined as $\text{TG}/\text{EG}/R/C$, and their contamination, defined as $(\text{EG}/\text{TG})/\text{EG}/C(1/R-1)$, which both can be greater or smaller than unity, are not physically meaningful, hence our preference for the reliability.
the nominally optimal FoF groups (blue) show a high (typically 35%) fraction of secondary fragments, both versions of MAGGIE lead to much lower fragmentation (typically 15%), even when realistic errors on galaxy luminosities and stellar masses are considered. The fragmentation by FoF and MAGGIE are both fairly independent of the chosen doubly complete subsample. The high fragmentation for the low and intermediate estimated mass bins was also found in Paper I (see the point labelled ‘R’ that corresponds to the linking lengths of Robotham et al. 2011 close to the optimal linking lengths used here).

4.2 Completeness and reliability

Figures 2 and 3 show that the EGs from MAGGIE-M (dark green lines) and MAGGIE-L (red lines) that are primary fragments are much more complete in galaxies than those from the FoF algorithm: for the nearby subsample (Fig. 2), in the worst performing among MAGGIE-M and MAGGIE-L, 100% completeness is achieved for 82%, 86%, and 78% of the groups, for the low, intermediate and high mass bins, respectively, and 90% completeness is reached for 82%, 92%, and 95% of the groups in the same respective mass bins. In comparison, only 73%, 55%, and 28% of FoF groups achieve just 90% completeness in these respective mass bins. Very similar trends are found for the distant subsample (Fig. 3).

The values of completeness for the FoF algorithm can be compared to fig. 5 of Paper I, which displays, for the same nearby subsample as used here ($M_r < -19$) with nearly the same optimal linking lengths (from Robotham et al. 2011) mean values of completeness of 90%, 80% and 72%,

---

**Figure 1.** Fraction of extracted groups that are secondary fragments as a function of their estimated group mass, for unflagged groups of at least 3 members (for both the extracted and true groups). The blue lines represent the optimal FoF ($b_\perp = 0.07$ and $b_\parallel = 1.1$), while the other coloured lines show the results for MAGGIE: the dark green and red lines show the results for MAGGIE-M and MAGGIE-L, with respective observational errors of 0.2 dex on stellar mass and 0.08 dex on luminosity, while the light green and orange lines show the results for MAGGIE-M and MAGGIE-L, with zero observational errors. The error bars are computed with the Wilson (1927) formula (see Wikipedia entry on Binomial proportion confidence interval). The points have their abscissa slightly shifted for clarity.

**Figure 2.** Cumulative distribution functions of the galaxy membership completeness (left, computed with eq. [29] for MAGGIE) and reliability (right, respectively computed with eqs. [26] and [27] for FoF and MAGGIE) in bins of estimated group masses, for the nearby subsample (unflagged galaxies in groups of at least 3 members that are not secondary fragments). The colours are the same as in Figure 1.

**Figure 3.** Same as Figure 2 but for the distant subsample.
While the completeness of the galaxy extraction with MAGGIE is fairly insensitive to the estimated group mass, the FoF algorithm produces increasingly incomplete galaxy extractions for increasing estimated group masses (see above). The lower galaxy completeness of FoF at high estimated group masses is a consequence of the higher fragmentation of high mass real-space groups (see Paper I). If a TG has non-negligible secondary fragments, then its primary fragment will tend to be incomplete. Consider, for example, a TG with 5 galaxies that is fragmented into an EG of 3 galaxies (containing the TG’s central) and another EG of 2 galaxies; the EG of 3 galaxies will have a completeness of 3/5 = 0.6, a reliability of unity, and will not be counted as a secondary fragment, while the EG of 2 galaxies will be considered a secondary fragment, but will not have completeness and reliability measured.

Figures 2 and 3 also show that, regardless of the inclusion of observational errors, the reliability (purity) of the galaxy membership of primary fragments is much higher for both MAGGIE-M and MAGGIE-L in comparison with FoF for low estimated group masses, and higher for intermediate group masses, while FoF produces higher reliabilities for high group masses. The galaxy reliability decreases with increasing EG mass, especially for MAGGIE: for the FoF method, the fractions of groups with 90% reliability are (for the distant subsample, which has better statistics) respectively 70%, 50% and 32%, for the three bins of estimated EG mass, in increasing order. For the worst performing among MAGGIE-M and MAGGIE-L with observational errors, the fractions of 90%-reliable groups are respectively 90%, 58% and 20%. The median galaxy reliabilities for the high-mass bin are more similar: while they are 83% and 84% for the nearby and distant subsamples with FoF, they are respectively 81% and 92% (75% and 82%) with MAGGIE-L (MAGGIE-M).

4.3 Accuracy in group total masses

Figure 4 shows how the estimated total masses (of the EGs) compare with the total masses (within the virial sphere) of the TGs (for clarity, we hereafter drop the term ‘total’ before ‘mass’ in this subsection). The FoF method (with the virial theorem to estimate masses) leads to frequent strong underestimation of the mass for low TG masses. This is analogous to what is found by most group finders (e.g., Old et al. 2014, although Old et al. 2015 find that a Bayesian fitted slope of the estimated versus true mass relation is typically unity). Figure 4 shows that MAGGIE does not underestimate the EG masses as frequently as does the FoF algorithm. This better behaviour is likely to be the result of the use of the median, rather than the FoF algorithm. The left panels of Figure 4 show the mass fragmentation of FoF EGs is biased low by a factor of 10 at the lowest estimated masses and by about 0.1 dex at high estimated masses. Similar trends of strong mass underestimation with FoF were found in Paper I for low-richness EGs. In contrast, both flavours of MAGGIE, without or with errors, have lower or equal absolute bias at all estimated group masses: the bias is never more than 0.1 dex in absolute value for groups with 

\[
\frac{\log_{10} M_{\text{est}}}{M_{\text{true}}} > 12.8.
\]

Since bias can, in principle, be corrected for, it is more important to consider the scatter in the mass estimation. In both doubly complete subsamples, MAGGIE (even with observational errors) shows lower scatter than FoF at all estimated group masses, except 

\[
\log_{10} M_{\text{est}} > 14.7
\]

for the distant subsample. While none of these differences in the 7 considered mass bins is statistically significant, the fact that MAGGIE has lower scatter than FoF in 6 or 7 mass bins out of 7 is highly significant. For 

\[
\log_{10} M_{\text{est}} \leq 14,
\]

the FoF mass errors tend to be of order 0.33 dex, while the mass errors with MAGGIE are almost always equal to or less than 0.25 dex, down to less than 0.2 dex for the highest mass bin.

