Mon. Not. R. Astron. Soc. **000**, 1–24 (2013) Printed 13 February 2014 (MN LATEX style file v2.2)

Galaxy Cluster Mass Reconstruction Project: I. Methods and first results on galaxy-based techniques

L. Old[1⋆], R. A. Skibba[2], F. R. Pearce[1], D. Croton[3], S. I. Muldrew[1], J. C. Muñoz-Cuartas[4], D. Gifford[5], M. E. Gray[1], A. von der Linden[6,7,8], G. A. Mamon[9], M.R. Merrifield[1], V. Müller[10], R. J. Pearson[11], T. J. Ponman[11], A. Saro[12], T. Sepp[13], C. Sifón[14], E. Tempel[13,15], E. Tundo[1], Y. O. Wang[1,16,17] and R. Wojtak[6]

[1]*School of Physics and Astronomy, University of Nottingham, Nottingham, NG7 2RD, UK*
[2]*Center for Astrophysics and Space Sciences, Department of Physics, University of California, 9500 Gilman Dr, San Diego, CA 92093, USA*
[3]*Centre for Astrophysics & Supercomputing, Swinburne University of Technology, PO Box 218, Hawthorn, VIC 3122, Australia*
[4]*Group for Computational Physics and Astrophysics, Instituto de Fisica, Universidad de Antioquia, Medellin, Colombia*
[5]*Department of Astronomy, University of Michigan, 500 Church St. Ann Arbor, MI, USA*
[6]*Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen , Denmark*
[7]*Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94305-4085, USA*
[8]*Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305-4060, USA*
[9]*Institut d'Astrophysique de Paris (UMR 7095 CNRS & UPMC), 98 bis Bd Arago, F-75014 Paris, France*
[10]*Leibniz-Institut für Astrophysik Potsdam, An der Sternwarte 16, D-14482 Potsdam, Germany*
[11]*School of Physics and Astronomy, University of Birmingham, Birmingham, B15 2TT, UK*
[12]*Department of Physics, Ludwig-Maximilians-Universitat, Scheinerstr. 1, D-81679 Munchen, Germany*
[13]*Tartu Observatory, Observatooriumi 1, 61602 Tõravere, Estonia*
[14]*Leiden Observatory, Leiden University, P.O. Box 9513, NL-2300 RA Leiden, The Netherlands*
[15]*National Institute of Chemical Physics and Biophysics, Rävala pst 10, Tallinn 10143, Estonia*
[16]*Key Laboratory for Research in Galaxies and Cosmology, Shanghai Astronomical Observatory, Shanghai 200030, China*
[17]*Graduate School, University of the Chinese Academy of Sciences, 19A, Yuquan Road, Beijing, China*

Accepted ??. Received ??; in original form ??

ABSTRACT

This paper is the first in a series in which we perform an extensive comparison of various galaxy-based cluster mass estimation techniques that utilise the positions, velocities and colours of galaxies. Our primary aim is to test the performance of these cluster mass estimation techniques on a diverse set of models that will increase in complexity. We begin by providing participating methods with data from a simple model that delivers idealised clusters, enabling us to quantify the underlying scatter intrinsic to these mass estimation techniques. The mock catalogue is based on a Halo Occupation Distribution (HOD) model that assumes spherical Navarro, Frenk and White (NFW) haloes truncated at R_{200}, with no substructure nor colour segregation, and with isotropic, isothermal Maxwellian velocities. We find that, above $10^{14}\mathrm{M}_{\odot}$, recovered cluster masses are correlated with the true underlying cluster mass with an intrinsic scatter of typically a factor of two. Below $10^{14}\mathrm{M}_{\odot}$, the scatter rises as the number of member galaxies drops and rapidly approaches an order of magnitude. We find that richness-based methods deliver the lowest scatter, but it is not clear whether such accuracy may simply be the result of using an over-simplistic model to populate the galaxies in their haloes. Even when given the true cluster membership, large scatter is observed for the majority non-richness-based approaches, suggesting that mass reconstruction with a low number of dynamical tracers is inherently problematic.

Key words: galaxies: clusters - cosmology: observations - galaxies: haloes - galaxies: kinematics and dynamics - methods: numerical - methods: statistical

⋆ E-mail: ppxlo@nottingham.ac.uk

1 INTRODUCTION

Deducing the masses of the largest gravitationally bound structures in the Universe, galaxy clusters, remains a complex problem that is at the focus of current and future cosmological studies. The characteristics of the galaxy cluster population provide crucial information for studies of large scale-structure (e.g., Bahcall 1988; Einasto et al. 2001; Yang et al. 2005b; Papovich 2008; Willis et al. 2013), constraining cosmological model parameters (see Allen et al. 2011 for a review) and galaxy evolution studies (e.g., Goto et al. 2003; Postman et al. 2005; Martínez et al. 2008). Despite the wealth of information clusters can provide, deriving strong constraints from cluster surveys is a non-trivial problem due to the complexity of estimating accurate cluster masses. The use of cluster surveys as a dark energy probe provides greater statistical power than other techniques (Dark Energy Task Force; Albrecht et al. 2006). However, enabling this statistical precision requires significant advances in treating the systematic uncertainties between survey observables and cluster masses.

Clusters can be detected across several different wavelength regimes using various techniques. They are identified in optical and infrared light as over-densities in the number counts of galaxies (e.g., Abell 1958, Zwicky et al. 1968), while colour information improves the contrast by selecting the red galaxies that dominate in these systems (e.g., Gladders & Yee 2005, Koester et al. 2007, Szabo et al. 2011, Ascaso et al. 2012). At X-ray wavelengths, the hot intra-cluster medium produces bright extended sources (e.g., Forman et al. 1972, Böhringer et al. 2000, Rosati et al. 2002, Vikhlinin et al. 2009), while at millimeter wavelengths, inverse Compton scattering of photons from this gas results in characteristic distortions in the cosmic microwave background (e.g., Sunyaev & Zeldovich 1972, Carlstrom et al. 2002, Planck Collaboration et al. 2013, Vanderlinde et al. 2010, Hasselfield et al. 2013). Finally, distortions of images of faint background galaxies through weak gravitational lensing offers perhaps the most direct measure of the huge masses of these systems (e.g., Applegate et al. 2012).

Despite these diverse methods of detecting clusters, no cluster observable *directly* delivers a mass. The cluster mass function is one key method to constrain the dark energy parameter. Ongoing and future dark energy missions plan to consider cluster counts in their analyses. Hence, it is crucial to be able to measure cluster masses as accurately as possible. Follow-up spectroscopy is of great importance to all group/cluster surveys, providing the kinematics of cluster galaxies, which is one of a few mass proxies that is directly related to cluster mass (by providing a direct measure of the dark matter potential well). This series of papers examines various observable - mass relations by testing an extensive range of galaxy-based cluster mass estimation techniques with the aim of calibrating follow-up mass proxies.

Galaxy-based mass estimation techniques commonly follow three general steps: first identify the cluster overdensity, second deduce cluster membership, and finally, using this membership, estimate a cluster mass. Common optical cluster finding methods include using the Huchra & Geller (1982) Friends-Of-Friends (FOF) group-finding algorithm (e.g., Berlind et al. 2006; Li & Yee 2008; Tempel et al. 2012; Jian et al. 2013) and methods based upon Voronoi tessellation (e.g., Marinoni et al. 2002; Lopes et al. 2004; van Breukelen & Clewley 2009; Soares-Santos et al. 2011). Also widely used are red-sequence filtering techniques (e.g., Gladders & Yee 2000; Murphy et al. 2012; Rykoff et al. 2013) and methods that rely on the bright central galaxy (BCG) to identify the presence of a cluster (e.g., Yang et al. 2005a; Koester et al. 2007). Cluster catalogues are also constructed using the positions and magnitudes of galaxies to search for over-densities via the matched filter algorithm (e.g., Postman et al. 1996; Olsen et al. 1999; Kepner et al. 1999; Menanteau et al. 2009).

Once the over-densities are identified, many methods select an initial cluster membership using the groups obtained via the FOF algorithm (e.g., Muñoz-Cuartas & Müller 2012; Pearson et al. in preparation; Tempel et al. 2012), whilst others select galaxies within a specified volume in phase space (e.g., von der Linden et al. 2007; Wojtak et al. 2009; Mamon et al. 2013; Gifford & Miller 2013; Sifon et al. 2013; Pearson et al. in preparation) or within a certain region of colour–magnitude space where cluster galaxies are known to reside (e.g., Saro et al. 2013). Once the initial set of member galaxies is chosen, it is common to iteratively refine membership using either the estimated velocity dispersion, radius and colour information, or even a combination of these properties. Deducing which galaxies are members of a cluster is non-trivial, and unfortunately the inclusion of even quite small fractions of interloper galaxies that are not gravitationally bound to the cluster can lead to a strong bias in velocity dispersion-based mass estimates (e.g., Lucey 1983; Borgani et al. 1997; Cen 1997; Biviano et al. 2006; Wojtak et al. 2007). For this reason, methods often employ careful interloper removal techniques, for example, by modelling interloper contamination when performing density fitting, by using the Gapper technique or via iterative clipping as described above.

Many methods then follow the classical approach of applying the virial theorem to the projected phase space distribution of member galaxies (e.g., Zwicky 1937; Yahil & Vidal 1977; Evrard et al. 2008), assuming that the system is in equilibrium. Other methods utilise the distribution of galaxies in projected phase space, assuming a Navarro, Frenk and White (NFW) density profile (Navarro et al. 1996; Navarro et al. 1997) to obtain an estimate of cluster mass. The number of galaxies associated with a cluster above a given magnitude limit (the richness) is also used as a proxy for mass (e.g., Yee & Ellingson 2003). In addition, the more-recently developed caustic method identifies the projected escape velocity profile of a cluster in radius-velocity phase space, delivering a measure of cluster mass (e.g., Diaferio & Geller 1997; Diaferio 1999; Gifford & Miller 2013).

The aim of this paper is to perform a comprehensive comparison of 23 different methods that employ variations of the techniques described above by deducing both the mass and membership of galaxy clusters from a mock galaxy catalogue. In order to simplify the problem, the clusters are populated with galaxies in a somewhat idealised manner, with cluster locations that are specified *a priori*; in this way, the basic workings of the various algorithms can be tested under optimal conditions, without the potential for confusion from more complex geometries or misidentified clusters.

The paper is organised as follows. We describe the mock galaxy catalogue in Section 2, and the mass reconstruction methods applied to this catalogue are described in Section 3. In Sections 4 and 5, we present our results on cluster mass and membership comparisons. We end with a discussion of our results and conclusions in Section 6. Throughout the paper we adopt a Lambda cold dark matter (ΛCDM) cosmology with $\Omega_0 = 0.25$, $\Omega_\Lambda = 0.75$, and a Hubble constant of $H_0 = 73\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, although none of the conclusions depend strongly on these parameters.

2 DATA

This paper forms the initial part (Phase I) of a large comparison programme aimed at studying how well halo masses can be recovered using a wide variety of group/cluster mass reconstruction techniques based on galaxy properties. As the first step, we intentionally use a very clean and straightforward set-up: a simple HOD galaxy mock catalogue built upon a nearby-Universe light-cone. Later stages of this project will involve more sophisticated mock galaxy catalogues using both more advanced HOD models (Skibba et al. in preparation) and semi-analytic modelling (Croton et al. 2006). This paper sets out to determine the simplest-case baseline by using a clean, well-defined dataset with idealised substructure, sharp boundaries, spherically symmetric haloes and a strong richness correlation. Given initial estimates for the location of the structures, just how bad can it get?

For Phase I, the dataset is the mock galaxy catalogue constructed in Muldrew et al. (2012). We briefly describe the catalogue here, and we refer the reader to the above paper and to Skibba & Sheth (2009) for more details. We begin with the Millennium Simulation (Springel et al. 2005), which tracks the evolution of 2160^3 dark matter particles of mass $8.6 \times 10^8\,h^{-1}\mathrm{M}_{\odot}$ from $z = 127$ to $z = 0$ within a comoving box of side length $500\,h^{-1}$Mpc, with a halo mass resolution of $\sim 5 \times 10^{10} h^{-1}\mathrm{M}_{\odot}$. The simulation adopts a flat ΛCDM cosmology with the following parameters: $\Omega_0 = 0.25$, $\Omega_\Lambda = 0.75$, $\sigma_8 = 0.9$, $n = 1$ and $h = 0.73$. Collapsed haloes at $z = 0$ with at least 20 particles are identified with the SUBFIND (Springel et al., 2001) FOF group-finding algorithm, although consistent results are found with other finders (Muldrew et al., 2011; Knebe et al., 2011). The haloes are populated with galaxies whose luminosities and colours follow the halo-model algorithm described in Skibba et al. (2006) and Skibba & Sheth (2009), which is constrained by the luminosity function, colour–magnitude distribution, and luminosity- and colour-dependent clustering (Zehavi et al., 2005) as observed in the Sloan Digital Sky Survey (SDSS; York et al., 2000). An important assumption in this HOD model is that all galaxy properties – their numbers, spatial distributions, velocities, luminosities, and colours – are determined by halo mass alone, again rendering the model as simple as possible. We specify a minimum r-band luminosity for the galaxies of $M_r = -19{+}5\log(h)$, to stay well above the resolution limit of the Millennium Simulation.[1] Haloes are assigned a 'central' galaxy which has the same position and velocity as the halo centre (Skibba et al. 2011). 'Satellite' galaxies are assumed to be fainter than this object and follow an NFW density profile (Navarro et al. 1997) that obeys the concentration–mass relation from Macciò et al. (2008), with the population extending out to R_{200} (the radius that encloses a density 200 times the critical density of the Universe), assuming isothermal, isotropic, Maxwellian velocity distributions.

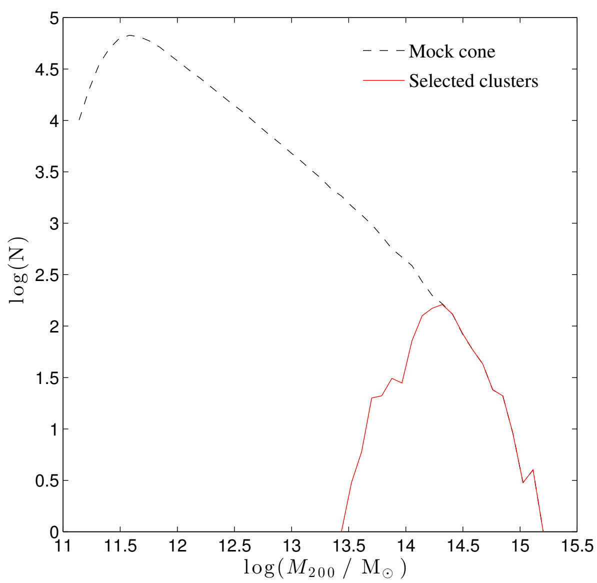

Figure 1. Cluster mass functions for the mock HOD light-cone (black dashed line) and the selected sample (red solid line). To deliver a sample with a sufficient number of cluster-sized haloes, the 1000 groups/clusters are selected by taking the 500 most massive, the next 300 richest and finally the groups/clusters with the most luminous brightest cluster galaxy (BCG) are taken to complete the sample.

