Galaxy Cluster Mass Reconstruction Project – IV. Understanding the effects of imperfect membership on cluster mass estimation

R. Wojtak [ID],[1,2,3]* L. Old [ID],[4] G. A. Mamon,[5] F. R. Pearce,[6] R. de Carvalho,[7]
C. Sifón,[8] M. E. Gray,[6] R. A. Skibba,[9,10] D. Croton,[11] S. Bamford,[6] D. Gifford,[12]
A. von der Linden,[13] J. C. Muñoz-Cuartas,[14] V. Müller,[15] R. J. Pearson,[16] E. Rozo,[3,17]
E. Rykoff,[3] A. Saro,[18] T. Sepp[19] and E. Tempel [ID][15,20]

[1]*Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen, Denmark*
[2]*Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94305-4085, USA*
[3]*SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*
[4]*Department of Astronomy and Astrophysics, University of Toronto, Toronto, M5S 3H4, Canada*
[5]*Institut d'Astrophysique de Paris (UMR 7095: CNRS and Sorbonne Université), 98 bis Bd Arago, F-75014 Paris, France*
[6]*School of Physics and Astronomy, University of Nottingham, Nottingham NG7 2RD, UK*
[7]*Instituto Nacional de Pesquisas Espaciais, MCT, São José dos Campos, SP-CEP 12227-010, Brazil*
[8]*Department of Astrophysical Sciences, Peyton Hall, Princeton University, Princeton, NJ 08544, USA*
[9]*Science Communication Program, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA*
[10]*Freelance Science Journalist, San Diego, CA 92101, USA*
[11]*Centre for Astrophysics and Supercomputing, Swinburne University of Technology, PO Box 218, Hawthorn, VIC 3122, Australia*
[12]*Department of Astronomy, University of Michigan, 500 Church St, Ann Arbor, MI 48109-1107, USA*
[13]*Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794, USA*
[14]*Group for Computational Physics and Astrophysics, Instituto de Fisica, Universidad de Antioquia, Medellin, 050010, Colombia*
[15]*Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, D-14482 Potsdam, Germany*
[16]*School of Physics and Astronomy, University of Birmingham, Birmingham B15 2TT, UK*
[17]*Department of Physics, University of Arizona, Tucson, AZ 85721, USA*
[18]*INAF – Osservatorio Astronomico di Trieste, via G. B. Tiepolo 11, I-34143 Trieste, Italy*
[19]*Tarkvara Tehnoloogia Arenduskeskus (STACC), Tartu 51003, Estonia*
[20]*Tartu Observatory, University of Tartu, Observatooriumi 1, 61602 Tõravere, Estonia*

ABSTRACT

The primary difficulty in measuring dynamical masses of galaxy clusters from galaxy data lies in the separation between true cluster members from interloping galaxies along the line of sight. We study the impact of membership contamination and incompleteness on cluster mass estimates obtained with 25 commonly used techniques applied to nearly 1000 mock clusters with precise spectroscopic redshifts. We show that all methods overestimate or underestimate cluster masses when applied to contaminated or incomplete galaxy samples, respectively. This appears to be the main source of the intrinsic scatter in the mass scaling relation. Applying corrections based on a prior knowledge of contamination and incompleteness can reduce the scatter to the level of shot noise expected for poorly sampled clusters. We establish an empirical model quantifying the effect of imperfect membership on cluster mass estimation and discuss its universal and method-dependent features. We find that both imperfect membership and the response of the mass estimators depend on cluster mass, effectively causing a flattening of the estimated–true mass relation. Imperfect membership thus alters cluster counts determined from spectroscopic surveys, hence the cosmological parameters that depend on such counts.

Key words: methods: numerical – methods: statistical – galaxies: clusters: general – galaxies: haloes – galaxies: kinematics and dynamics – cosmology: observations.