4.4 Accuracy in group luminosities and stellar masses

Figures 5 and 7 show the bias and inefficiency of the recovered group luminosities and stellar masses, respectively (using eqs. 24 and 25). The FoF groups with 

\[
\log_{10} M_{\text{est}}/M_{\odot} < 13
\]

have zero median bias, as expected since the median completeness and reliability of the primary EG FoF groups are both unity at low mass, i.e. the galaxy membership is perfect more than half the time. On the other hand, the MAGGIE groups with low estimated total masses are slightly biased low, by up to 0.03 dex (7%) at 

\[
\log_{10} M_{\text{est}}/M_{\odot} = 12.
\]

Although these MAGGIE groups also have median completeness and reliability of unity, the argument applied above for the FoF groups cannot apply for the MAGGIE groups, since the latter use a probabilistic definition of total group luminosity (eq. 23) and stellar mass (eq. 25). At high estimated group masses, the bias is less well measured (given the lower number of groups in the higher mass bins), and some fluctuations appear: in the nearby subsample, MAGGIE-M luminosities and stellar masses are respectively biased by 0.10 ± 0.03 and 0.08 ± 0.02 dex at 

\[
\log_{10} M_{\text{est}}/M_{\odot} = 14.4,
\]

while...
Figure 4. Estimated mass versus true mass for the unflagged non-secondary (filled coloured circles) and secondary (black crosses) EGs with at least 3 members for five group finders (from left to right): Friends-of-Friends, MAGGIE-M with no observational errors on stellar mass, MAGGIE-M with 0.2 dex errors on stellar mass, and MAGGIE-L with 0.08 dex errors on luminosity. The diagonal lines indicate perfect mass recovery. There are roughly as many MAGGIE groups as there are FoF groups, but the former occupy identical positions in the plots due to the replications of the simulation boxes causing identical stellar masses, hence identical group masses after the abundance matching used to infer group masses.

FoF luminosities and stellar masses are biased by $-0.3 \pm 0.05$ dex at $\log M_{\text{est}}/M_\odot = 14.8$ for the distant subsample.

The inaccuracies in total group luminosity or stellar mass estimation with both flavours of MAGGIE are smaller than with FoF at all masses (FoF inefficiencies are lower for $\log M_{\text{est}}/M_\odot = 14.8$ of the distant subsample, but consistent with those of MAGGIE within the errors). For example, at intermediate estimated group masses, the scatter in luminosities and stellar masses estimated by MAGGIE-M are systematically 0.05 or 0.03 dex lower than when estimated with FoF, for the nearby and distant subsamples, respectively.

5 DISCUSSION

5.1 General features of MAGGIE

Thanks to its probabilistic nature, MAGGIE generally performs well with galaxy membership reliability, since the least reliable galaxies along the LOS are assigned low probabilities (FoF shows better reliability in the high estimated mass bin, which will discuss below). In general, extracted global properties of groups are less biased and more reliable. This is also the case for FoF in a more indirect manner (through the linking lengths). The bloated sizes of high-mass groups, as witnessed by the mass bias going from negative to positive for MAGGIE groups (left panels of Fig. 5), explains the strong decrease in reliability with increasing estimated group mass (Figs. 2 and 3), especially in MAGGIE. In Paper I, we found that for the near-optimal FoF (Robotham et al. 2011), group merging increased from the lowest to highest estimated mass bin from 10% to 75% for the nearby subsample, and from 5% to 25% for the distant one. In turn, group merging is caused by overestimated group masses, and indeed, an increasingly larger fraction of FoF groups are biased high as shown by the upper (blue) error bars in top panel (nearby subsample) of Figure 5.

Our tests have been performed in an idealized situation for MAGGIE, with perfectly known scaling relations, yet with realistic measurement errors on galaxy luminosities and stellar masses. We discuss the effects of the observational errors in Sect. 5.2 and the choice of initial mass – luminosity scaling relation in Sect. 5.3.1.

5.2 Robustness of MAGGIE to observational errors

We have run MAGGIE both with and without the observational errors on galaxy luminosities and stellar masses. Figures 4 and 5 show the effects of going from 0 (light green) to 0.2 dex (dark green) errors in stellar mass for MAGGIE-M. Including 0.2 dex errors on stellar masses produces negligible extra group fragmentation (Fig. 1) and only slightly lower galaxy completeness and reliability, with the greatest effects seen at high group masses in the nearby subsample (lower panels of Fig. 2), where the median galaxy reliability decreases from 80% (no errors) to 73% (0.2 dex errors). The effects of 0.2 dex errors are stronger on the accuracy of total group masses (Fig. 5): increasing the bias (whether positive or negative) by a factor 1.5, and, especially, increasing the scatter by roughly 0.08 dex at all group masses. The effects of 0.2 dex observational errors on stellar masses are negligible on group luminosities (Fig. 6) and stellar masses (Fig. 7).

The effects of 0.08 dex luminosity errors on MAGGIE-L are even smaller, as expected. Moving from no errors (orange) to 0.08 dex errors (red) on luminosities has a negligible effect for group fragmentation (Fig. 1), for the bias in group total masses (Fig. 5), and for the accuracy of group luminosities (Fig. 6) and stellar masses (Fig. 7).
In summary, while geometric based algorithms such as FoF or Voronoi-Delaunay methods are immune to observational errors on luminosities and stellar masses, such errors will affect group finders that derive total group masses by abundance matching with group luminosities or stellar masses, as those of Yang et al. (2007), Muñoz-Cuartas & Müller (2012), Domínguez Romero et al. (2012) and MAGGIE. Our tests indicate that the effects of observational errors are small on galaxy completeness and reliability and negligible for group fragmentation, and the accuracy of group luminosities and stellar masses. However, including errors of 0.2 dex on stellar masses for MAGGIE-M and 0.08 dex on luminosities for MAGGIE-L leads to an increased scatter in total group masses by 0.08 and 0.03 dex, respectively.