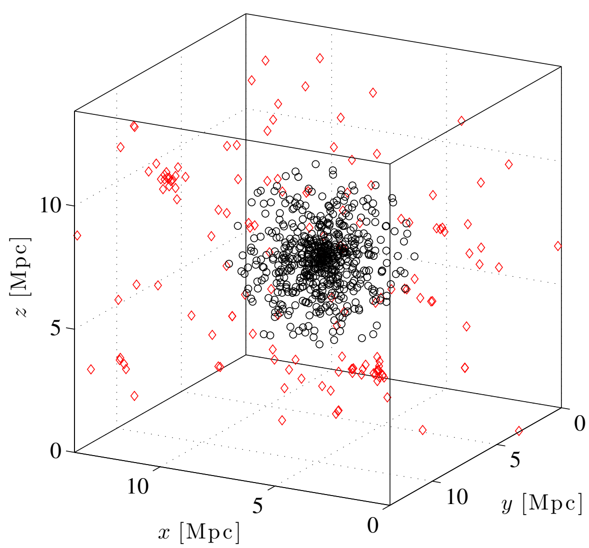

Figure 2. Real-space distribution of the HOD galaxies contained within the largest cluster (black circles) and surrounding galaxies contained within smaller haloes (red diamonds). The galaxy distribution is, by definition, spherical and lacking in substructure.

In the model of galaxy colours, central and satellite galaxies have different colour–luminosity distributions. The central galaxy is usually the reddest galaxy in a given halo, though satellites are redder than central galaxies at a given luminosity (van den Bosch et al. 2008; Skibba 2009). Satellites are assumed to follow a particular sequence on the colour–magnitude diagram, which approaches the red sequence with increasing luminosity, consistent with what is found in the SDSS group/cluster catalogues (Skibba 2009). Note that alternative approaches to modelling galaxy colours and colour-dependent clustering have recently appeared in the literature (Hearin & Watson 2013; Masaki et al. 2013; Gerke et al. 2013; Phleps et al. 2013).

[1] The mass resolution of the simulation is sufficient that haloes that host galaxies as faint as $0.1\,L_*$ ($M_r = 18 + 5\log(\mathrm{h})$) are typically resolved with more than $\sim$100 particles (Springel et al. 2005), which corresponds to a stellar mass threshold of $M_* \sim 10^{9.5} h^{-1}\mathrm{M}_{\odot}$.

We also allow for the expected scatter in the relation between host halo mass and central galaxy luminosity (Zheng et al., 2007). The number of satellite galaxies in a halo of given mass, $P(N_{\rm gal}|M)$, is approximated well by a Poisson distribution (Kravtsov et al., 2004), with a mean HOD that increases approximately linearly with mass, $\langle N_{\rm sat}|M\rangle = [(M - M_0)/M_1^{'}]^{\alpha}$, hence we adopt this distribution to populate the haloes[2]. The value of $M_1^{'}/M_{\rm min} \approx 17$ (where $M_{\rm min}$ is the mass corresponding to the luminosity threshold: $M_r < 19 + 5\log h$), which determines the critical mass above which haloes typically host at least one satellite, is approximately independent of luminosity, and $\alpha \approx 1$ for most of this luminosity range (Zehavi et al. 2011). M_0 determines the shape of the satellite HOD at low halo masses and is typically smaller than $M_{\rm min}$. The HOD parameters are described in detail in Appendix A of Skibba & Sheth (2009).

A galaxy's velocity is given by the sum of the velocity of its parent halo plus an internal motion contribution within the halo. The internal motions are well approximated by a Maxwellian distribution (admittedly, ΛCDM haloes have more complex velocity distribution functions, see, e.g., Sheth & Diaferio 2001 and Beraldo et al. 2013), with velocities that are independent Gaussians in each of the three Cartesian coordinates. The dispersion depends on halo mass and radius through the scaling:

$$\sigma_{200}^2 = GM_{200}/(2R_{200}). \quad (1)$$

Note that this yields velocity dispersions that are 7% greater than expected for NFW models with realistic anisotropic velocities (Mamon et al. 2013). This overestimate of velocity dispersion and the assumption that it is independent of radius cause a violation of local dynamical equilibrium. In this phase of the project, we also neglect the effects of galaxy velocity bias (Skibba et al. 2011; Munari et al. 2013; Old et al. 2013).

A 90×90 degree light-cone, $500\,h^{-1}$Mpc deep, is constructed by taking a slice through the zero-redshift simulation cone. To deliver a dataset with a sufficient number of cluster-sized haloes, the 1000 groups/clusters are selected by taking the 500 most massive, the next 300 richest and finally the groups/clusters with the most luminous brightest cluster galaxy (BCG) are taken to complete the sample. The mass functions of the selected sample and the full light-cone are shown in Figure 1 via the solid red and black dashed line respectively. An example of the underlying galaxy distribution inserted by the HOD model is shown in Figure 2. Black circles indicate the member galaxies for the largest cluster in the sample. By construction, this spatial distribution is smooth and spherical, lacking any imposed substructure, again to keep the test as simple as possible. The red diamonds indicate galaxies in other haloes. As Figure 2 demonstrates, there are many small haloes that have been populated with HOD galaxies that are not part of our target list but which form the background galaxy distribution for this initial phase of the project, to give the mass measuring algorithms a simple contaminant to reject in their analysis. Once a light-cone is generated, the internal velocity dispersion of this large object will be added to the Hubble recession, stretching it out along the line of sight, generating the usual "Finger of God" effect.

In summary, the model for this simplified initial test generates data where clusters are spherically-symmetric, there is no internal substructure, no galaxy velocity bias, no large-scale streaming motions, galaxies follow isotropic orbits and have effectively zero size so there is no blending of objects on the sky.

[2] Note that the true relationship between richness and mass is determined by assumptions about the shape of the halo occupation distribution and its mean as a function of mass. The true number distribution at fixed mass is not Poisson, however, because of the group/cluster selection procedure (see Section 2).

3 MASS RECONSTRUCTION METHODS

In this section, we present the halo mass reconstruction methods used in this comparison project as listed in Table 1, which also summarises some basic properties of each method. The following subsections provide brief descriptions of each method, and are headed by an identifying acronym used throughout this paper, as well as giving the names of the developers who participated in this project and the type of method involved. The two main steps performed by each method, the initial galaxy selection and the mass estimation, are separated into broad classes (which are specified in parentheses in the subsection titles). For the procedure of deducing the initial member galaxy sample, methods are categorised as either FOF (star), red sequence (diamond) or phase space (circle) -based. The mass estimation procedures are classed as either richness (magenta), phase space (black), radius (blue), abundance matching (green) or velocity dispersion (red) -based. Further details can be found in the paper references that are provided in each description, and a more extensive summary of the method characteristics is provided in the appendix.

3.1 PCN (Pearson & Ponman, phase space, Richness)

All PCx methods are based on a cylindrically selected galaxy sample. Starting with the halo positions, galaxies are initially selected from a $5\,{\rm Mpc}$ radius cylinder about each halo with a depth of $\pm 1000\ {\rm km\,s^{-1}}$. The velocity depth is then iterated with a robust $3\,\sigma$ clipping using galaxies within $1\,{\rm Mpc}$. To derive masses, the PCN method uses an aperture richness of each cluster as discussed in Pearson et al. (in preparation). Richness is defined as the number of galaxies above a threshold absolute magnitude within $1\,{\rm Mpc}$, subtracting an interloper contribution estimated using galaxies in a background annulus of radii $3 - 5\,{\rm Mpc}$. Mass is then estimated using a M_{500}-richness relation calibrated on a sample of clusters with SDSS galaxy data and X-ray estimates for M_{500} from Sun et al. (2009) and Sanderson & Ponman (2010). The estimated M_{500} is converted to M_{200} for this project using the mass-concentration relation of Duffy et al. (2008). We estimate statistical errors through bootstrap resampling the observed mass proxy and systematic errors by propagating errors on our calibration relation.

3.2 PFN (Pearson & Ponman, FOF, Richness)

All PFx methods are based on a FOF-selected cluster sample. We apply a FOF analysis using the scheme of Eke et al. (2004) which utilizes the positions and velocities of galaxies. The linked clusters clusters are matched to the given cluster center positions. All linked galaxies are assumed to be cluster members, so we do not include any corrections for interloper contamination. For the PFN method, masses are derived based on the FOF richness of each cluster as discussed in Pearson et al. (in preparation), and calibrated against X-ray masses using the same sample as described for the PCN method, and is converted to M_{200} in the same way. Statistical errors are estimated from Poisson errors propagated through the calibrated mass relation; systematic errors are derived from calibrating uncertainties as for PCN.

Table 1. Summary of the participating cluster mass estimation methods. Listed is an acronym identifying the method, an indication of the scheme used to undertake member galaxy selection and an indication of the method used to convert this membership list to a mass estimate. Note that acronyms denoted with a star indicate that the method did not use our initial 1000 object target list but rather matched these locations at the end of their analysis. Please see Tables A1 and A2 in the appendix for more details on each method.

Method	Initial Galaxy Selection	Mass Estimation	Reference
PCN	Phase space	Richness	Pearson et al. (in prep.)
PFN*	FOF	Richness	Pearson et al. (in prep.)
NUM	Phase space	Richness	Mamon et al. (in prep.)
ESC	Phase space	Phase space	Gifford & Miller (2013)
MPO	Phase space	Phase space	(Mamon et al., 2013)
MP1	Phase space	Phase space	(Mamon et al., 2013)
RW	Phase space	Phase space	Wojtak et al. (2009)
TAR*	FOF	Phase space	Tempel et al. (2014)
PCO	Phase space	Radius	Pearson et al. (in prep.)
PFO*	FOF	Radius	Pearson et al. (in prep.)
PCR	Phase space	Radius	Pearson et al. (in prep.)
PFR*	FOF	Radius	Pearson et al. (in prep.)
HBM*	FOF	Abundance matching	Muñoz-Cuartas & Müller (2012)
MVM*	FOF	Abundance matching	Muñoz-Cuartas & Müller (2012)
AS1	Red sequence	Velocity dispersion	Saro et al. (2013)
AS2	Red sequence	Velocity dispersion	Saro et al. (2013)
AvL	Phase space	Velocity dispersion	von der Linden et al. (2007)
CLE	Phase space	Velocity dispersion	Mamon et al. (2013)
CLN	Phase space	Velocity dispersion	Mamon et al. (2013)
SG1	Phase space	Velocity dispersion	Sifon et al. (2013)
SG2	Phase space	Velocity dispersion	Sifon et al. (2013)
PCS	Phase space	Velocity dispersion	Pearson et al. (in prep.)
PFS*	FOF	Velocity dispersion	Pearson et al. (in prep.)

3.3 NUM (Mamon, phase space, Richness)

The radius R_{200} is estimated using the richness measured in a rectangular area of projected phase space within $1\,\mathrm{Mpc}$ and $1333\,\mathrm{km\,s^{-1}}$ from the halo centre, with a linear relation between $\log R_{200}$ and $\log N(1\ \mathrm{Mpc}, 1333\,\mathrm{km\,s^{-1}})$ deduced from a robust linear fit to the mock clusters analysed by CLE (see Sect. 19 below). The membership is deduced by selecting all galaxies within R_{200} and with velocities, relative to the central halo, smaller (in absolute value) than $2.7\,\sigma_{\rm los}(R)$ (computed from an NFW model, as in the CLE method). See Mamon et al. (in preparation).

3.4 ESC (Gifford & Miller, phase space)

The caustic technique utilizes the radius-velocity phase space information of galaxies in clusters, as well as their dispersion, to estimate the escape velocity profile of the host haloes. The mass profile is inferred by integrating the square of the escape velocity profile multiplied by a parameter $\mathcal{F}_\beta$ which contains information on the potential, density, and velocity anisotropy profiles of the halo along with fundamental constants. $\mathcal{F}_\beta$ is treated as approximately constant (see Diaferio 1999 and Serra et al. 2011) with a value of 0.65 as found in Gifford et al. (2013). Member galaxies are identified as those within the escape velocity envelope in radius-velocity phase space and within the estimated R_{200} of the halo. This technique is described in both Gifford & Miller (2013) and Gifford et al. (2013). * Sect. 3.5 (MPO): "while the red and blue populations follow NFW" -¿ "() follow NFW models, each with its scale radius"

3.5 MPO (Mamon, phase space)

Starting from the sample of members obtained with the CLN algorithm, the virial radius, R_{200}, total mass scale radius, R_ρ, red and blue galaxy population scale radii, $R_{\rm red}$ and $R_{\rm blue}$, and the velocity anisotropies at the virial radius of these red and blue populations are computed using the Bayesian MAMPOSSt method (Mamon et al., 2013). This method jointly fits the positions of the red and blue galaxies in projected phase space. Here, it is assumed that the system is spherically symmetric and that the total mass distribution follows the NFW model, while the red and blue galaxy populations follow NFW models, each with its scale radius. The red and blue populations are assumed to have isotropic orbits at the centre, but increasingly radial or tangential beyond this (with different free outer anisotropies, but a transition scale fixed to be the scale radius of the tracer). The 3D velocities are assumed to be Gaussian at all radii.

3.6 MP1 (Mamon, phase space)

MP1 is like MPO, but is colour-blind: a single tracer population is assumed.

3.7 RW (Wojtak, phase space)

In this method, the halo mass M_{200} is derived from the distribution of galaxies in phase space. It is assumed that the galaxies follow a combination of a spherical NFW model (where number follows mass) with a distribution function of energy and angular momentum derived from ΛCDM haloes (Wojtak et al. 2008), forcing here the inner and outer anisotropies to match those of ΛCDM haloes, and a constant projected density background term that is kept as a free parameter. See Wojtak et al. (2009) for details. The membership is determined by restricting to galaxies within $v_{\rm los} < \sqrt{-2\Phi(R)}$, where R is the projected distance of the galaxy.

3.8 TAR (Tempel, FOF, phase space)

TAR groups/clusters are based upon the conventional FOF group finding algorithm, where the linking-length is calibrated based on the mean distance to nearest galaxy in the plane of the sky. For the current dataset $d_\perp = 0.44\,h^{-1}\,\mathrm{Mpc}$ and $d_{||} = 440\,\mathrm{km\,s^{-1}} = 4.4\,\mathrm{h^{-1}}\,\mathrm{Mpc}$ (assuming $d_{||}/d_\perp = 10$). More details of the group finding algorithm are explained in Tago et al. (2008, 2010) and Tempel et al. (2012). The masses of groups/clusters are estimated by applying the virial theorem to the sphere of radius R_{200}:

$$M = \frac{3\,R_{\mathrm{G}}\sigma_v^2}{G} = 7.0\times10^{12}\,\frac{R_{\mathrm{G}}}{\mathrm{Mpc}}\left(\frac{\sigma_v}{100\,\mathrm{km\,s^{-1}}}\right)^2 M_\odot, \quad (2)$$

where σ_v is the 1D velocity dispersion. The gravitational radius R_G is estimated from the RMS projected radius. For that we assume a NFW profile and find the theoretical relationship between these two parameters. Since the concentration parameter of the NFW profile depends on the halo mass (we use the mass-concentration relation from Macciò et al. 2008), we find the final mass iteratively. See Tempel et al. (2014) for more details of the method.

3.9 PCO (Pearson & Ponman, phase space, Radius)

Using the galaxy membership of PCN, the galaxy overdensity profiles of clusters are modelled and fitted as described in Pearson et al. (in preparation). A projected NFW profile (Bartelmann 1996) plus a uniform background term to allow for interloper contamination, is fitted to all galaxies within $5\,\mathrm{Mpc}$. From the fitted NFW profile a radius R_{500} is found, within which the cumulative number density is $500/\Omega_m$ times the mean cosmic number density of galaxies. This number density is estimated from the SDSS luminosity function of Blanton et al. (2003) where galaxies are counted above a threshold luminosity of $M_r - \log h = -19$. The mass M_{500} within R_{500} is then deduced from R_{500}. These overdensity masses have been calibrated against the X-ray masses described under the PCN method, and as a result a linear scaling is applied to determine the final M_{500} estimate, which is then extrapolated to M_{200}. Error analysis is as for PCN.