Yet, once the observational errors are included, MAGGIE-M performs as well as MAGGIE-L on the scatter in total group masses. As can be seen in Figure 5, MAGGIE-M produces less scatter than MAGGIE-L when the observational errors are not included (light green and orange symbols). This result is expected at high group (halo) total mass (where it appears to be only marginally significant). Indeed, the increasing preponderance of the Red Sequence and its slope combine to increase the stellar mass-to-light ratio with increasing stellar or group mass, hence the halo (total group) mass varies faster with group luminosity than with group stellar mass, leading to more scattered total masses with MAGGIE-L. At lower group total masses, Figure 5 indicates an even greater superiority of MAGGIE-M relative to MAGGIE-L. This may be caused by the increased importance of the Blue Cloud at low group masses (Weinmann et al. 2006), and its much wider scatter in colours is linked to a much wider scatter in stellar mass-to-light ratios, which adds to the scatter in stellar to total masses.

5.3 Robustness of MAGGIE to details of the model

5.3.1 Initial halo mass – central stellar mass relation

We tested how MAGGIE is affected by our initial relation between halo mass and central stellar mass (item 2a of Sect. 2.1). We found that MAGGIE is insensitive to our adopted scheme of relating luminosity or stellar mass to halo mass in its first pass: the final variation of group $M/L_r$ vs. $L_r$ is precisely the same whether one adopts $M/L_r = 300$ or the relation of $M/L_r$ vs. $L_r$ that Behroozi et al. (2010) derived from abundance matching. The same effect has been previously noticed by Yang et al. (2007).

5.3.2 Halo mass function model

The estimation of the virial mass (or virial radius) is a crucial step (item 2 in Sect. 2.1) of MAGGIE (and of other methods that use priors such as Yang et al. 2007 and Domínguez Romero et al. 2012). A biased estimate of group masses will affect the observed trends of galaxy properties with the global environment. The abundance matching technique, used in MAGGIE (as well as by Yang et al., Muñoz-Cuartas & Müller 2012 and Domínguez Romero et al.) appears to be a good way to estimate the virial mass of galaxy group haloes. There are, however, three issues that need to be considered.

First, there may be haloes with no galaxies that may perturb the halo-group bijectivity assumption of abundance matching. We checked that no haloes above $10^{11} \, M_{\odot}$ in the Millennium-II simulation have zero galaxies assigned to them in the SAM of Guo et al. (2011).

Second, deriving group total masses by abundance matching between a halo mass function and the inferred distribution of galaxy group luminosities or stellar masses should cause inefficient estimation of group masses when these are in the high range (14 < log$_{10}$ $M_{\text{halo}}$/$M_{\odot}$ < 15), because of the lower slope of the high mass end of the group luminosity (or stellar mass) as a function of halo mass at high halo mass (e.g., Yang et al. 2008, 2009).

Third, most analytical halo mass functions described in the literature are based on fits to the FoF mass of the haloes instead of the spherical over-density mass, which is how we defined the virial mass of the halo. Since we used the galaxy catalogue from Guo et al. (2011), whose semi-analytical code was applied onto the Millennium-II run, we fit the halo virial mass function directly on its output.

Figure 8 shows the cumulative halo mass function computed in various ways. The figure clearly shows that the cumulative FoF
halo mass function computed from the Millennium-II Simulation is typically 0.2 dex above the cumulative virial mass function computed from the same simulation. While the analytical approximation of Courtin et al. (2011) (cyan) matches very well the halo FoF mass function, Figure 8 shows that the cumulative halo virial mass function obtained from the analytical approximation of Tinker et al. (2008) (purple) is slightly offset, at low masses, relative to cumulative halo virial mass function extracted from the Millennium-II simulation.

We therefore chose, in our present tests, to fit ourselves the halo virial mass function of the Millennium-II Simulation with the Tinker et al. (2008) model. Our maximum likelihood fit (green curve) produces Tinker et al. parameters $a = 2.13$, $b = 1.97$, and $c = 1.75$, with normalization $A = 0.188$ (instead of the corresponding values of $a = 1.87$, $b = 1.59$, $c = 1.58$, and $A = 0.248$ that Tinker et al. found for $\Delta_m = 800$, i.e. $\Delta = 200$ given $\Omega_m = 0.25$ of the Millennium-II simulation).

Figures 9 and 10 show how the performance of MAGGIE is affected by the choice of halo mass function, respectively for the galaxy completeness and reliability and for the estimates of group total and stellar masses as well as total luminosities. Figure 9 shows that adopting the analytical approximation to the halo FoF mass function (cyan lines) leads to higher galaxy completeness, but lower galaxy reliability. Interestingly, despite the difference between the Tinker et al. (2008) fit (purple lines) and ours (green lines) to the halo virial mass function (Fig. 9), there are no distinguishable differences in the completeness and reliabilities of the galaxy memberships. This suggests that one can use the Tinker et al. analytical fits to the halo virial mass function for the abundance matching in MAGGIE.

Figure 10 indicates that, in comparison to the group total masses obtained by fits of the halo virial mass function, the masses obtained with the analytical approximation to the halo FoF mass function are biased high (up to 0.25 dex), leading to absolute mass biases of 0.35 dex at high masses. This is expected, given the higher FoF (monotonically decreasing) halo mass function (i.e. higher FoF halo masses at equal halo ranks, as seen in Fig. 8), while there is again no difference in mass bias between the two adopted fits to the halo virial mass function. The scatter in total mass is unaffected as are the accuracies of group luminosities and stellar masses.

The fragmentation is not shown but is also not affected by the choice of halo mass function.

5.3.3 Cosmological parameters

The distances used by group finders such as MAGGIE and FoF depend on the choice of cosmological parameters. For example, when computing the projected radius of a galaxy at the redshift of the group, we implicitly need to compute the cosmological angular distance, $d_A(z) = d_L(z)/(1 + z)^2$, hence the luminosity distance, $d_L(z)$, which is cosmology dependent. We assume a flat Universe, for which the luminosity distance is computed using elliptic integrals [Liu et al. 2011] see also Eisenstein (1997). Moreover, for all four analytical halo mass functions tested in Sect 5.5.2, we have assumed the same cosmological parameters as in our mock (based on those from the Millennium-II simulation). The observer may choose a slightly different set of cosmological parameters.