3.10 PFO (Pearson & Ponman, FOF, Radius

Using the linked galaxy membership of PFN, the galaxy overdensity profiles of clusters are modelled and fitted as described in Pearson et al. (in preparation). We fit a projected NFW (Bartelmann 1996) profile, assuming that the linked galaxy membership is subject to no interloper contamination. M_{200} is then derived from the fitted profile as for PCO.

3.11 PCR (Pearson & Ponman, phase space, Radius)

Using the galaxy membership within $1\,\mathrm{Mpc}$, as derived for PCN, this method is based on the RMS radius of each cluster as discussed in Pearson et al (2013, in preparation). Note, however, that since we have no way of knowing which galaxies are interlopers, we are unable to make any statistical allowance for them (in contrast to the PCN method). As for PCN, we apply a relation calibrated on X-ray derived masses to estimate M_{500}, which is then extrapolated to M_{200}. Error analysis is as for PCN.

3.12 PFR (Pearson & Ponman, FOF, Radius)

The method is the same as PCR, except that it is applied to the FOF-selected galaxy membership described for PFN.

3.13 HBM (Muñoz-Cuartas, FOF, Abundance Matching)

HBM is based upon an ellipsoidal FOF method with linking lengths adapted according to the estimated halo mass. The linking length along the line of sight is controlled by the (theoretical) velocity dispersion of the halo. Cluster masses are determined by abundance matching between the cluster r-band luminosity function and the theoretical halo mass function of (Warren et al. 2006). The centre of the halo is set to the galaxy with the largest r-band luminosity. The method is described in detail in Muñoz-Cuartas & Müller (2012).

3.14 MVM (Müller, FOF, Abundance Matching)

MVM is the same as HBM with the difference that the group/cluster centre of mass, virial radius and velocity dispersion are derived from the group catalogue according to the virial theorem. The procedure is described in more detail in Muñoz-Cuartas & Müller (2012)

3.15 AS1 (Saro, Red sequence, Velocity dispersion)

AS1 was developed to study possible systematics affecting follow-up dynamical mass estimation of high-redshift massive galaxy clusters. By construction, it assumes that the centre of the cluster is known, along with an initial estimate of R_{200} from other observables. It also assumes an intrinsic scatter of $\approx 30\%$ in mass at fixed velocity dispersion, mainly driven by the triaxial properties of DM haloes. As the simulated clusters in this work are spherical, this is largely overestimated. As the total estimated errors on individual clusters mass could be larger than $\approx\ 60\%$, it does not iterate to solve for R_{200}, but is focussed more on obtaining an average unbiased mass estimation for an ensemble of clusters. Since, for the purpose of this work, no initial R_{200} was given, it assumes a fiducial value of $1\,\mathrm{Mpc}$ for all the mass range. Galaxies must lie within 0.1 magnitude in colour from a model given by Song et al. (2012), which has proven to be a good fit to the observational data and they must also lie within $4000\,\mathrm{km\,s^{-1}}$ from the cluster centre. A final clipping of $\pm3\,\sigma$ is then performed to remove interlopers and provide a robust estimate of the velocity dispersion. A scaling relation, provided in Saro et al. (2013), is then used to convert the velocity dispersion into M_{200}. The model is cosmologically dependent at a background level and assumes a cosmology of $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$ and $\mathrm{h_0} = 0.7$. More details of this method are described in Saro et al. (2013).

3.16 AS2 (Saro, Red sequence, Velocity dispersion)

AS2 follows the same procedure as AS1 but the estimated velocity dispersion is corrected by taking into account the number of galaxies as described by Equation 6 in Saro et al. (2013). Note: the values of the constants a and b of the relation $\log \mathrm{M_{200}} = \mathrm{a} + \mathrm{b}\log\sigma_{\mathrm{v}}$ employed by the ASx methods can be found in Table 2.

3.17 AvL (von der Linden, phase space, Velocity dispersion)

This method is a relatively simple velocity dispersion estimator, as used for SDSS clusters in von der Linden et al. (2007) and for EDisCS clusters in Milvang-Jensen et al. (2008). Galaxies are iteratively selected to lie within $2.5\,\sigma_v$ and $0.8\,R_{200}$ – the latter is estimated from σ_v by assuming the virial theorem. These cuts are chosen to make the method relatively insensitive to contamination from nearby structures; R_{200} and the final σ_v are corrected for the expected bias from sigma-clipping. Note: the values of the constants a and b of the relation $\log \mathrm{M}_{200} = \mathrm{a} + \mathrm{b}\log\sigma_\mathrm{v}$ employed by AvL can be found in Table 2.

3.18 CLE (Mamon, phase space, Velocity dispersion)

The initial membership is limited to $R < 3\,\mathrm{Mpc}$ and $|v| < 4000\,\mathrm{km\,s^{-1}}$. A relative velocity gap technique (Girardi et al. 1993), with gapper coefficient $C = 4$, is initially applied to remove obvious interlopers (keeping the largest subsample). The radius R_{200} is first estimated from the aperture velocity dispersion, where the measured value (using the robust Median Absolute Deviation, see Beers et al., 1990), is matched to the aperture velocity dispersion at R_{200}. This is predicted for a spherical single-component NFW model with concentration of $c = 4$, with the Mamon & Łokas (2005) velocity anisotropy profile (with anisotropy radius equal to the scale radius of the NFW model, as found for ΛCDM haloes by Mamon et al. 2010). The membership is recovered by selecting all galaxies within R_{200} with velocities relative to the central one smaller (in absolute value) than $2.7\sigma_{\rm los}(R)$ (computed from an NFW model with the same velocity anisotropy model, but now assuming a concentration obtained from the ΛCDM concentration-mass relation of Macciò et al., 2008). The virial radius and membership are iterated, now measuring the aperture velocity dispersion using the unbiased standard deviation. The method is described in the appendix of Mamon et al. (2013). Note: the values of the constants a and b of the relation $\log \mathrm{M}_{200} = \mathrm{a} + \mathrm{b}\log\sigma_\mathrm{v}$ employed by CLE can be found in Table 2.

3.19 CLN (Mamon, phase space, Velocity dispersion)

CLN is similar to CLE, but now using the output of NUM as input. Note: the values of the constants a and b of the relation $\log \mathrm{M}_{200} = \mathrm{a} + \mathrm{b}\log\sigma_\mathrm{v}$ employed by CLN can be found in Table 2.

3.20 SG1 (Sifón, phase space, Velocity dispersion)

Both SG1 and SG2 implement the shifting gapper of Fadda et al. (1996) and the velocity dispersion-mass relation of Evrard et al. (2008). All galaxies within $4000\,\mathrm{km\,s^{-1}}$ (rest-frame) of the cluster redshift are binned in projected radial annuli, each of which has at least 15 galaxies and a minimum width of $250\,\mathrm{kpc}$. Galaxies within each bin are ordered by the modulus of the velocity and a main body is defined by finding a gap between two successive velocities of $500\,\mathrm{km\,s^{-1}}$ or more. All galaxies within $1000\,\mathrm{km\,s^{-1}}$ of this boundary are considered halo members. The velocity dispersion is the bi-weight estimate of scale (Beers et al. 1990) of all members. From this velocity dispersion, a mass, M_{200}, is estimated from the velocity dispersion-mass relation of Evrard et al. (2008), and the radius, R_{200}, is estimated from this mass. A new velocity dispersion is computed using only members within R_{200}, and this process is repeated until convergence (usually ~ 3 iterations). The full description of the implementation is in Sifon et al. (2013).

Table 2. Values of the constants a and b of the relation $\log\left(\mathrm{M}_{200}/1\mathrm{x}10^{14}\mathrm{M}_\odot\right) = a + b\log\left(\sigma_\mathrm{v}/1000\mathrm{kms}^{-1}\right)$ employed by methods that utilise the group/cluster velocity dispersion. Please see Tables A1 and A2 in the appendix for more details on each method.

Method	a	b
AS1	1.080	2.910
AS2	1.080	2.910
AvL	1.220	3.000
CLE	1.064	3.000
CLN	1.064	3.000
SG1	1.034	2.975
SG2	1.034	2.975
PCS	0.608	2.280
PFS*	0.797	2.750

3.21 SG2 (Sifón, phase space, Velocity dispersion)

SG2 is the same algorithm as SG1 but with different parameters for the shifting gapper method: radial bins have a minimum width of $150\,\mathrm{kpc}$ and 10 galaxies; the main body boundary is $300\,\mathrm{km\,s^{-1}}$ and all galaxies within $500\,\mathrm{km\,s^{-1}}$ of this boundary are considered members. Consequently, SG1 and SG2 only differ in the membership selection. Note: the values of the constants a and b of the relation $\log \mathrm{M}_{200} = \mathrm{a} + \mathrm{b}\log\sigma_\mathrm{v}$ employed by the SGx methods can be found in Table 2.

3.22 PCS (Pearson & Ponman, phase space, Velocity dispersion)

Using the galaxy membership within $1\,\mathrm{Mpc}$, as for PCN, this method is based on the velocity dispersion of each cluster as discussed in Pearson et al. (in preparation). The velocity dispersion is determined using the Gapper estimator (Beers et al. 1990). From the virial theorem, we expect $M \propto \sigma^3$. In practice, both the normalisation and power law index of the relation between mass and velocity dispersion has been calibrated to the X-ray derived masses, extrapolated to M_{200}, and errors estimated, as described for PCN. Note: the values of the constants a and b of the relation $\log \mathrm{M}_{200} = \mathrm{a}{+}\mathrm{b}\log\sigma_\mathrm{v}$ employed by PCS can be found in Table 2.

3.23 PFS (Pearson & Ponman, FOF, Velocity dispersion)

This algorithm is identical to PCS, except that it uses the FOF-linked galaxy membership of PFN. Note: the values of the constants a and b of the relation $\log \mathrm{M}_{200} = \mathrm{a} + \mathrm{b}\log\sigma_\mathrm{v}$ employed by PFS can be found in Table 2.

4 RESULTS: CLUSTER MASS COMPARISON

In this section, we present results comparing the recovered group/cluster masses of the different reconstruction methods to the 'true' masses in the catalogues. We first make comparisons when the galaxy members are selected by the algorithms, before examining the simpler case where actual galaxy membership is specified (i.e., the case in which membership is known *a priori*). The differences between these results will allow us to distinguish between uncertainties due to the mass estimates and uncertainties due to the identification of group and cluster members. We note that supplying the 'true' galaxy membership does not necessarily guarantee an improvement, nor should it be expected: the methods have at most one or two free parameters, and are not generally tuned to each

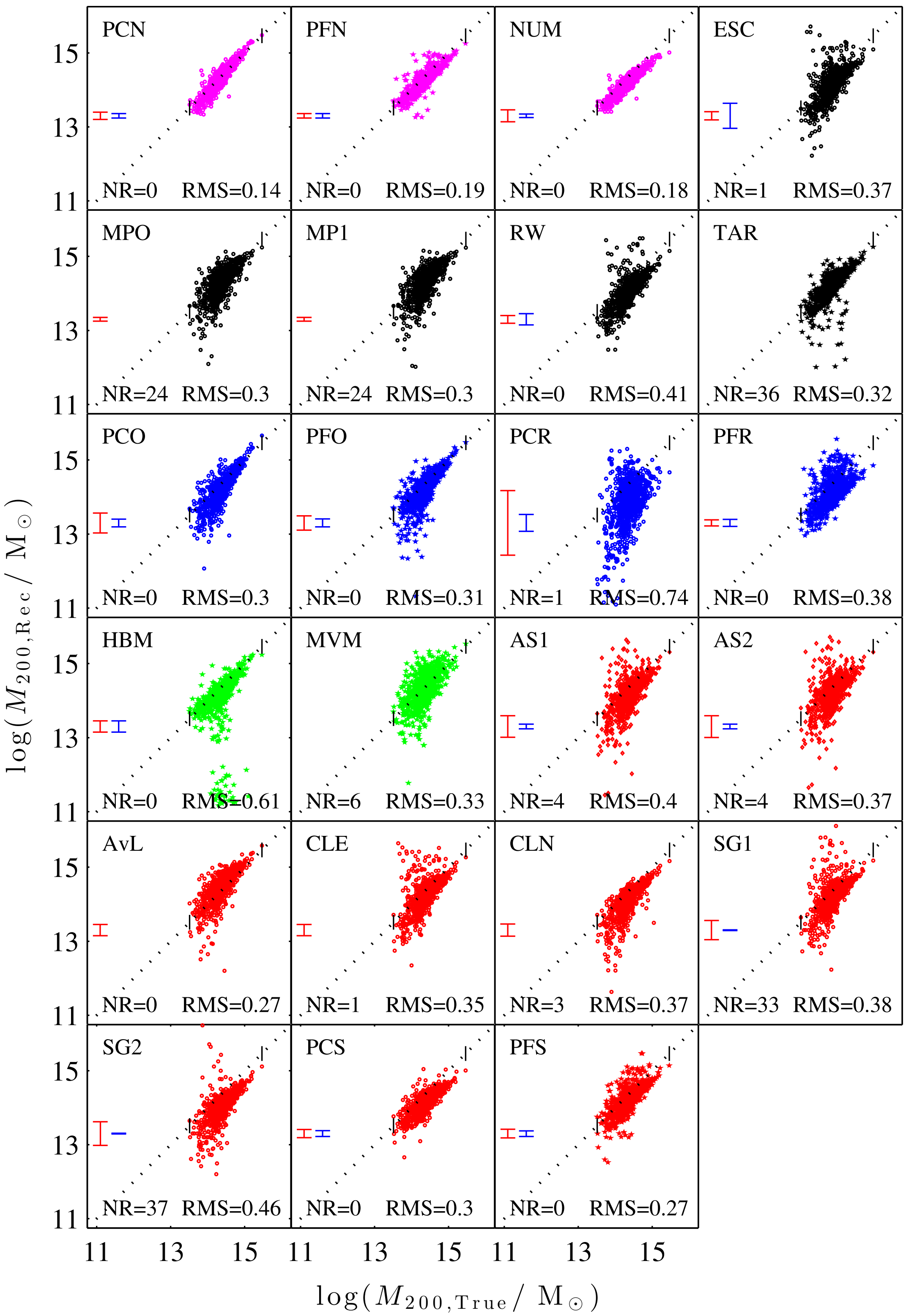

Figure 3. Recovered versus true mass when the group/cluster membership is not known. The black dotted line represents the 1:1 relation. ‘NR’ in the legend represents groups/clusters that are not recovered because they are found to have very low ($< 10^{10}\,\mathrm{M}_\odot$) or zero mass. The black ticks that lie across the 1:1 relation represent the minimum and maximum ‘true’ halo M_{200}. The vertical red bar represents the mean statistical error delivered by methods and the vertical blue bar represents the mean systematic error delivered by methods.