We tested the sensitivity of MAGGIE to the choice of cosmological parameters, by comparing the results of MAGGIE-M (with 0.2 dex errors) with the “true” cosmology ($\Omega_m = 0.25$, $h = 0.73$ as assumed in the Millennium-II simulation, on which our mock is based) with analogous runs of MAGGIE-M assuming instead two “false” cosmologies (i.e. inconsistent with our mock): WMAP9, with $\Omega = 0.279$, $h = 0.70$ (Bennett et al. 2013) and Planck with $\Omega_m = 0.315$, $h = 0.67$ (Planck Collaboration et al. 2014). As seen in Figures 11 and 12 the choice of cosmological parameters affects very little the performance of MAGGIE-M on galaxy completeness and reliability, and accuracy in total group mass, with no statistically significant trends with $\Omega_m$. The strongest effect is that the 90% galaxy completeness in the nearby subsample, at high estimated group masses, decreases from 95% to 86% when moving from the “true” cosmology to the “false” (WMAP9 or Planck) one.

This lack of sensitivity to cosmological parameters is expected, given the low maximal redshift of our subsamples ($z = 0.1$, see Table 1). One may expect MAGGIE to be more sensitive to the choice of cosmological parameters when applied to deeper galaxy surveys.
Figure 9. Effects of the choice of halo mass function on the cumulative distribution functions of galaxy completeness and reliability of the groups extracted with MAGGIE-M (with 0.2 dex observational errors). The lines show the three adopted halo mass functions: our fit to the halo virial mass function (green, as used throughout article), and the analytical halo mass functions of Tinker et al. (2008, virial purple) and Courtin et al. (2011, FoF, cyan). The Tinker et al. completeness and reliability cumulative distribution functions are usually indistinguishable from those obtained with our fit.

Figure 10. Effects of the choice of halo mass function on the accuracy of the properties of the groups extracted with MAGGIE-M (with 0.2 dex observational errors). The points have their abscissa slightly shifted for clarity. Same line colours as in Figure 9.
Figure 11. Effects of the choice of the cosmological density parameter (for flat universes) and dimensionless Hubble constant on the cumulative distribution functions of the galaxy completeness and reliability of the groups extracted with MAGGIE-M with 0.2 dex observational errors. The different lines represent $\Omega_m = 0.25, h = 0.73$ (Millennium-II, green, as used throughout this article), $\Omega_m = 0.279, h = 0.70$ (WMAP9, magenta), and $\Omega_m = 0.315, h = 0.67$ (Planck, royal blue).

Figure 12. Effects of the choice of the cosmological density parameter (for flat universes) and dimensionless Hubble constant on the accuracy of the properties of the groups extracted with MAGGIE-M (with 0.2 dex observational errors). The points have their abscissa slightly shifted for clarity. Same line colours as in Figure 11.
5.4 Comparison of MAGGIE to other group finders

5.4.1 Friends-of-Friends

Our tests often show different behaviours of the relative performances of FoF, MAGGIE-L and MAGGIE-M in the three bins of estimated mass (log_{10}M_{est}/M_☉ = 12–13, 13–14, and 14–15). For this reason we now compare the performances for each mass bin in turn.

At low estimated group masses (12 < log_{10}M_{est}/M_☉ < 13), in comparison with FoF, both flavours of MAGGIE lead to much less fragmentation, and for the non-secondary fragments, the galaxy membership is more complete and reliable, the total group masses are less biased and suffer less scatter. The inefficiency of group luminosity and stellar masses are roughly the same for FoF and MAGGIE, while FoF has zero median bias in group luminosity and stellar mass. There are only very small differences between MAGGIE-L and MAGGIE-M: MAGGIE-L is slightly more complete while MAGGIE-M is slightly more reliable, MAGGIE-M is slightly less biased in total mass with slightly less scatter.

At intermediate estimated group masses (13 < log_{10}M_{est}/M_☉ < 14), both flavours of MAGGIE are greatly superior to FoF in terms of fragmentation: the fraction of groups that are secondary fragments is 2.5 times greater with FoF in comparison to MAGGIE. For the non-secondary fragments, in comparison with FoF, the galaxy membership of MAGGIE is much more complete and more reliable, total masses are extracted with less bias and much less scatter, while group luminosities and stellar masses are extracted with roughly similar bias (slightly lower in FoF), and much less scatter again. MAGGIE-M and MAGGIE-L perform in similar ways on most tests, except that MAGGIE-M produces slightly more reliable galaxy membership and less scatter in group stellar masses, while MAGGIE-L is slightly less biased in total mass.

At large estimated group masses (14 < log_{10}M_{est}/M_☉ < 15), group fragmentation is much more (≈ 3x) severe in the distant subsample for the FoF groups relative to the MAGGIE groups (lower panel of Fig. 7). For the nearby subsample, the situation is less clear because of poor statistics. For the non-secondary fragments, while the galaxy membership of MAGGIE groups is highly complete (all groups have > 80% completeness), the galaxy membership of FoF groups is highly incomplete (80% median completeness). On the other hand, the FoF groups are somewhat more reliable than the MAGGIE groups (although the median galaxy reliabilities are very similar). However, FoF is more biased in group total mass, stellar mass and luminosity and slightly more inefficient, for all three quantities. The fraction of secondary fragments is smaller in MAGGIE-M than in MAGGIE-L, while for the non-secondary fragments, MAGGIE-L is more complete and slightly more reliable in the galaxy membership. Finally, both flavours of MAGGIE have similar accuracies on group total masses, stellar masses, and luminosities.

These tests point to the superior performance of MAGGIE in comparison with FoF, in all three mass bins considered, with a very slight edge for MAGGIE-M over MAGGIE-L thanks to lower group fragmentation in the former for log_{10}M_{est}/M_☉ ≥ 14.

One may ask whether the superior performance of MAGGIE-M relative to FoF is caused by its reliance on priors (e.g. its assumptions of reasonable NFW priors on group number and mass density profiles), the abundance matching used to derive group virial masses, or its probabilistic nature. We thus now compare MAGGIE to two group finders based upon priors and that use abundance matching to derive group sizes and masses, those of Yang et al. (2005, 2007), which are standard group finders, and that of Domínguez Romero et al. (2012) which is, in great part, probabilistic, but uses hard assignments to galaxies in the end. We also compare to the modified FoF algorithm of Muñoz-Cuartas & Müller (2012) in which the masses are obtained by abundance matching. Unfortunately, only a small subset of our tests have been carried out by these three authors.