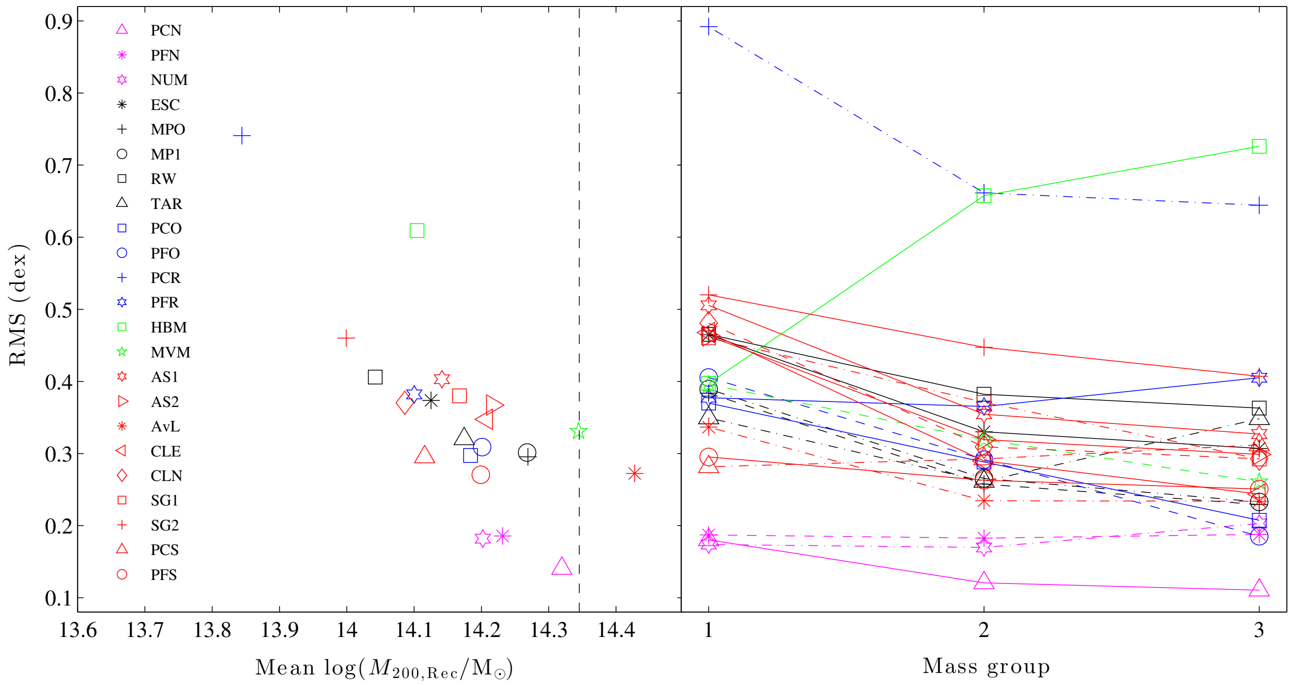

Figure 4. Left hand side: RMS error on log mass versus mean recovered mass (in dex) when the true galaxy membership is not known. The dashed black line identifies where the mean of the true mass distribution lies. Right hand side: RMS error on log mass versus mean recovered mass (in dex) for three mass bins when the true galaxy membership is not known. Mass groups 1, 2 and 3 represent clusters with 'true' M_{200} within the ranges $\log(M_{200}) \leqslant 14.25$, $14.25 < \log(M_{200}) \leqslant 14.45$ and $14.45 < \log(M_{200})$, respectively.

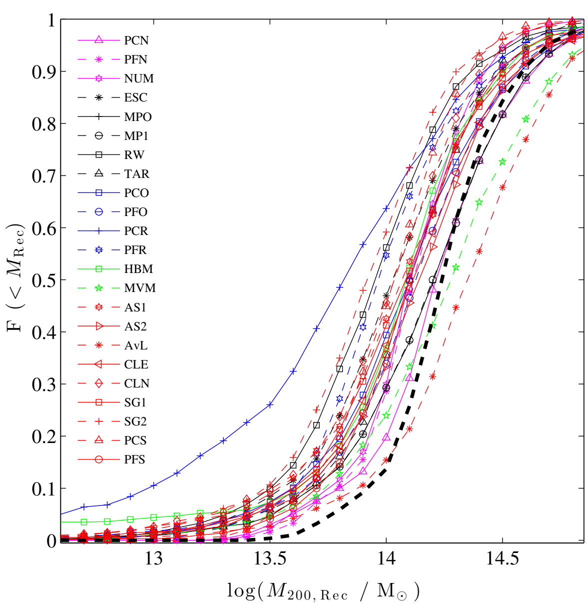

Figure 5. The cumulative distribution functions (CDF) of recovered halo mass delivered by the 23 methods when the group/cluster membership is not known. The true M_{200} CDF is shown via the thick black dashed line. It is clear that all methods are recovering lower halo masses than that of the 'true' mass. A two-sample Kolmogorov-Smirnov (KS) test also demonstrates that the recovered mass distributions are not statistically similar to the true mass distribution (delivering p-values for all methods $p \leqslant 0.01$).

specific dataset but are rather run using their default settings. As such, if the default setting assumes a more restrictive or extensive galaxy membership, the calculation of mass from galaxy members selected in a different way is not guaranteed to be reliable. However, the level of scatter will indicate whether the calculation of mass from galaxy members selected in a different way is a serious issue.

Figure 3 shows the recovered mass versus input mass when the group/cluster membership is *not* given in advance. The colour scheme separates the methods into the broad classifications introduced in the last section and are classed as either richness (magenta), phase space (black), radius (blue), abundance matching (green) or velocity dispersion (red) -based. Methods where membership is based on an FOF linking method have star-shaped markers, red sequence-based methods have diamond markers and phase space-based methods have circle markers. If returned, the vertical bars at the left hand side of each panel indicate the statistical (red) and systematic (blue) errors estimated internally by the mass reconstruction methods themselves without reference to this plot.

Encouragingly, we see a correlation across the input mass range $13 < \log(M_{200}/\mathrm{M}_\odot) < 15$. There is generally good agreement, at least for the inferred masses of massive galaxy clusters. Nonetheless, one can see substantial scatter, especially at masses of $\log(M_{200}/\mathrm{M}_\odot) < 14$, typically associated with groups. Although for some methods or mass regimes the masses may be slightly overestimated, the masses of groups and poor clusters appear to be more often underestimated, except for the methods based upon richness (PCN, PFN and NUM), as well as HBM. These biases are also apparent in Figure B2, which shows the residual recovered mass in dex via $\log(M_{200,\mathrm{Rec}}/M_{200,\mathrm{True}})$. In some cases, these masses may be underestimated by more than an order of magnitude. Phase space methods MPO, MP1, TAR and velocity dispersion methods SGx fail to recover masses for some groups/clusters, but the number of such cases amounts to a very low fraction of the sample. The PCR method fails to recover reliable masses. This method uses the RMS radius of the galaxy distribution extracted within a $1\,\mathrm{Mpc}$ aperture (and velocity range). However, this parameter is inflated by the presence of interlopers (which can removed statistically when calculating richness, for example) and is reduced by the imposed $1\,\mathrm{Mpc}$ aperture. It is noticeable that the PFR method, which is also based on RMS radius, but is less affected by interlopers and has no restrictive aperture imposed, performs significantly better.

In the left hand side of Figure 4, we quantify the error in the estimated masses by calculating the RMS of the difference between the recovered mass and the input mass (in dex) and we display this versus the mean of the recovered mass distribution. The black dashed vertical line identifies the mean of the true mass distribution. It is clear that the majority of methods, with the exception of AvL and MVM, are systematically biased to lower halo masses. For the majority of the methods, the RMS error on M_{200} is of the order of $0.3\,\mathrm{dex}$ (i.e. a factor of 2). Richness-based methods NUM, PFN and PCN produce the lowest RMS values indicating lower scatter. Both PCR and HBM are outliers, delivering substantially higher RMS values than other methods. For HBM, this higher scatter is most likely due to the large tail of groups/clusters recovered with low masses as seen in Figure 3. As we will see below, these lower masses are not seen when the galaxy membership is defined, and they seem to be due to the galaxy selection algorithm returning very few galaxy members (most likely due to a mis-matching cluster centres when HBM performs the initial step of cluster finding).

We present the cumulative distribution functions (CDFs) of the recovered halo masses in Figure 5. The CDFs illustrate the mass range over which a given method tends to under/overestimate halo masses: most methods are biased low over the entire mass range, while a few methods (AvL and MVM) are biased high for massive clusters $\log\,(M_{200}/\mathrm{M}_{\odot}) \geqslant 14.2$. While only $\sim 10\%$ of the input groups/clusters have a mass of $\log\,(M_{200}/\mathrm{M}_{\odot}) \leqslant 14$, some methods assign $\sim 65\%$ of the population a mass of $\log\,(M_{200}/\mathrm{M}_{\odot}) \leqslant 14$, highlighting further that the majority of methods are recovering lower group/cluster masses than one would expect. Those derived using the 23 methods reveal that none of the algorithms return a measured mass distribution consistent with the input data (with p-values for all methods $\leqslant 0.01$). These recovered mass distributions for all 23 methods can be seen in Figure B1 in the appendix.

To quantitatively compare how well the different methods reconstruct group/cluster mass, we calculate the difference between the recovered mass and the true group/cluster M_{200} via $|\log\,(\mathrm{M}_{200,\mathrm{True}}/\mathrm{M}_{200,\mathrm{Rec.}})|$. The mean of these values is taken to calculate the mean deviation along with the dispersion of the deviations. The RMS of these values is used to rank the 23 methods as shown in the final column of Table 3. The method producing the lowest RMS is given a ranking of 1 and the method producing the highest mean deviation is given a ranking of 23. The RMS ranking where the average bias of a given method has been subtracted is also given in the second to last column (Rank_{σ}). Additionally, we separate the groups/clusters into three 'true' M_{200} mass bins: $\log(M_{200}) \leqslant 14.25$, $14.25 < \log(M_{200}) \leqslant 14.45$ and $14.45 < \log(M_{200})$, to explore whether the mean deviation values for each method are consistent across all masses. As seen earlier, we find that the majority of methods have a higher mean deviation for groups/clusters in the lowest group/cluster mass bin, this is highlighted in the right hand side of Figure 4 where the RMS of the difference between the recovered and input masses (in dex) is shown for the three mass groups. It is also clear the three richness-based methods recover the group/cluster masses well. The majority of the remaining methods are very similar, with typical mass estimation errors of a factor of 2 to 3.

In Figure 6, we show the recovered mass versus the input mass when each group/cluster's galaxy membership is specified in advance. Note that though this group/cluster catalogue is derived from the same light-cone, the sample is not exactly the same as those in the previous figures. In order to maintain a blind set-up for comparison, new clusters from the light-cone were added into the sample. As a result, this sample has, on average, poorer groups/clusters. One can see qualitatively similar correlations as in the previous results, but detailed comparisons indicate interesting differences between them. We immediately see that when the methods are not allowed to restrict the galaxy membership according to their usual scheme, for the majority of methods, the recovered masses are poorer (AvL, SG1, SG2, PFS, PCS, PFN, PCN, NUM, CLE, CLN, ESC, PFR, HBM, PCO, PCR, RW, MPO and MP1). This phenomenon is highlighted in Figure 7, where the RMS of the difference between the recovered mass and true group/cluster mass is calculated for both cases of unknown and known membership (in dex). Note that groups/clusters that are not recovered by the methods are excluded in this calculation. The majority of methods show a higher RMS when estimating mass using the true membership as opposed to selecting their own member galaxies. Furthermore, the widths of the distributions (as shown in Figure B3 in the appendix) are not significantly decreased; indeed in some cases they are increased. Moreover, the tail of underestimated group masses is more pronounced. Some of the methods (e.g., ASx, SGx, CLE, ESC and RW) exhibit occasional large mass overestimates when they select their own membership in a manner that does not occur when the galaxy membership is fixed. These overestimates are not driven by nearby large objects, as the number of member galaxies in these objects is recovered approximately correctly. It seems likely that the dynamical mass estimator is failing due to the influence of a small number of interloper galaxies.

Those methods which perform significantly better when provided with the true galaxy membership (HBM, MVM and the two ASx methods) have been calibrated using the true membership of haloes derived from cosmological simulations, so it is natural that they should perform best when provided with a set of galaxies which is not contaminated by interlopers. In contrast, the PCx and PFx methods, for example, have been calibrated using galaxy samples which contain interlopers, and so one would expect their results to be biased when given only the true group/cluster members.

5 RESULTS: CLUSTER MEMBERSHIP

We now examine the galaxy cluster membership delivered by the various methods and compare the richnesses of the recovered systems. Figure 8 presents the richness of the recovered groups and clusters compared to the number of member galaxies in the source catalogue. In general, the recovery of galaxy membership is very good and we find tighter relations with somewhat lower levels of scatter in comparison to the mass recovery results (also highlighted in Figure C2 in the appendix). Certain methods tend to miss members of massive clusters, such as both the AS and PC approaches. This deficit is intrinsic to these methods, in that they are deliberately conservative in their membership selection, focussing on the

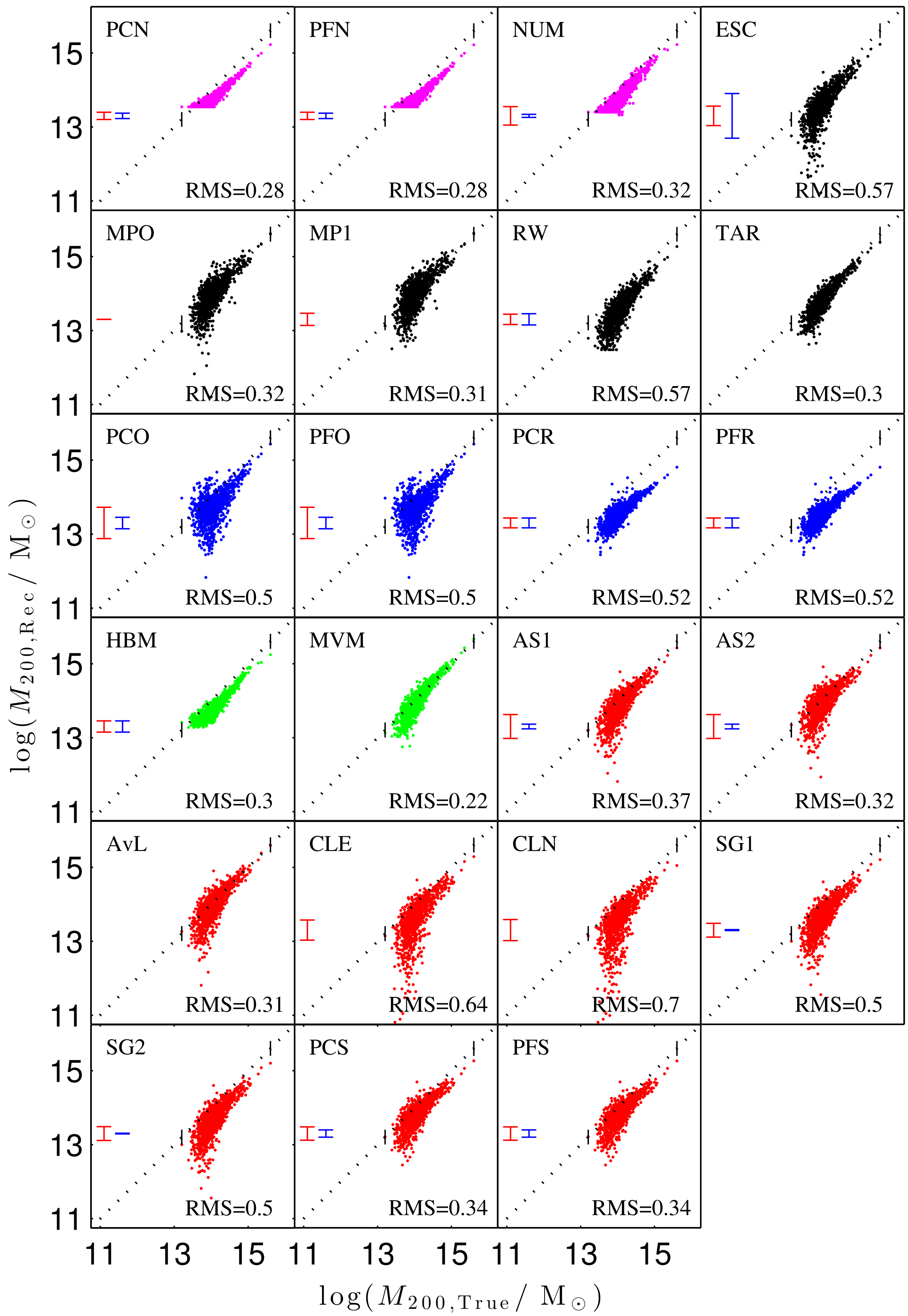

Figure 6. The true cluster mass versus recovered mass when the group/cluster membership is known. The black dotted line represents the 1:1 relation. 'NR' in the legend represents groups/clusters that are not recovered because they are found to have very low ($< 10^{10}\mathrm{M}_\odot$) or zero mass. The black ticks that lie across the 1:1 relation represent the minimum and maximum input group/cluster M_{200}. The vertical red bar represents the mean statistical error delivered by methods and the vertical blue bar represents the mean systematic error delivered by methods.

Table 3. The mean, dispersion, RMS and ranking of $|\log(M_{200,\mathrm{True}}/M_{200,\mathrm{Rec.}})|$ for three 'true' mass bins: $\log(M_{200}) \leqslant 14.25$, $14.25 < \log(M_{200}) \leqslant 14.45$ and $14.45 < \log(M_{200})$. The bins are chosen so that there are roughly equal numbers of clusters in each mass bin. Here '1' represents the method with the lowest RMS and '23' represents the method with the highest RMS. Groups/clusters that are not recovered by the methods are excluded in this calculation. The overall RMS ranking calculated for groups/clusters of all masses where the average bias of a given method has been subtracted, Rank_σ, is given in the second to last column column. The overall RMS ranking calculated without bias subtraction is given in the final column.

Method	$\log(M_{200}) \leqslant 14.25$				$14.25 < \log(M_{200}) \leqslant 14.45$				$\log(M_{200}) \geqslant 14.45$				All masses	
	Mean	σ	RMS	Rank	Mean	σ	RMS	Rank	Mean	σ	RMS	Rank	Rank_σ	Rank
PCN	0.14	0.12	0.18	2	0.10	0.07	0.12	1	0.08	0.07	0.11	1	2	1
PFN	0.14	0.13	0.19	3	0.15	0.11	0.18	3	0.17	0.08	0.19	3	3	3
NUM	0.14	0.11	0.17	1	0.15	0.09	0.17	2	0.18	0.09	0.20	4	1	2
ESC	0.36	0.30	0.46	17	0.28	0.18	0.33	16	0.27	0.15	0.31	15	16	16
MPO	0.28	0.26	0.38	10	0.20	0.17	0.26	5	0.17	0.15	0.23	6	12	7
MP1	0.28	0.27	0.39	11	0.20	0.17	0.27	8	0.17	0.16	0.23	7	13	9
RW	0.38	0.26	0.47	18	0.34	0.18	0.38	20	0.33	0.15	0.36	19	9	20
TAR	0.23	0.26	0.35	7	0.19	0.18	0.26	6	0.21	0.28	0.35	18	10	11
PCO	0.28	0.24	0.37	8	0.23	0.18	0.29	9	0.15	0.15	0.21	5	6	8
PFO	0.26	0.31	0.41	14	0.21	0.21	0.29	11	0.13	0.13	0.19	2	11	10
PCR	0.68	0.57	0.89	23	0.53	0.40	0.66	23	0.55	0.34	0.64	22	22	23
PFR	0.32	0.20	0.38	9	0.33	0.16	0.37	18	0.37	0.17	0.41	20	14	18
HBM	0.19	0.35	0.40	13	0.32	0.58	0.66	22	0.31	0.66	0.73	23	23	22
MVM	0.29	0.27	0.40	12	0.24	0.21	0.32	14	0.21	0.16	0.26	11	19	12
AS1	0.37	0.35	0.51	21	0.27	0.23	0.35	17	0.25	0.21	0.33	17	21	19
AS2	0.32	0.33	0.46	16	0.23	0.22	0.32	15	0.22	0.20	0.30	14	20	14
AvL	0.25	0.22	0.34	6	0.19	0.14	0.23	4	0.17	0.16	0.23	8	7	5
CLE	0.34	0.32	0.47	19	0.23	0.18	0.29	10	0.20	0.14	0.24	9	17	13
CLN	0.34	0.34	0.48	20	0.25	0.18	0.31	13	0.23	0.18	0.29	13	8	15
SG1	0.35	0.29	0.46	15	0.29	0.24	0.37	19	0.24	0.16	0.29	12	18	17
SG2	0.40	0.33	0.52	22	0.39	0.22	0.45	21	0.37	0.17	0.41	21	15	21
PCS	0.23	0.16	0.28	4	0.25	0.15	0.29	12	0.28	0.13	0.31	16	4	6
PFS	0.23	0.19	0.30	5	0.21	0.16	0.26	7	0.22	0.12	0.25	10	5	4

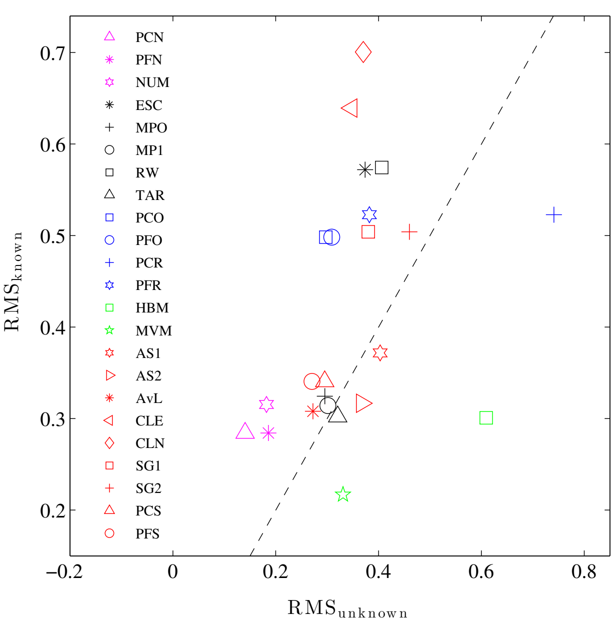

Figure 7. RMS difference in recovered versus true log mass when the membership is known versus that when the membership is not known (in dex). Groups/clusters that are not recovered by the methods are excluded in this calculation. The black dashed line represents a 1:1 relation. The majority of methods have a higher RMS when estimating mass using the true membership.

very central regions of each object; the bias that this selection introduces in the recovered mass is then calibrated out of the estimator. Other methods, such as CLE, ESC, PFx, RW, and SGx, are more contaminated by interlopers and consequently have richness estimates that are biased high. As mentioned earlier, this plot also illustrates that the high recovered mass up-scattered clusters seen in the ASx, ESC and SGx methods in Figure 3 do not seem to be due to line-of-sight contamination by higher mass objects as very few objects have spuriously high numbers of recovered galaxies for these methods.

Most methods have distributions similar in shape to the true sample (as shown in Figure C1 in the appendix) although some distributions are, as noted above, offset either to high values due to interloper inclusion or low values due to conservative membership criteria. Finally, Figure 9 compares the recovered richnesses as a function of recovered mass for the different methods. The three richness-based methods, PFN, PCN, and NUM, have, as expected, very tight relations. For both the input catalogue and these methods, there is a close relationship between richness and cluster mass which may not hold in the real universe. This strong correlation is a consequence of the simple model that we have adopted for this initial part of the comparison project. HBM, a velocity-dispersion-based method, also has a particularly narrow distribution combined with some catastrophic failures due to the mis-matching of a small number of groups/clusters. In contrast, many other methods have more scatter in both recovered number and recovered mass. Moreover, the slope of the $N(M)$ relation is not well recovered for some methods (e.g., ASx, PCx) because of the variation in completeness as a function of mass seen in Figure 8 for these methods: although they recover the halo masses to a similar level of accuracy as the other methods, they should not be used for reliable member galaxy determination.

6 DISCUSSION

The initial set-up used for this project was kept deliberately simple. We began with a simulated dark matter halo catalogue, and a model that inserts galaxies via smooth, spherically-symmetric NFW distributions centred at the centre of the dark matter potential well and scaled by the mass of the halo. Within the $z = 0$ snapshot, haloes of mass above $10^{11.5}\mathrm{M}_{\odot}$ (Figure 1) are populated and a light-cone is then drawn through this distribution to create the "observations" used for this test. Once this baseline study has quantified and minimised the uncertainties intrinsic in mass estimation, we will move on to a more sophisticated cluster model to identify the additional levels of uncertainty that such complexity introduces. Due to the simplicity of the model used for this initial phase and the use of a single cosmology, we cannot comment on the absolute calibration of each model, other than noting that the values have been calibrated to at least approximate reality. The main focus of this paper is to quantify the underlying scatter inherent in cluster mass estimation techniques that use the positions, velocities, and colours of galaxies.

There are three general stages involved in galaxy-based cluster mass estimation. The first stage is the identification of a group/cluster overdensity, the second is the selection of galaxies deemed to be group/cluster members, and the third is the estimation of cluster properties based on this membership. These steps are not, in practice, independent from each other. For instance, a cluster mass estimation method based on dynamical properties might be very sensitive to contamination by unrelated field galaxies. As such, it is perhaps better in such a method to be very conservative with the membership selection at the expense of completeness and then recalibrate the mass estimate based on this incomplete galaxy sample. Conversely, a method based on the volume covered might not be sensitive to interlopers but highly reliant on obtaining a nearly complete galaxy sample.

Following the philosophy of this study of making things as simple as possible, and to aid the inter-comparison of the results of different methods, we supplied the participants with a list of initial centres (i.e. the first stage of this process) about which to look for structures. We further note that not all methods taking part in this study include this step. The centres of the group/cluster sample correspond with the location of the brightest cluster galaxy in all cases and are the "true" location of the halo centre in the DM simulation (the HOD model used places the brightest galaxy at the location of the most bound DM particle in the halo). Some methods (indicated by an asterisk in Table 1) chose not to use this information, and instead used the full galaxy catalogue detecting initial centres themselves. After calculating the properties of the identified groups/clusters, these methods then matched to our supplied coordinates. This is admirable and a more stringent test of these methods. We aim to investigate the issue of initial search location further in subsequent work.

We conclude from this study that, for clusters with masses above $10^{14}\ M_{\odot}$, the uncertainty in the methods seems to be around a factor of two. Richness-based methods have the smallest uncertainties, but this reliability may be due to the underlying simplicity of the HOD model, which includes no a-sphericity, dynamical substructure or large scale velocity distortions. However, we note that low scatter in the richness–mass relation has been observed for photometric samples (e.g., Rozo et al. 2014). Below $10^{14}\ \mathrm{M}_{\odot}$, the scatter rises as the number of member galaxies drops, and the uncertainty rapidly approaches an order of magnitude. This level of error has severe implications for studies of cosmology based on cluster masses given the steeply-falling cluster mass function: there are many more $10^{13}\ \mathrm{M}_{\odot}$ clusters than $10^{14}\ \mathrm{M}_{\odot}$ clusters such that a large scatter in mass estimates will introduce very unpleasant Malmquist-like biases that will render the answers meaningless unless the biases can be very well modelled and controlled.

In order to pinpoint the primary source of the errors, we also supplied the participants with the "true" galaxy cluster membership, as the halo has been initially populated by the HOD model. We then asked the participants to return the group/cluster properties based on this galaxy list rather than the one they had calculated. This simplification did not improve mass estimates; for the majority of methods, the level of scatter was increased. The key factor here is the way in which methods have been calibrated. Those which have been tuned to return unbiased results on the basis of galaxies lying within the 3D 'virial' radius will naturally perform best when provided with such data, whilst methods attuned to the more practical situation in which interlopers cannot be avoided have adopted a variety of approaches to deal with this (aperture selection, background subtraction etc.) and are likely to perform worse in the absence of the expected interlopers. We note that the masses of the cluster sample used for this 'known' membership test are, on average, slightly lower than the 'unknown' membership test. This may deliver a small contribution to the higher levels of scatter, as we have seen previously, that the level of scatter is higher for the lower mass clusters.

The bottom line is that, with the exception of the richness-based methods whose accuracy is unlikely to be realised in a more realistic scenario, the limited number of cluster tracers for the lower-mass systems (typically only $\sim 10-20$) results in an irreducible large uncertainty in the cluster mass estimate. We stress that this experiment has been carried out on the most unchallenging possible test case of spherical systems with known locations and no imposed substructure. Observational challenges such as spectroscopic target selection, incompleteness, and slit/fibre collisions are also not considered. With a more realistic model for the galaxy population and a more observationally challenging set-up, it is likely that accurate group/cluster mass reconstruction will be even more problematic.

ACKNOWLEDGMENTS

The Millennium Simulation used in this paper was carried out by the Virgo Supercomputing Consortium at the Computing Centre of the Max-Planck Society in Garching. The halo merger trees used in the paper are publicly available through the GAVO interface, found at http://www.mpa-garching.mpg.de/millennium/. We would like to acknowledge funding from the Science and Technology Facilities Council (STFC). DC would like to thank the Australian Research Council for receipt of a QEII Research Fellowship. The Dark Cosmology Centre is funded by the Danish National Research Foundation. SIM acknowledges the support of the STFC Studentship Enhancement Program (STEP). E.T. acknowledges the ESF grant MJD272. YOW acknowledges the support of the EU LaceGal grant: PIRSES-GA-2010-269264.

The authors contributed in the following ways to this paper: LO, RAS, FRP, & DC designed and organised this project. LO performed the analysis presented and wrote the majority of the paper.