5.4.2 Yang et al.

Only two tests allow for comparison of Yang et al. group finder with MAGGIE, galaxy completeness and mass accuracy. Yang et al. (2007) study the completeness of the galaxy membership for the primary fragments (which they define as the most massive ones). Their figure 2 indicates that 90% galaxy completeness is reached in 97%, 94%, 83%, 72% and 68% of their groups in mass bins log M_{true}/M_☉ = 12.5 – 13, 13 – 13.5, 13.5 – 14, 14 – 14.5, and 14.5–15. In comparison, Figure 13 shows that the fractions of groups with at least 90% galaxy completeness are 97% and 100% (depending on subsample), 82% and 92%, and 94% and 91% for MAGGIE-M in the three mass bins log M_{est}/M_☉ = 12 – 13, 13 – 14, and 14–15, respectively. The corresponding numbers for MAGGIE-L are 93% and 99%, 78% and 93%, and 97% and 100%. Therefore, both flavours of MAGGIE lead to comparable galaxy completeness at low and intermediate TG masses, while the galaxy completeness is much greater in MAGGIE compared to Yang et al. at high TG masses.

The scatter in group total mass of the Yang et al. algorithm is shown in figure 7 of Yang et al. (2007). Considering the case where no correction for luminosity incompleteness is required (their panel c), one finds that the scatter divided by √2 is σ_σ ≈ 0.23 for group estimated log masses in the range 12 to 14.4 (solar units), leading

© 2013 RAS, MNRAS 000 0

Figure 13. Cumulative distribution functions of the galaxy membership completeness in three bins of true group mass for the nearby (left) and distant (right) subsamples. The colours are the same as in Figure 7. The completeness of all methods for the low true group masses in the distant subsample are 100%.
to a scatter of $\sqrt{2} \times 0.23 = 0.33 \text{[15]}$. Figure 14 overplots the values of their scatter with those that we found for FoF and MAGGIE. In both subsamples, the Yang et al. scatter is higher than that of both flavours of MAGGIE at all estimated group masses, and even slightly worse (nearby subsample) or comparable (distant subsample) to the scatter of the FoF masses.

In summary, in comparison to the Yang et al. group finder, MAGGIE has comparable (low and intermediate group mass) or much higher (high group mass) galaxy completeness, and much lower inefficiency (scatter) in total mass estimation at all group masses.

5.4.3 Domínguez Romero et al.

Similarly, we can compare the performance of MAGGIE with the other similar group finder by Domínguez Romero et al. (2012), for which two tests can be compared: fragmentation and galaxy completeness. In the right panel of their figure 1, they obtain zero median group fragmentation at all TG masses, while we found typically 15% of group fragmentation in bins of estimated group mass. As we discussed in Paper I, the statistics of group fragmentation are very different in terms of EG mass in comparison with those in terms of TG mass: fragmentation occurs in the highest mass TGs, but the secondary fragments contaminate most the low estimated mass groups. However, Domínguez Romero et al. restricted their secondary fragments to those accounting for at least 10% of the TG mass, and defined their primary fragments as the most massive, while our definition of primary is the fragment containing the central galaxy.

Figure 14 shows that a small minority of our secondary fragments are more massive than the TG, so to make a clean comparison with Domínguez Romero et al., we show in Figure 15 the mean number of secondary fragments with mass between one-tenth and one times the TG mass. Fragmentation worsens with increasing TG mass, as we had found in Paper I for FoF. But Figure 15 also indicates that, using the measure of group fragmentation of Domínguez Romero et al., MAGGIE (as FoF) is unable to match the zero mean number of secondary fragments per TG that Domínguez Romero et al. found. This discrepancy would be even stronger had we used errors on the means instead of standard deviations for the points of Domínguez Romero et al.

Moving on to galaxy completeness, the left panel of figure 2 of Domínguez Romero et al. indicates median galaxy completeness of 96%, 91% and 92% for the three mass bins we used here (but for TG masses). In contrast, according to Figure 14 both flavours of MAGGIE lead to median galaxy completeness of 100% in all three of our mass bins.

In summary, MAGGIE cannot match the zero group fragmentation of Domínguez Romero et al. (2012), while they cannot match the 100% median galaxy completeness of MAGGIE. Now group fragmentation is caused by either negative mass bias producing too narrow virial cones, or by too restrictive group membership along the LOS. Since MAGGIE has higher galaxy completeness than Domínguez Romero et al. one must conclude that the value

---

**Figure 14.** Scatter of extracted total group mass within the virial sphere (for unflagged primary groups of 3 or more galaxies), for FoF (blue), MAGGIE-M (green), and MAGGIE-L (red), all using 16th and 84th percentiles, and for Yang et al. (2007, black, based on standard deviation), multiplying the values of their Figure 7c (bottom) by $\sqrt{2}$. The upper and lower panels correspond to the nearby and distant subsamples, respectively, with the Yang et al. values the same in both. The error bars indicate errors on means (MAGGIE and FoF) or standard deviations (Domínguez Romero et al.). The points have their abscissa slightly shifted for clarity.

**Figure 15.** Mean number of secondary fragments (estimated mass between 0.1 and 1 times the true group mass) per true group as a function of true group mass for FoF (blue), MAGGIE-M (with 0.2 dex errors, green), and MAGGIE-L (with 0.08 dex errors, red), and for Domínguez Romero et al. (2012, black). The upper and lower panels correspond to the nearby and distant subsamples, respectively, with the Domínguez Romero et al. values the same in both. The error bars indicate errors on means (MAGGIE and FoF) or standard deviations (Domínguez Romero et al.). The points have their abscissa slightly shifted for clarity.
of $p_{\text{cen}} = 0.001$ adopted here for MAGGIE, although conservative in its precise value, is sufficiently large to fragment TGs along the LOS. Alternatively, since Domínguez Romero et al. did not provide tests of mass bias and group merging, they may have overestimated their group sizes and masses and merged their groups.

5.4.4 Muñoz-Cuartas & Müller

Muñoz-Cuartas & Müller (2012) have tested their modified FoF approach using a mock constructed from a doubly complete subsample with $M_r < -19$, which matches our nearby doubly complete subsample. Only their galaxy completeness can be compared to ours. The fraction of their groups with over 90% completeness is 82%, while for the nearby subsample, our Figure 2 indicates that the fractions of groups with 90% galaxy completeness are 87%, 93%, and 96% for MAGGIE-M and 82%, 92%, and 100% for MAGGIE-L. These fractions suggest that the overall galaxy completeness (summing over the groups of all masses) is well over 87% (MAGGIE-M) or 82% (MAGGIE-L). Therefore, it seems that MAGGIE produces even more complete galaxy membership than the modified FoF algorithm of Muñoz-Cuartas & Müller. One must note that these authors considered groups of at least 2 members, while we have considered groups of at least 3 members.