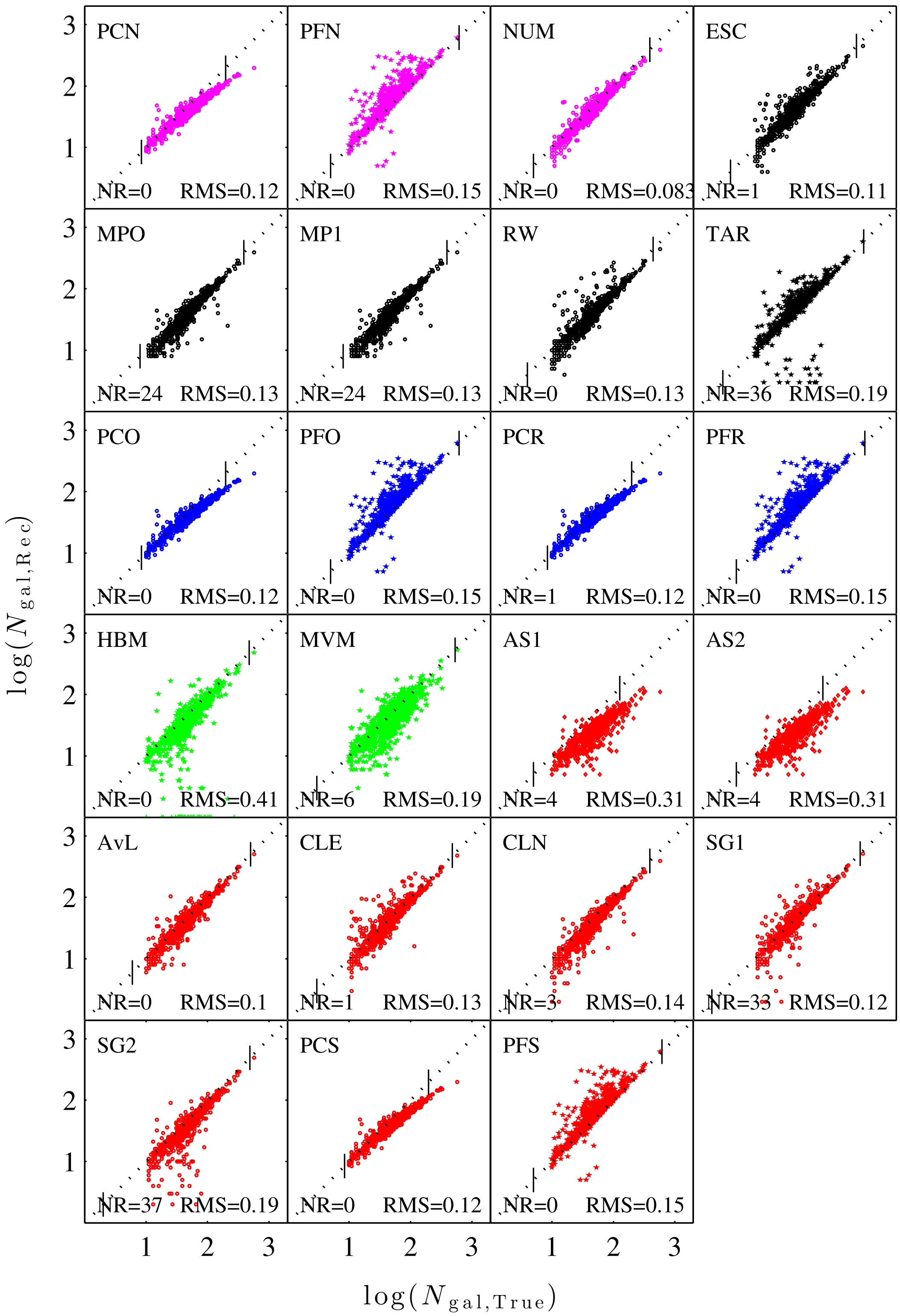

Figure 8. Recovered number of galaxies associated with each group/cluster versus the true number of galaxies when the group/cluster membership is not known. The black dotted line represents the 1:1 relation and the black ticks represent the true minimum and maximum number of galaxies associated with the input groups/clusters. 'NR' in the legend represents groups/clusters that are not recovered because they are found to have very low ($< 10^{10}\mathrm{M}_{\odot}$) or zero mass.

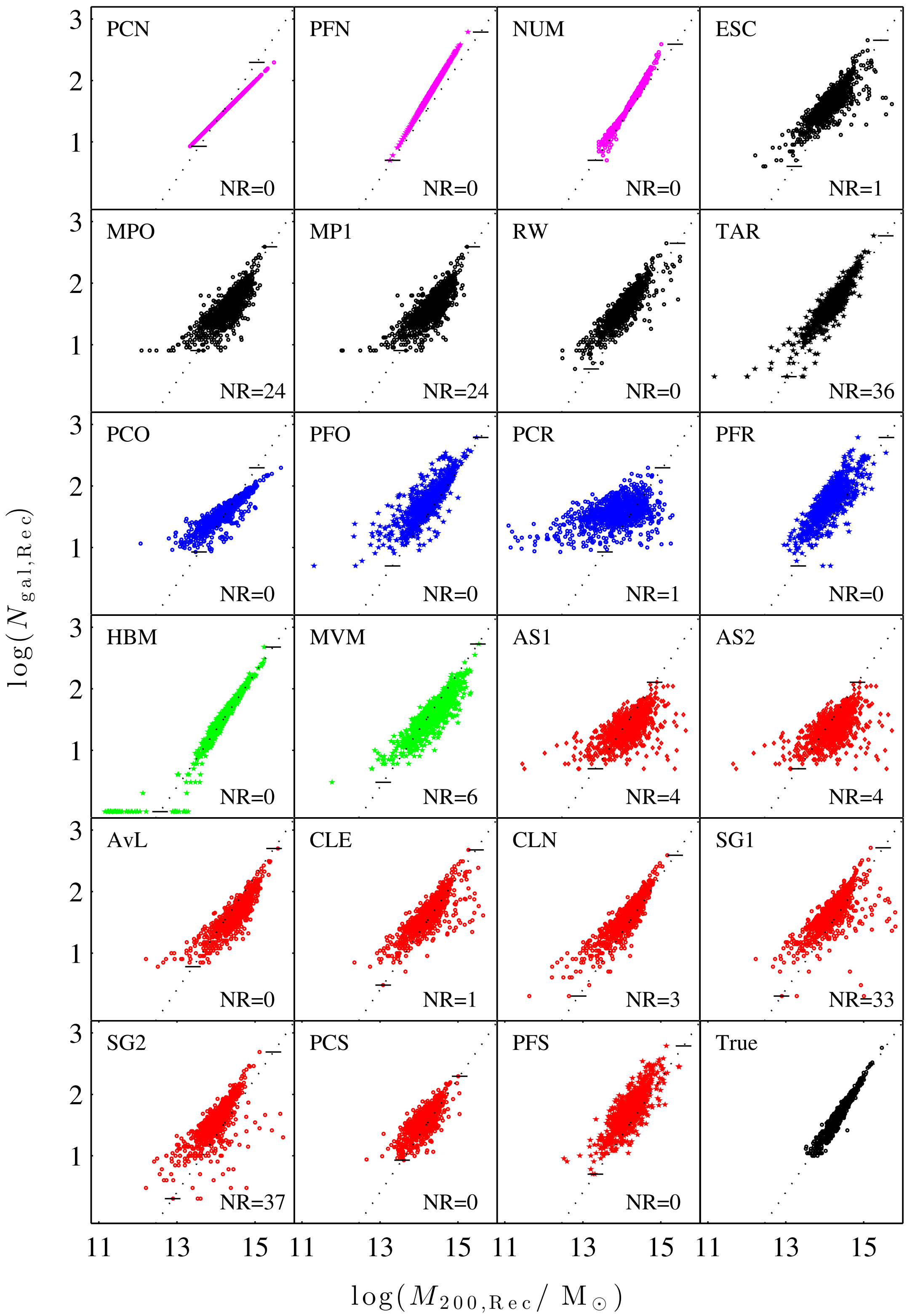

Figure 9. Recovered richness versus recovered mass for each halo, when the group/cluster membership is not known. The black dotted line represents the true mass versus the true number of galaxies associated with each halo and the black ticks represent the true minimum and maximum number of galaxies associated with the input groups/clusters. 'NR' in the legend represents groups/clusters that are not recovered because they are found to have very low ($< 10^{10}\rm M_\odot$) or zero mass. The bottom right panel displays the input HOD mass-richness distribution.

LO, ET, SIM, MEG, RP, TP & FRP organised the workshop that initiated this project. MRM & YW contributed to the analysis. The other authors (as listed in section 3) provided results and descriptions of their respective algorithms. All authors helped proof-read the paper.

References

Abell G. O., 1958, ApJS, 3, 211
Albrecht A. et al., 2006, ArXiv e-prints, astro-ph/0609591
Allen S. W., Evrard A. E., Mantz A. B., 2011, ARA&A, 49, 409
Applegate D. E. et al., 2012, ArXiv e-prints, 1208.0605
Ascaso B., Wittman D., Benítez N., 2012, MNRAS, 420, 1167
Bahcall N. A., 1988, ARA&A, 26, 631
Bartelmann M., 1996, A&A, 313, 697
Beers T. C., Flynn K., Gebhardt K., 1990, AJ, 100, 32
Beraldo L. J., Mamon G. A., Duarte M., Peirani S., Boué G., 2013, ArXiv e-prints, 1310.6756
Berlind A. A. et al., 2006, ApJS, 167, 1
Biviano A., Murante G., Borgani S., Diaferio A., Dolag K., Girardi M., 2006, A&A, 456, 23
Blanton M. R., Hogg D. W., Bahcall, 2003, ApJ, 592, 819
Böhringer H. et al., 2000, ApJS, 129, 435
Borgani S., Gardini A., Girardi M., Gottlober S., 1997, New. Ast., 2, 119
Carlstrom J. E., Holder G. P., Reese E. D., 2002, ARA&A, 40, 643
Cen R., 1997, ApJ, 485, 39
Croton D. J. et al., 2006, MNRAS, 365, 11
Diaferio A., 1999, MNRAS, 309, 610
Diaferio A., Geller M. J., 1997, ApJ, 481, 633
Duffy A. R., Schaye J., Kay S. T., Dalla Vecchia C., 2008, MNRAS, 390, L64
Einasto M., Einasto J., Tago E., Müller V., Andernach H., 2001, AJ, 122, 2222
Eke V. R. et al., 2004, MNRAS, 348, 866
Evrard A. E. et al., 2008, ApJS, 672, 122
Fadda D., Girardi M., Giuricin G., Mardirossian F., Mezzetti M., 1996, ApJ, 473, 670
Forman W., Kellogg E., Gursky H., Tananbaum H., Giacconi R., 1972, ApJ, 178, 309
Gerke B. F., Wechsler R. H., Behroozi P. S., Cooper M. C., Yan R., Coil A. L., 2013, ApJS, 208, 1
Gifford D., Miller C., Kern N., 2013, ApJ, 773, 116
Gifford D., Miller C. J., 2013, ApJL, 768, L32
Girardi M., Biviano A., Giuricin G., Mardirossian F., Mezzetti M., 1993, ApJ, 404, 38
Gladders M. D., Yee H. K. C., 2000, AJ, 120, 2148
Gladders M. D., Yee H. K. C., 2005, VizieR Online Data Catalog, 215, 70001
Goto T., Yamauchi C., Fujita Y., Okamura S., Sekiguchi M., Smail I., Bernardi M., Gomez P. L., 2003, MNRAS, 346, 601
Hasselfield M. et al., 2013, JCAP, 7, 8
Hearin A. P., Watson D. F., 2013, MNRAS, 435, 1313
Huchra J. P., Geller M. J., 1982, ApJ, 257, 423
Jian H.-Y. et al., 2013, ArXiv e-prints, 1305.1891
Kepner J., Fan X., Bahcall N., Gunn J., Lupton R., Xu G., 1999, ApJ, 517, 78
Knebe A. et al., 2011, MNRAS, 415, 2293
Koester B. P. et al., 2007, ApJ, 660, 221
Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, ApJ, 609, 35
Li I. H., Yee H. K. C., 2008, AJ, 135, 809
Lopes P. A. A., de Carvalho R. R., Gal R. R., Djorgovski S. G., Odewahn S. C., Mahabal A. A., Brunner R. J., 2004, AJ, 128, 1017
Lucey J. R., 1983, MNRAS, 204, 33
Macciò A. V., Dutton A. A., van den Bosch F. C., 2008, MNRAS, 391, 1940
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
Marinoni C., Davis M., Newman J. A., Coil A. L., 2002, ApJ, 580, 122
Martínez H. J., Coenda V., Muriel H., 2008, MNRAS, 391, 585
Masaki S., Lin Y.-T., Yoshida N., 2013, MNRAS, 436, 2286
Menanteau F. et al., 2009, ApJ, 698, 1221
Milvang-Jensen B. et al., 2008, A&A, 482, 419
Muñoz-Cuartas J. C., Müller V., 2012, MNRAS, 423, 1583
Muldrew S. I. et al., 2012, MNRAS, 419, 2670
Muldrew S. I., Pearce F. R., Power C., 2011, MNRAS, 410, 2617
Munari E., Biviano A., Borgani S., Murante G., Fabjan D., 2013, MNRAS, 430, 2638
Murphy D. N. A., Geach J. E., Bower R. G., 2012, MNRAS, 420, 1861
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Old L., Gray M. E., Pearce F. R., 2013, MNRAS, 434, 2606
Olsen L. F. et al., 1999, A&A, 345, 681
Papovich C., 2008, ApJ, 676, 206
Phleps S., Wilman D. J., Zibetti S., Budavári T., 2013, ArXiv e-prints, 1313.1340
Planck Collaboration et al., 2013, ArXiv e-prints, 1303.5089
Postman M. et al., 2005, ApJ, 623, 721
Postman M., Lubin L. M., Gunn J. E., Oke J. B., Hoessel J. G., Schneider D. P., Christensen J. A., 1996, AJ, 111, 615
Rosati P., Borgani S., Norman C., 2002, ARA&A, 40, 539
Rozo E., Rykoff E. S., Bartlett J. G., Melin J. B., 2014, ArXiv e-prints
Rykoff E. S. et al., 2013, ArXiv e-prints, 1303.3562
Sanderson A. J. R., Ponman T. J., 2010, MNRAS, 402, 65
Saro A., Mohr J. J., Bazin G., Dolag K., 2013, ApJ, 772, 47
Serra A. L., Diaferio A., Murante G., Borgani S., 2011, MNRAS, 412, 800
Sheth R. K., Diaferio A., 2001, MNRAS, 322, 901
Sifon C. et al., 2013, ApJ, 772, 25
Skibba R., Sheth R. K., Connolly A. J., Scranton R., 2006, MNRAS, 369, 68
Skibba R. A., 2009, MNRAS, 392, 1467
Skibba R. A., Sheth R. K., 2009, MNRAS, 392, 1080
Skibba R. A., van den Bosch F. C., Yang X., More S., Mo H., Fontanot F., 2011, MNRAS, 410, 417
Soares-Santos M. et al., 2011, ApJ, 727, 45
Song J., Mohr J. J., Barkhouse W. A., Warren M. S., Dolag K., Rude C., 2012, ApJ, 747, 58
Springel V. et al., 2005, Nature, 435, 629
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Sun M., Voit G. M., Donahue M., Jones C., Forman W., Vikhlinin A., 2009, ApJ, 693, 1142
Sunyaev R. A., Zeldovich Y. B., 1972, CoASP, 4, 173

Szabo T., Pierpaoli E., Dong F., Pipino A., Gunn J., 2011, ApJ, 736, 21
Tago E., Einasto J., Saar E., Tempel E., Einasto M., Vennik J., Müller V., 2008, A&A, 479, 927
Tago E., Saar E., Tempel E., Einasto J., Einasto M., Nurmi P., Heinämäki P., 2010, A&A, 514, A102
Tempel E., Tago E., Liivamägi L. J., 2012, A&A, 540, A106
Tempel E. et al., 2014, ArXiv e-prints
van Breukelen C., Clewley L., 2009, MNRAS, 395, 1845
van den Bosch F. C., Aquino D., Yang X., Mo H. J., Pasquali A., McIntosh D. H., Weinmann S. M., Kang X., 2008, MNRAS, 387, 79
Vanderlinde K. et al., 2010, ApJ, 722, 1180
Vikhlinin A. et al., 2009, ApJ, 692, 1033
von der Linden A., Best P. N., Kauffmann G., White S. D. M., 2007, MNRAS, 379, 867
Warren M. S., Abazajian K., Holz D. E., Teodoro L., 2006, ApJ, 646, 881
Willis J. P. et al., 2013, MNRAS, 430, 134
Wojtak R., Łokas E. L., Mamon G. A., Gottlöber S., 2009, MNRAS, 399, 812
Wojtak R., Łokas E. L., Mamon G. A., Gottlöber S., Klypin A., Hoffman Y., 2008, MNRAS, 388, 815
Wojtak R., Łokas E. L., Mamon G. A., Gottlöber S., Prada F., Moles M., 2007, A&A, 466, 437
Yahil A., Vidal N. V., 1977, ApJ, 214, 347
Yang X., Mo H. J., van den Bosch F. C., Jing Y. P., 2005a, MNRAS, 356, 1293
Yang X., Mo H. J., van den Bosch F. C., Jing Y. P., 2005b, MNRAS, 357, 608
Yee H. K. C., Ellingson E., 2003, ApJ, 585, 215
York D. G. et al., 2000, AJ, 120, 1579
Zehavi I. et al., 2011, ApJ, 736, 59
Zehavi I. et al., 2005, ApJ, 630, 1
Zheng Z., Coil A. L., Zehavi I., 2007, ApJ, 667, 760
Zwicky F., 1937, ApJ, 86, 217
Zwicky F., Herzog E., Wild P., 1968, Catalogue of galaxies and of clusters of galaxies

APPENDIX A: PROPERTIES OF THE MASS RECONSTRUCTION METHODS

Table A1. Table illustrating the member galaxy selection process for all methods. The second column details how each method selects an initial member galaxy sample, while the third column outlines the member galaxy sample refining process. Finally, the fourth column describes how methods treat interloping galaxies that are not associated with the clusters.

Methods	Member galaxy selection methodology		
	Initial Galaxy Selection	Refine Membership	Treatment of Interlopers
PCN	Within 5 Mpc, 1000 $\mathrm{km\,s^{-1}}$	Clipping of $\pm 3\,\sigma$, using galaxies within 1 Mpc	Use galaxies at $3-5$ Mpc to find interloper population to remove
PFN	FOF	No	No
NUM	Within 1 Mpc, 1333 $\mathrm{km\,s^{-1}}$	1) Estimate R_{200} by a relationship between R_{200} and richness deduced from CLE; 2) Galaxies within R_{200} and with velocities less than $2.7\,\sigma_{\rm los}(R)$ are selected	No
ESC	Within preliminary R_{200} estimate and $\pm 3500\,\mathrm{km\,s^{-1}}$	Gapper technique	Removed in refining by Gapper technique
MPO	Input from CLN	1) Calculate R_{200}, R_ρ, $R_{\rm red}$, $R_{\rm blue}$ by MAMPOSSt method; 2) Select members within radius according to colour	No
MP1	Input from CLN	Same as MPO except colour blind	No
RW	Within 3 Mpc, 4000 $\mathrm{km\,s^{-1}}$	Within R_{200}, $\lvert 2\Phi(R)\rvert^{1/2}$, where R_{200} obtained iteratively	No
TAR	FOF	No	No
PCO	Input from PCN	Input from PCN	Include interloper contamination in density fitting
PFO	Input from PFN	Input from PFN	No
PCR	Input from PCN	Input from PCN	Same as PCN
PFR	Input from PFN	Input from PFN	No
HBM	FOF (ellipsoidal search range, centre of most luminous galaxy)	Increasing mass limits, then FOF, loops until closure condition	No
MVM	Same as HBM	Same as HBM	No
AS1	Within 1 Mpc, 4000 $\mathrm{km\,s^{-1}}$, constrained by colour-magnitude relation	Clipping of $\pm 3\,\sigma$	Removed by clipping of $\pm 3\,\sigma$
AS2	Within 1 Mpc, 4000 $\mathrm{km\,s^{-1}}$, constrained by colour-magnitude relation	Clipping of $\pm 3\,\sigma$	Removed by clipping of $\pm 3\,\sigma$
AvL	Within $2.5\,\sigma_v$ and $0.8\,R_{200}$	Obtain R_{200} and σ_v by σ-clipping	No
CLE	Within 3 Mpc ,4000 $\mathrm{km\,s^{-1}}$	1) Estimate R_{200} by aperture velocity dispersion; 2) galaxies within R_{200} and with velocities less than $2.7\,\sigma_{\rm los}(R)$ are selected; 3) Iterate steps 1 and 2 until convergence	Obvious interlopers are removed by velocity gap technique, then further treated in iteration by σ clipping
CLN	Input from NUM	Same as CLE	Same as CLE
SG1	Within 4000 $\mathrm{km\,s^{-1}}$	1) Measure $\sigma_{\rm gal}$, estimate M_{200} and R_{200}; 2) Select galaxies within R_{200}; 3) Iterate steps 1 and 2 until convergence	Shifting gapper with minimum bin size of 250 kpc and 15 galaxies; velocity limit 1000 $\mathrm{km\,s^{-1}}$ from main body
SG2	Within 4000 $\mathrm{km\,s^{-1}}$	1) Measure $\sigma_{\rm gal}$, estimate M_{200} and R_{200}; 2) Select galaxies within R_{200}; 3) Iterate steps 1 and 2 until convergence	Shifting gapper with minimum bin size of 150 kpc and 10 galaxies; velocity limit 500 $\mathrm{km\,s^{-1}}$ from main body
PCS	Input from PCN	Input from PCN	Same as PCN
PFS	Input from PFN	Input from PFN	No

Table A2. Table showing the characteristics of the mass reconstruction process of methods used in this comparison. The second, third, fourth and fifth columns illustrate whether a method calculates/utilizes the velocities, velocity dispersion, radial distance of galaxies from cluster centre, the richness and the projected phase space information of galaxies respectively. If a method assumed a mass or number density profile it is indicated in columns six and seven.

Methods	Properties used to recovering halo mass						
	Velocities	Velocity dispersion	Radial distance	Richness	Projected phase space	Mass density profile	Number density profile
PCN	Yes	No	No	Yes	No	No	No
PFN	Yes	No	No	Yes	No	No	No
NUM	Yes	No	No	Yes	Yes	No	No
ESC	Yes	Yes	Yes	No	No	Caustics	No
MPO	Yes	No	Yes	No	Yes	NFW	Yes
MP1	Yes	No	Yes	No	Yes	NFW	Yes
RW	Yes	No	Yes	No	Yes	NFW	Yes
TAR	Yes	Yes	Yes	No	No	NFW	No
PCO	Yes	No	No	No	No	NFW	Yes
PFO	Yes	No	No	No	No	NFW	Yes
PCR	Yes	No	Yes	No	No	No	No
PFR	Yes	No	Yes	No	No	No	No
HBM	Yes	Yes	Yes	No	No	NFW	No
MVM	Yes	Yes	Yes	No	No	NFW	No
AS1	Yes	Yes	No	No	No	No	No
AS2	Yes	No	Yes	No	Yes	No	No
AvL	Yes	Yes	Yes	No	No	No	No
CLE	Yes	Yes	No	No	No	NFW	NFW
CLN	Yes	Yes	No	No	No	NFW	NFW
SG1	Yes	Yes	Yes	No	No	No	No
SG2	Yes	Yes	Yes	No	No	No	No
PCS	Yes	Yes	No	No	No	No	No
PFS	Yes	Yes	No	No	No	No	No

2 APPENDIX B: RECOVERED MASS DISTRIBUTIONS AND RESIDUALS

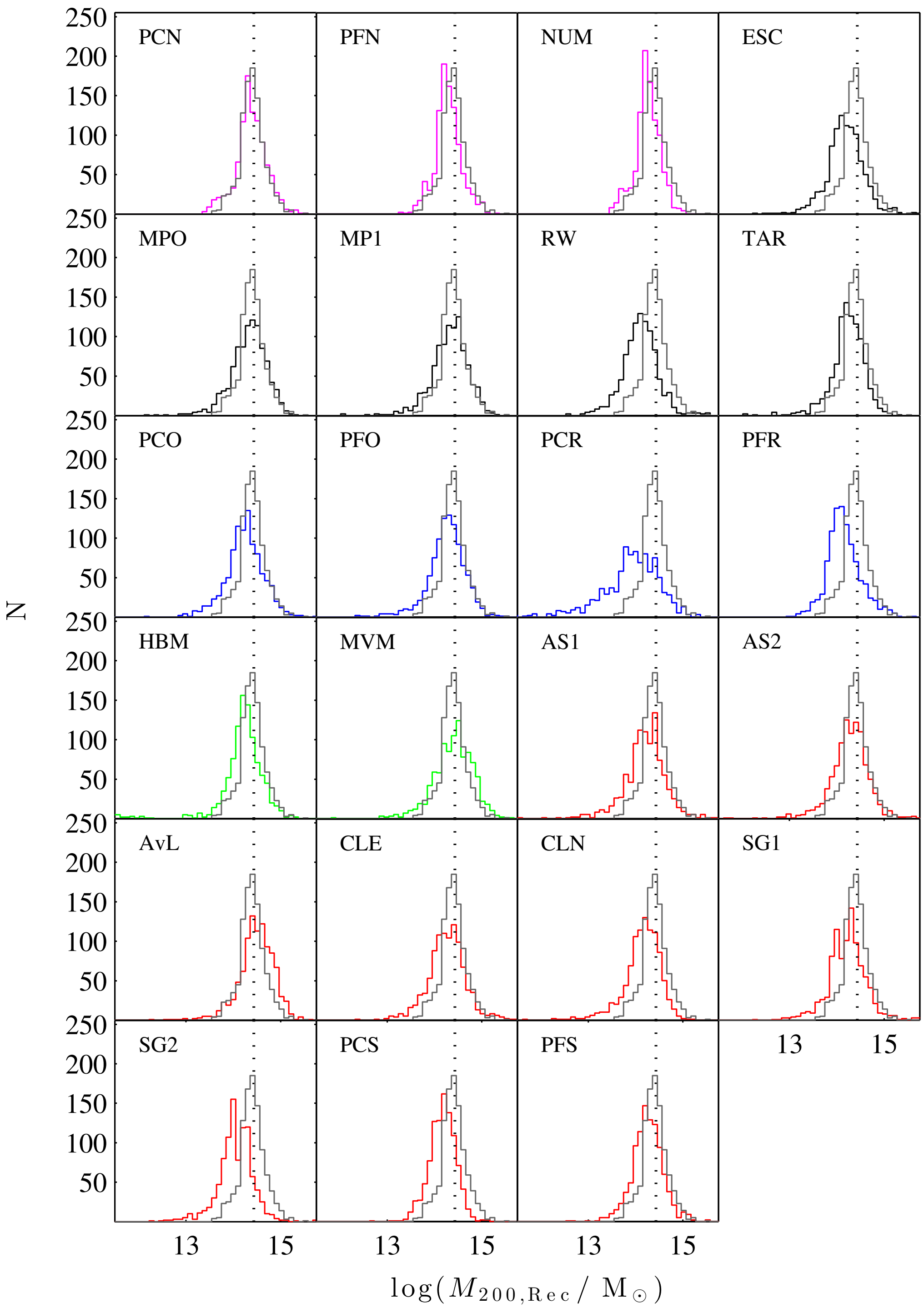

Figure B1. Recovered mass distributions when the group/cluster membership is not known. The black dotted line represents the mean of the true mass distribution and the grey distributions on each subplot represent the true mass distributions.

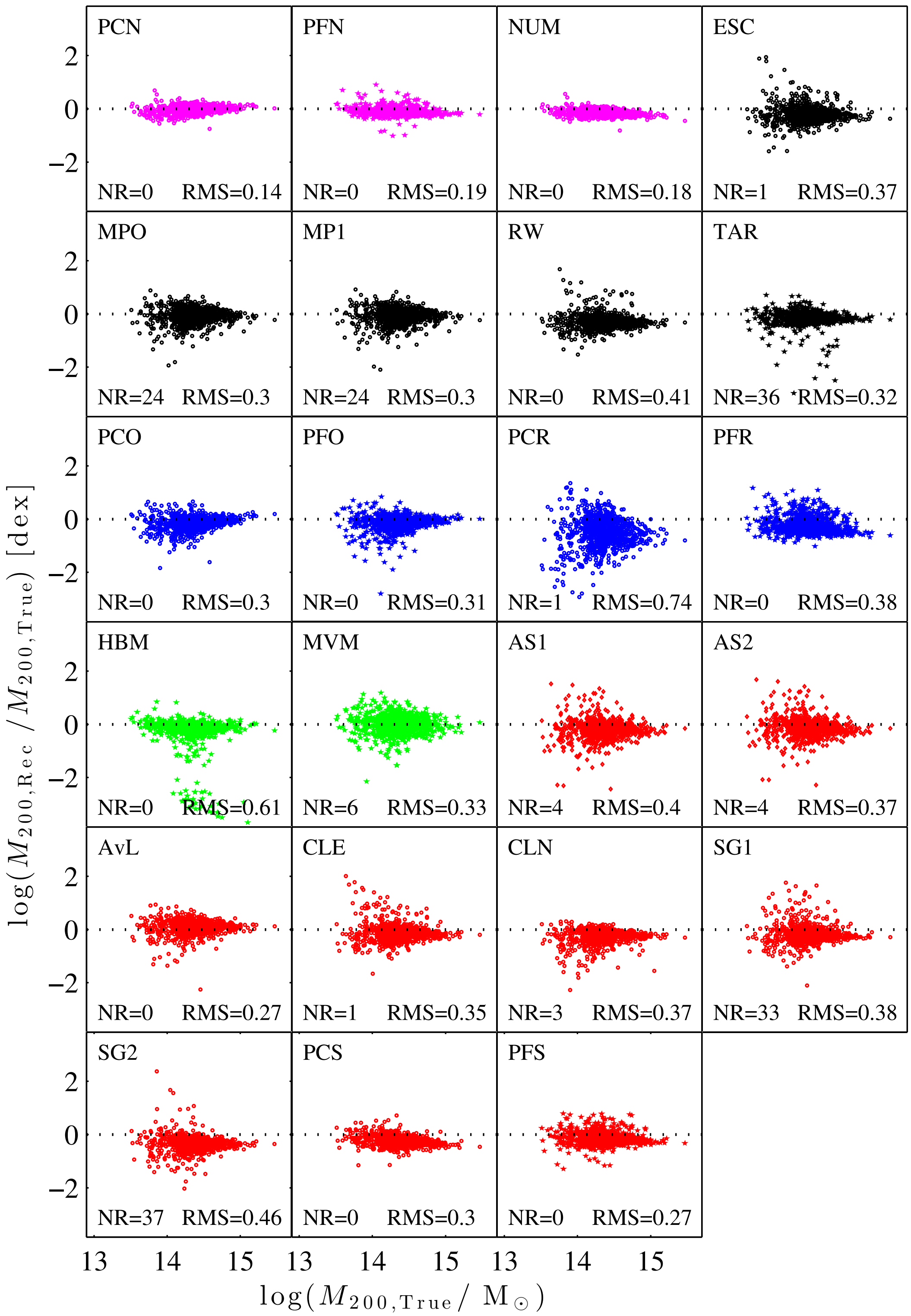

Figure B2. Mass bias versus true mass when the group/cluster membership is not known. The black dotted line represents a residual of zero. 'NR' in the legend represents groups/clusters that are not recovered because they are found to have very low ($< 10^{10}\ \mathrm{M}_\odot$) or zero mass.