5.4.5 Synthesis

As mentioned above, the superior performance of MAGGIE relative to the optimal FoF group finder can be explained by the 3 main improvements of MAGGIE relative to FoF: its use of realistic priors, its use of abundance matching to measure group masses and radii, and its probabilistic galaxy membership. The Yang et al. (2007) and Domínguez Romero et al. (2012) group finders also use abundance matching. Therefore the superior mass accuracy of MAGGIE, relative to Yang et al. suggests that the more realistic priors and probabilistic memberships improve the performance of the group finder.

6 CONCLUSIONS

We have introduced a new prior-based and fully probabilistic group finder called MAGGIE, where the total group/cluster masses are obtained by abundance matching between the assumed known halo mass function and the derived group luminosity (MAGGIE-L) or stellar mass (MAGGIE-M) function. This grouping algorithm is similar to that of Yang et al. (2005, 2007), but uses a more refined and probabilistic membership criterion, and is meant to be applied to subsamples that are complete in both luminosity and distance, to avoid the unavoidable luminosity incompleteness in flux limited samples, which are very difficult to accurately correct for.

We extensively tested both the FoF group finder with the optimal linking lengths derived by Duarte & Mamon (2014) and MAGGIE, using a mock SDSS Legacy spectroscopic survey derived from the Guo et al. (2011) semi-analytical model of galaxy formation, itself run on the Millennium-II cosmological dark matter simulation. We also compared MAGGIE with the similar group finders of Yang et al. (2007) and Domínguez Romero et al. (2012), as well as with the modified FoF of Muñoz-Cuartas & Müller (2012) in all instances where this could be done. We find that both flavours of MAGGIE perform as well or better than these four other group finders for all tests, with two exceptions: MAGGIE produces less reliable galaxy membership for high estimated mass groups than the optimal FoF, and it cannot match the zero mean fragmentation of true groups found by Domínguez Romero et al. In particular, MAGGIE suffers from much less group fragmentation than does the optimal FoF, despite our having incorporated 0.2 (0.08) dex errors on the stellar masses (luminosities) of galaxies. MAGGIE also performs better than FoF (Yang et al. and Domínguez Romero et al. in galaxy completeness at all masses (and than Muñoz-Cuartas & Müller averaged over all masses), better than FoF in galaxy reliability (except at high estimated group mass), better than both FoF and Yang et al. in the scatter the derived total group masses, and better than FoF in total group mass bias, as well as accuracy (bias and scatter) in group luminosity and stellar mass.

The strength of MAGGIE appears to be the consequence of its use of realistic priors, abundance matching and probabilistic galaxy membership.

This makes MAGGIE-M an ideal grouping algorithm to be applied on large galaxy spectroscopic surveys such as the Sloan Digital Sky Survey (SDSS) and the Galaxy And Mass Assembly (GAMA), for several applications, in particular the environmental effects on galaxy properties such as SSFR and mass/orbit modeling of groups and clusters (possibly stacking the groups). Moreover, MAGGIE should in principle be able to work for much deeper spectroscopic surveys, possibly including surveys based upon photometric redshifts (since MAGGIE naturally handles redshift errors), with applications to the evolution of environmental effects, dark matter properties (normalization, concentration), velocity anisotropy (orbital shapes).

In particular, MAGGIE should be very useful for dark energy surveys such as the Dark Energy Survey (DES), Euclid, and the Wide-Field Infrared Survey Telescope (WFIRST, yet to be approved) that will constrain dark energy parameters not only with cosmic shear and baryonic acoustic oscillations, but also by measuring the mass function and clustering of galaxy clusters. However, the abundance matching method – used to determine group masses – involves an assumption on the halo mass function, which is cosmology-dependent. This implies that the current implementation of MAGGIE cannot be used as a cosmographic tool to determine cosmological parameters from the derived halo mass function. However, MAGGIE is an excellent tool to optimally detect and measure groups and clusters in dark energy surveys, if a given cosmology is assumed. Moreover, by replacing abundance matching by other techniques, MAGGIE could be adapted into a powerful cosmographic tool for such surveys.

ACKNOWLEDGMENTS

Paper I and this article were part of the doctoral thesis of MD. We thank Radek Wojtak for encouraging us to avoid hard assignments of galaxies to their final groups in our probabilistic group finder, at a time when we were still hesitant on the approach. We also warmly thank Reinaldo de Carvalho for a thorough reading of an earlier draft with constructive criticisms, Xiaohu Yang and Mariano Domínguez Romero for explaining details of their respective group finders, as well as David Valls-Gabaud, Andrea Biviano, and Florence Durret for useful comments. The Millennium-II Simulation database used in this paper and the web application providing online access to them were constructed as part of the activities of the German Astrophysical Virtual Observatory (GA VO). We are grateful to Michael Boylan-Kolchin and Qi Guo for respectively allowing the outputs of the Millennium-II simulation and the Guo semi-analytical model to be available to the public, and Gerard Lemson.
for maintaining the GAVO database and for useful discussions. We also made use of HMFCalc [Murray et al., 2013] to check our computations of halo mass functions.

REFERENCES

Abazajian K. N. et al., 2009, ApJS, 182, 543
Abell G. O., 1958, ApJS, 3, 211
Beers T. C., Tonry J. L., 1986, ApJ, 300, 557
Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379
Bennett C. L. et al., 2013, ApJS, 208, 20
Bernaldo L. J., Mamon G. A., Duarte M., Peirani S., Boué G., 2013, arXiv:1310.6756
Berlind A. A. et al., 2006, ApJS, 167, 1
Blaizot J., Wadadekar Y., Guiderdoni B., Colombi S. T., Bertin E., Bouchet F. R., Devriendt J. E. G., Hatton S., 2005, MNRAS, 360, 159
Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, MNRAS, 398, 1150
Brinchmann J., Charlot S., Heckman T. M., Kauffmann G., Tremonti C., White S. D. M., 2004, ArXiv Astrophysics e-prints, arXiv:astro-ph/0406220
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Maccìò A. V., Dutton A. A., van den Bosch F. C., 2008, MNRAS, 391, 1940
Mamon G. A., 1992, ApJ, 401, L3
Mamon G. A., Biviano A., Boué G., 2013, MNRAS, 429, 3079
Mamon G. A., Biviano A., Murante G., 2010, A&A, 520, A30
Mamon G. A., Łokas E. L., 2005, MNRAS, 363, 705
Maraston C., Strömberg G., Thomas D., Wake D. A., Nichol R. C., 2009, MNRAS, 394, L107
Marinoni C., Davis M., Newman J. A., Coil A. L., 2002, ApJ, 580, 122
Marinoni C., Hudson M. J., 2002, ApJ, 569, 101
Mendel J. T., Simard L., Palmer M., Ellison S. L., Patton D. R., 2014, ApJS, 210, 3
Merritt D., 1987, ApJ, 313, 121
Muñoz-Cuartas J. C., Müller V., 2012, MNRAS, 423, 1583
Murray S. G., Power C., Robotham A. S. G., 2013, Astronomy and Computing, 3, 23, http://hmf.icrar.org
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Old L. et al., 2014, MNRAS, 441, 1513
Old L. et al., 2015, MNRAS, submitted
Peng Y.-j. et al., 2010, ApJ, 721, 193
Planck Collaboration et al., 2014, A&A, 571, A16
Robotham A. S. G. et al., 2011, MNRAS, 416, 2640
Rose J. A., 1977, ApJ, 211, 311
Schechter P., 1976, ApJ, 203, 297
Shapley H., 1926, Harvard College Observatory Bulletin, 838, 1
Skibba R. A., van den Bosch F. C., Yang X., More S., Mo H., 2005, MNRAS, 360, 1409
Cone S., Lacey C., 1996, MNRAS, 281, 716
Colless M. et al., 2001, MNRAS, 328, 1039
Conroy C., Gunn J. E., White M., 2009, ApJ, 699, 486
Courtn J., Rasera Y., Alimi J.-M., Corasaniti P.-S., Boucher V., Füzfa A., 2011, MNRAS, 410, 1911
Crocc M., Fosalba P., Castander F. J., Gaztañaga E., 2010, MNRAS, 403, 1353
Domínguez Romero M. J. d. L., García Lambas D., Muriel H., 2012, MNRAS, 427, L6
Duarte M., Mamon G. A., 2014, MNRAS, 440, 1763, paper I
Eisenstein D. J., 1997, ArXiv Astrophysics e-prints, arXiv:astro-ph/9709054
Eke V. R. et al., 2004, MNRAS, 348, 866
Falco M., Mamon G. A., Wojtak R., Hansen S. H., Gottlöber S., 2013, MNRAS, 436, 2639
Gerke B. F. et al., 2005, ApJ, 625, 6
Gunn J. E., Gott J. R., 1972, ApJ, 176, 1
Guo Q. et al., 2011, MNRAS, 164
Hickson P., 1982, ApJ, 255, 382
Hopkins A. M. et al., 2013, MNRAS, 430, 2047
Hubble E. P., 1936, Realm of the Nebulae
Hubble E. P., 1936, Realm of the Nebulae
Hubble E. P., 1936, Realm of the Nebulae
Hugmore F. J., Geller M. J., 1982, ApJ, 257, 423
Jackson J. C., 1972, MNRAS, 156, 1P
Jenkins A., Kent C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, MNRAS, 321, 372
Jones D. H. et al., 2009, MNRAS, 399, 683
Kennicutt, Jr. R. C., Bothun G. D., Schommer R. A., 1984, AJ, 89, 1279
Larson R. B., Tinsley B. M., Caldwell C. N., 1980, ApJ, 237, 692
Liu D.-Z., Ma C., Zhang T.-J., Yang Z., 2011, MNRAS, 412, 2685
Liu H. B., Hsieh B. C., Ho P. T. P., Lin L., Yan R., 2008, ApJ, 681, 1046

<citation_text>
The halo component is limited to the virial radius. The SDSS-DR10 database contains 8 measures of stellar mass for the primary spectroscopic sample. Figure B1 compares these 8 different SDSS errors on stellar mass.

Despite its very high quality, the SDSS survey is not immune to errors on galaxy stellar mass and luminosity. We estimate these errors below.

APPENDIX B: SDSS ERRORS ON GALAXY LUMINOSITY AND STELLAR MASS

For our choice of $r$ (anisotropy model of equation (18), the exponential in equation (A1) becomes

$$\sigma^2(r) = \frac{G}{\nu(r)} \int_0^\infty \exp \left[ 2 \int_r^\infty \frac{\beta(t)}{t} \nu(s) \frac{M(s)}{s^2} \, ds \right],$$

where the term in brackets is expressed in analytical form for simple anisotropy models in an appendix of Mamon et al. (2013). With the anisotropy model of equation [18], the exponential in equation (A1) becomes $(s + r_{200}/c)/(r + r_{200}/c)$. The solution of equation (A1) for a pure NFW model (eq. [16]) with ML velocity anisotropy (eq. [18]) is then

$$\frac{\sigma^2(r)}{GM_{200}/r_{200}} = \frac{c/[6y(y + b)]}{\ln(e + 1) - c/(c + 1)} \times \left\{ 6(3b - 2)y^2(y + 1)^2L_{i2}(-y) + 6b y^4 \coth^{-1}(2y + 1) - 3b y^2(2y + 1) \ln y + 3 \left[ 2y(y + 1)(2y + 1) - b \left( 4y^3 + 8y^2 + 2y - 1 \right) \right] \ln y(1 + y) + 3(3 - 2)y^2(y + 1)^2 \left[ \pi^2 + 3 \ln^2(y + 1) \right] + 3y \left( 4 - 7b \right) y^2 + (5 - 9b) y - b \right\},$$

where $y = c r/r_{200}$, $b = c r_{r200}/r_{200}$, while $L_{i2}$ is the dilogarithm or Spence function:

$$L_{i2}(x) = -\int_0^x \ln(1 - u) \frac{du}{u} = \sum_{i=1}^\infty \frac{x^i}{i^2}.$$  