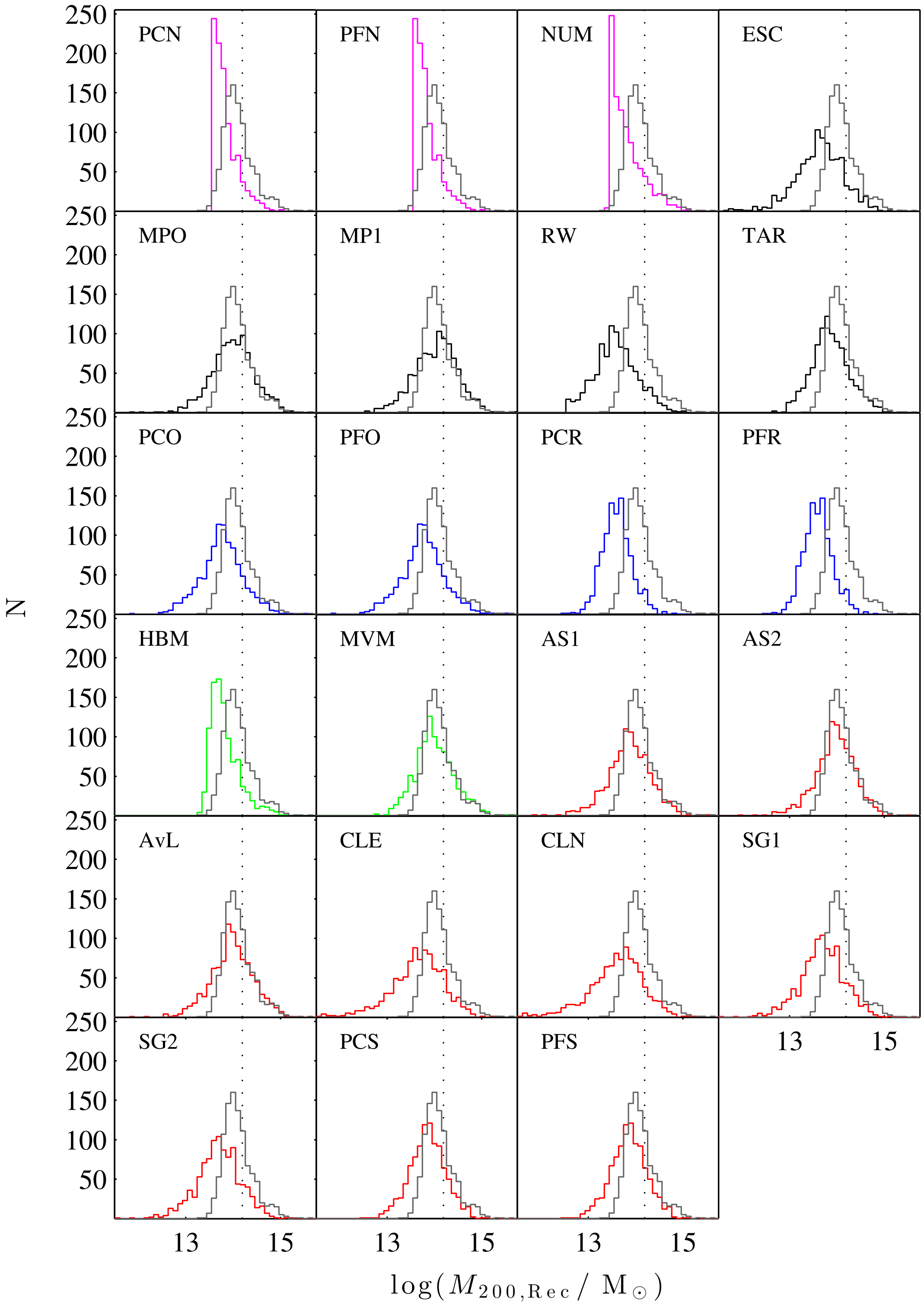

Figure B3. Recovered mass distributions when the group/cluster membership is known. The black dotted line represents the mean of the true mass distribution and the grey distributions on each subplot represent the true mass distributions.

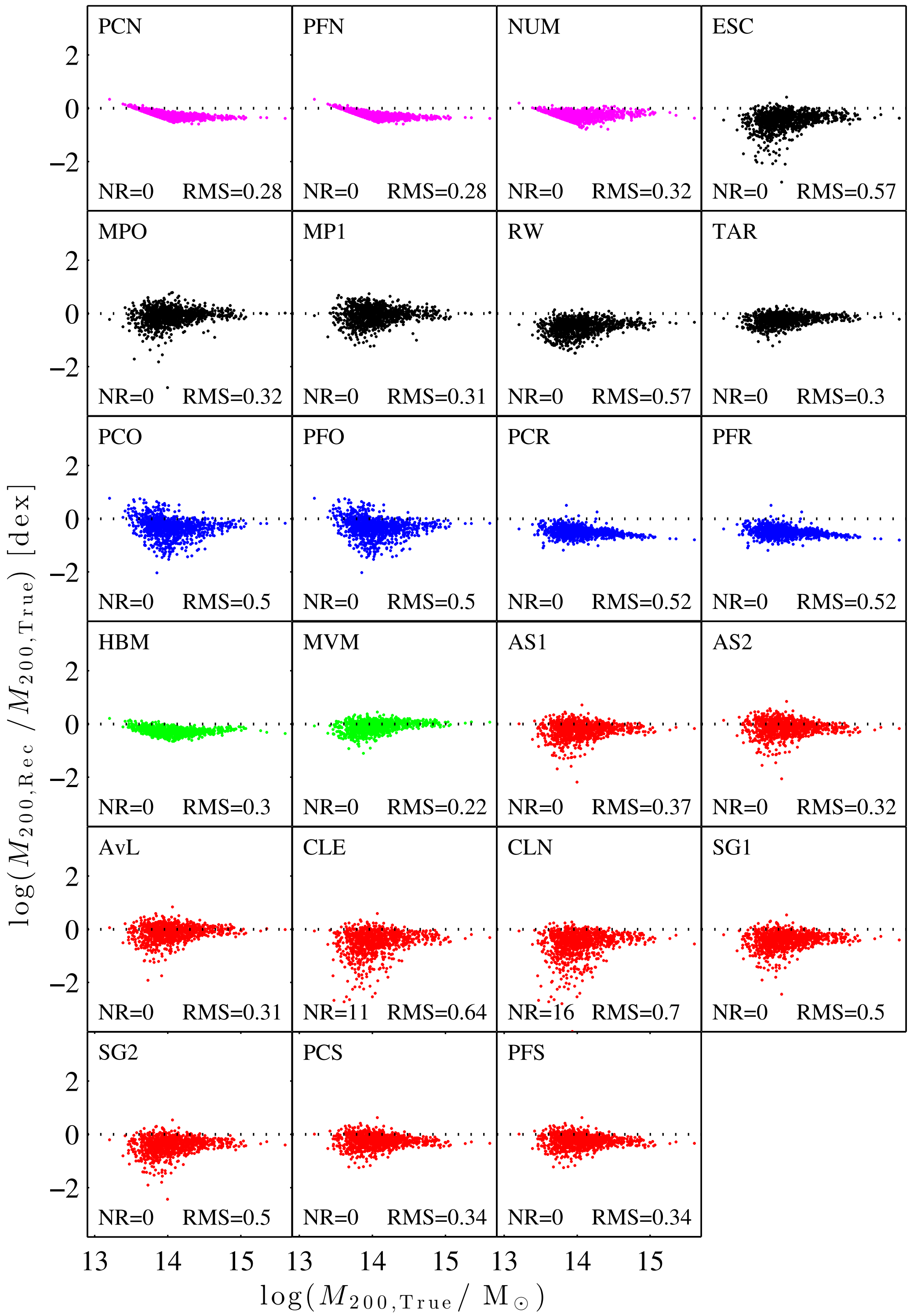

Figure B4. Mass bias versus true mass when the group/cluster membership is known. The black dotted line represents a residual of zero. ‘NR’ in the legend represents groups/clusters that are not recovered because they are found to have very low ($< 10^{10}$ $\mathrm{M}_\odot$) or zero mass.

3 APPENDIX C: RECOVERED NUMBER OF GALAXIES DISTRIBUTIONS AND RESIDUALS

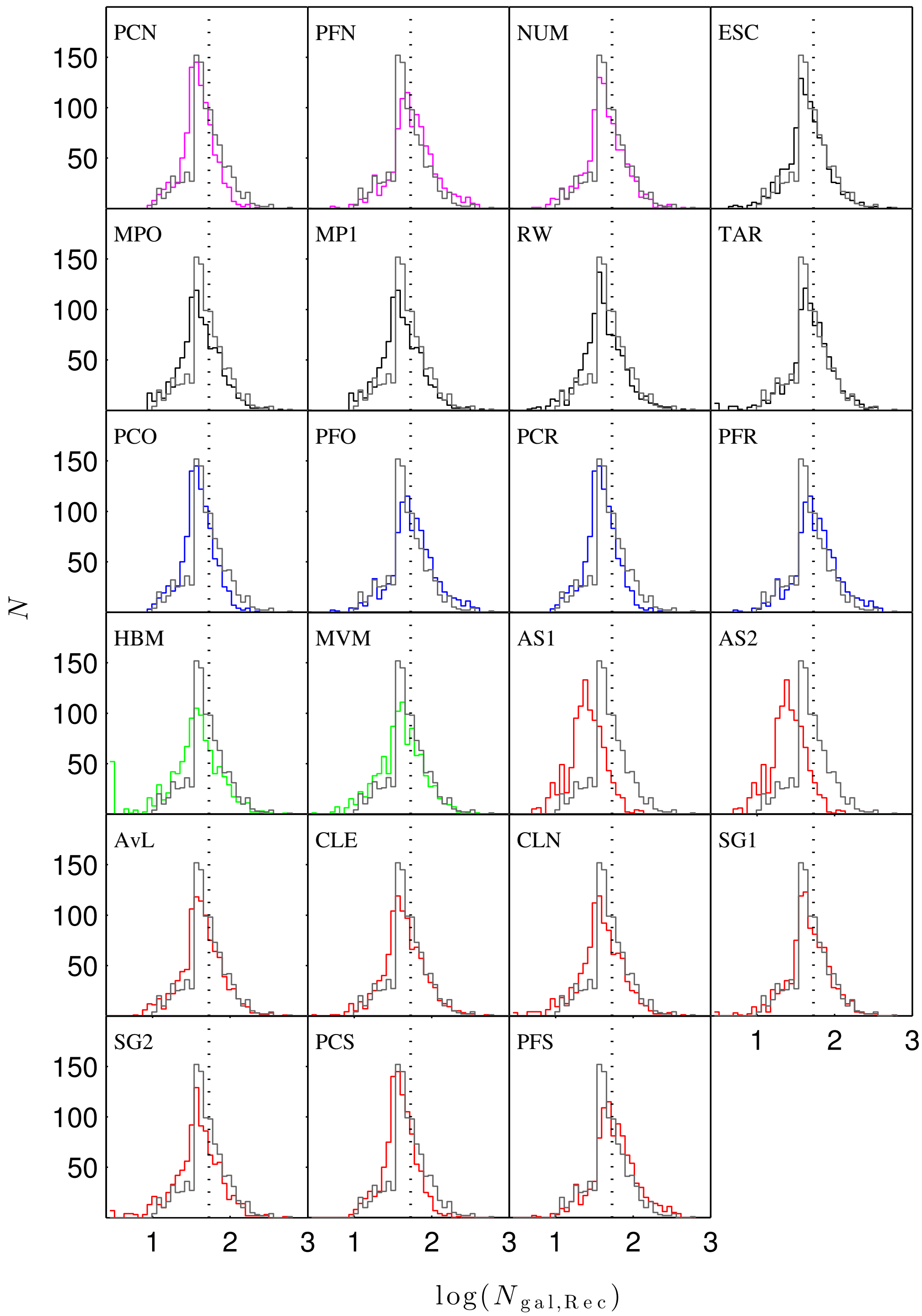

Figure C1. Distributions of the recovered number of galaxies associated with each group/cluster when the membership is not known. The grey distribution on each subplot represents the true richness distribution of the groups/clusters. The black dotted line presents the mean of this input richness distribution.

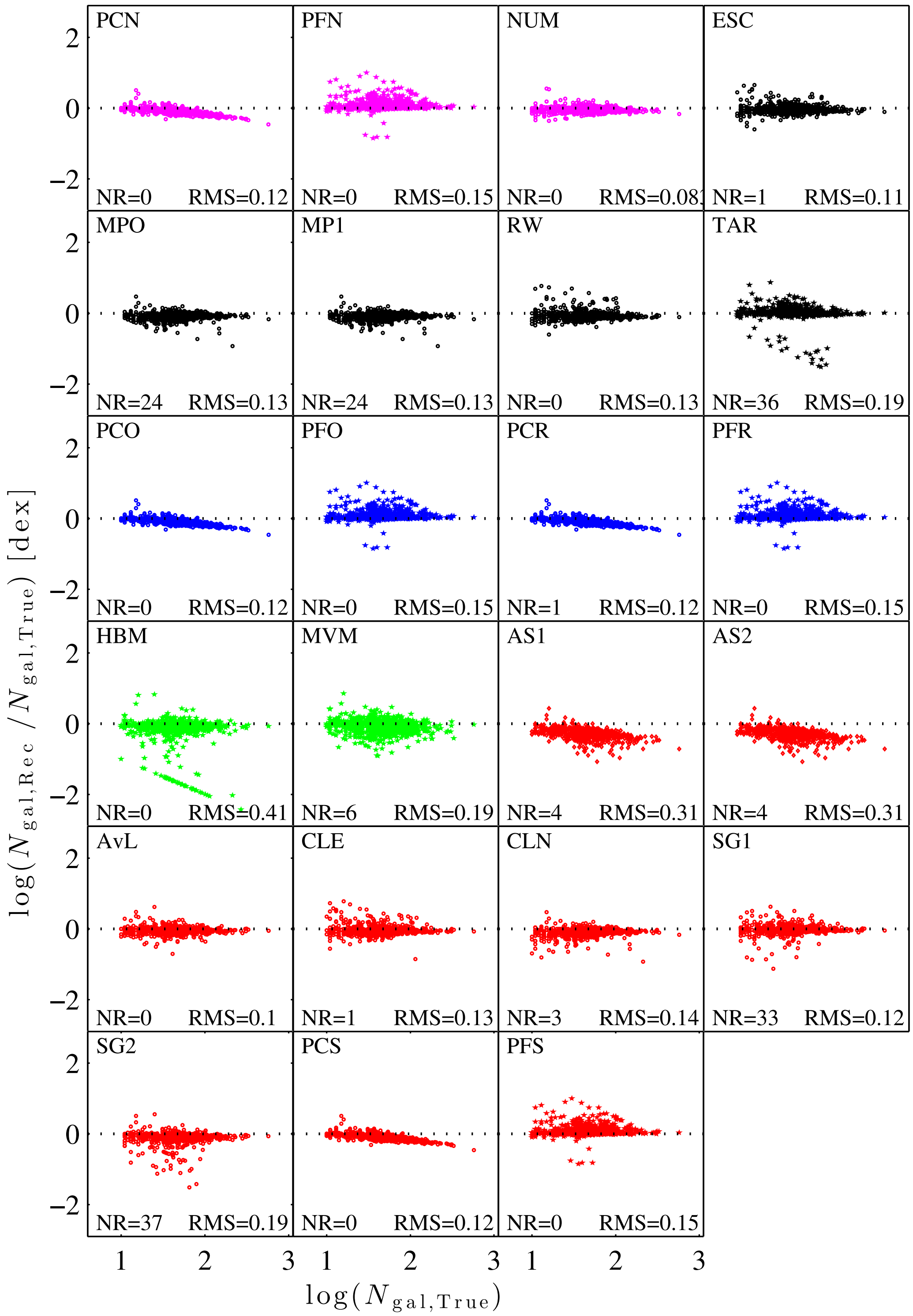

Figure C2. Richness bias versus true richness when the group/cluster membership is not known. The black dotted line represents a residual of zero. 'NR' in the legend represents groups/clusters that are not recovered because they are found to have very low ($< 10^{10}\mathrm{M}_{\odot}$) or zero mass.