For our choice of $r_\beta = r_{-2}$, i.e. $b = 1$, equation (A2) simplifies to

$$\frac{\sigma^2(r)}{GM_{200}/r_{200}} = \frac{c/[6y(y + 1)]}{\ln(e + 1) - c/(c + 1)} \times \left\{ 6y^2(y + 1)^2L_{i2}(-y) + 6y^4 \coth^{-1}(2y + 1) - 3y^2(2y + 1) \ln y + y^2(2y + 1)^2 \left[ \pi^2 + 3 \ln^2(y + 1) \right] - 3(2y^2 - 1) \ln(y + 1) - 3y(y + 1)(3y + 1) \right\}.$$  

In equations (A2) and (A4), the dilogarithm of negative argument, $L_{i2}(-x)$ can be approximated using series expansions around $x = 0$, $x = 1$, and $x \to \infty$, yielding

$$L_{i2}(-x) \sim \begin{cases} \sum_{i=1}^{10} (-1)^{i+1} \frac{x^i}{i^2} & x < 0.35 \\ \frac{\pi^2}{12} + \sum_{i=1}^{10} \left( \ln2 - \frac{a_i}{b_i} \right) (1 - x)^i & 0.35 \leq x < 1.95 \\ -\frac{\pi^2}{6} - \ln^2(x) + \sum_{i=1}^{10} (-1)^{i+1} \frac{x^{i-1}}{i^2} & x \geq 1.95 \end{cases}$$

where the coefficients $a_i$ and $b_i$ given in Table A1. Equation (A5) has relative accuracy better than $2.5 \times 10^{-6}$ for all $x$. With the approximation of equation (A5) for $L_{i2}(-x)$, the radial velocity dispersion $\sigma_r$ in equation (A4) has relative accuracy better than $10^{-4}$ for all $r$.

APPENDIX B: SDSS ERRORS ON GALAXY LUMINOSITY AND STELLAR MASS

Despite its very high quality, the SDSS survey is not immune to errors on galaxy stellar mass and luminosity. We estimate these errors below.

B1 SDSS errors on stellar mass

The SDSS-DR10 database contains 8 measures of stellar mass for the primary spectroscopic sample. Figure B1 compares these 8 different measures. Apart from those from the Wisconsin group, the models generally agree to better than 0.3 dex, i.e. the errors on individual masses

\footnote{Even if the halo component is limited to the virial radius $r_{200}$, the upper integration limit in equation (A1) must be infinity.}

© 2013 RAS, MNRAS 000, 0-0
Figure B1. Comparison of the 8 measures of log stellar mass (solar units) in the SDSS-DR10 database. The biases ($\mu$) and bias-corrected differences ($\sigma$) are highlighted. These measures are from the following models: FSPSGranEarlyDust, FSPSGranEarlyNoDust, FSPSGranWideDust, and FSPSGranWideNoDust: logMass respectively from $\text{stellarMassFSPSGranEarlyDust}$, $\text{stellarMassFSPSGranEarlyNoDust}$, $\text{stellarMassFSPSGranWideDust}$, and $\text{stellarMassFSPSGranWideNoDust}$ (Conroy et al. 2009); PassivePort and StarFormingPort: logMass respectively from $\text{stellarMassPassivePort}$ and $\text{stellarMassStarFormingPort}$ (Maraston et al. 2009); PCAWiscM11 and PCAWiscBC03: $\text{mstellar\_median}$ respectively from $\text{stellarMassPCAWiscM11}$ and $\text{stellarMassPCAWiscBC03}$ (Chen et al. 2012), respectively using the Maraston & Stromback (2011) and Bruzual & Charlot (2003) stellar population synthesis models; MPA/JHU: lgm\_tot\_p50 from GalSpecExtra (Brinchmann et al. 2004) using the Bruzual & Charlot (2003) stellar population synthesis model.

is of order $0.3/\sqrt{2} = 0.2$ dex. In particular, the MPA/JHU masses agree with all others to typically better than 0.2 dex for $\sigma$ and 0.3 dex for the rms ($\sqrt{\mu^2 + \sigma^2}$). We therefore adopt an error of 0.2 dex on stellar mass.

B2 SDSS errors on galaxy luminosity

Writing the $r$-band absolute magnitude of a galaxy as
\[ M_r = r - \mu(z) - k_r(z) - A_{\text{Gal}}^r - A_{\text{int}}^r \]  

(B1)

where \( \mu \) is the distance modulus, while \( r, k_r, A_{\text{Gal}}^r, \) and \( A_{\text{int}}^r \) are respectively the apparent magnitude, k-correction, Galactic extinction and internal extinction, all in the \( r \) band. The photometric errors are expected to be less than 0.05 mag, i.e. less than 0.02 dex on luminosity. The error caused by the uncertain distance can be written as the quadratic sum of the error on redshift (as a distance indicator) and the neglect of group peculiar velocities relative to the observer. We do not consider here the galaxy peculiar velocities within a group, as the group finders handle this.

\[
\epsilon(\log_{10} L_r) = \frac{1}{\ln 10} \left[ \left( \frac{\epsilon(v)}{cz} \right)^2 + \left( \frac{\sigma(v_p)}{cz} \right)^2 \right]^{1/2} \lesssim 0.056 \text{ dex}
\]

for \( \epsilon(v) \simeq 30 \text{ km s}^{-1}, \sigma(v_p) \simeq 200 \text{ km s}^{-1} \), and \( z > 0.01 \) (where the assumption of zero difference in peculiar velocity between the galaxy and the observer dominates the error). According to Figure 2 of [Chilingarian et al. (2010)](https://doi.org/10.1093/mnras/stp314), the intrinsic scatter in the k-correction is of order 0.015 mag, i.e. 0.006 dex. Admittedly, the k-correction of [Chilingarian et al.] suffers from some catastrophic errors, but since 99.9% of the galaxies with \( z < 0.12 \) have k-corrections between \(-0.15\) and \(0.25\), it suffices to impose these limits to \( k_r \). Finally, since SDSS spans high galactic latitudes, the uncertainty on the Galactic extinction should be \( \gtrsim 0.075 \text{ mag} \) (the median \( r \)-band extinction of SDSS/Legacy galaxies), i.e. 0.03 dex. The uncertainty on internal extinction is more difficult to measure, but can be estimated to be 0.1 mag, i.e. 0.04 dex. Combining these 6 errors (photometry, redshift, assumption of no peculiar velocity, k-correction, Galactic extinction and internal extinction) in quadrature, we deduce that the error on luminosity is of order of 0.08 dex.