* E-mail: wojtak@dark-cosmology.dk

1 INTRODUCTION

Virtually all studies of clusters of galaxies rely on accurate measurements of their mass. Cluster global masses (defined within some physical radius, such as the virial radius, within which the cluster should lie near dynamical equilibrium) and cluster mass profiles can be extracted by various astronomical techniques (see below). The primary challenge is that clusters are viewed in projection in the sky: our distance estimators are insufficient to enable the reconstruction of a precise three-dimensional (3D) mass distribution. This projected view of clusters affects all methods of cluster mass estimation, including the analysis of X-ray observations of the hot diffuse cluster gas (assumed to be in hydrostatic equilibrium) or of the statistical shear of galaxy shapes caused by weak gravitational lensing. Mass measurements of clusters using (optically selected) galaxies as kinematic tracers of their gravitational potential is particularly difficult, because of the small number of accurate measurements on one hand, and because of the inevitable confusion, caused by the Hubble flow, with other galaxies (some belonging to groups or clusters) up to 10–20 virial radii along the line of sight (LOS; Mamon, Biviano & Murante 2010). This makes most of the galaxy-based methods for measuring dynamical masses particularly prone to the effects of inaccurate membership (Wojtak et al. 2007). Despite these limitations imposed by imperfect membership, many galaxy-based methods of cluster mass estimation have been used for a variety of applications. Various case studies based on kinematics have provided numerous constraints on the cluster mass profiles, the mass–concentration relation and the anisotropy of galaxy orbits (Wojtak & Łokas 2010; Munari, Biviano & Mamon 2014; Mamon et al., in preparation).

Moreover, clusters are cosmological tools, since the evolution of the cluster mass function depends strongly on cosmological parameters. In the optical domain, these parameters were constrained using scaling relations of mass with richness for optical galaxy clusters detected in the Sloan Digital Sky Survey (SDSS) data (Rozo et al. 2010) or with velocity dispersion using clusters detected through the Sunyaev–Zel'dovich (Sunyaev & Zeldovich 1970) effect with the South Pole Telescope (Bocquet et al. 2015). Undoubtedly, the role of galaxy-based methods for constraining cluster mass profiles on one hand and cosmological parameters on the other will grow even more with the advent of large-scale imaging and spectroscopic cosmological surveys such as *Euclid*[1] and the Large Synoptic Survey Telescope (LSST).[2] Such 'cluster cosmology' will improve the present constraints on extensions of a standard Λ cold dark matter (ΛCDM) model, driven thus far primarily by X-ray observations of galaxy clusters, such as modified gravity (Cataneo et al. 2015), neutrino physics (Mantz et al. 2015), or the dark energy equation of state (Mantz et al. 2014).

Assigning cluster membership to galaxies observed in the cluster field can be performed in several ways. First, clusters are selected as galaxy concentrations on the sky or directly in redshift space. Moreover, thanks to the prominence of red galaxies in low-redshift clusters combined with the narrowness of the red sequence of galaxies in colour–luminosity diagrams, one can select cluster members appearing on a narrow red sequence in the colour–magnitude diagram (several narrow sequences may signify several clusters aligned along the LOS; Gladders & Yee 2000; Rykoff et al. 2014). LOS velocities provide extra information, and thanks to the Hubble flow,

obvious interlopers can be identified in several methods, such as (i) global 3σ clipping (Yahil & Vidal 1977), (ii) searching for gaps in the global LOS velocity distribution (Fadda et al. 1996), (iii) selecting maximum local absolute LOS velocities, (iv) using best estimates of the infall velocity (den Hartog & Katgert 1996), (v) the escape velocity identified again as LOS velocity gaps now called 'caustics' (Diaferio 1999), or (vi) a local $2.7\,\sigma_{LOS}$ clipping (Mamon, Biviano & Boué 2013).[3] These methods (except caustics) are iterative (first guessing a virial radius, and, for the latter one, assumed mass and velocity anisotropy profiles). But as mentioned above, all these methods suffer from inevitable contamination from galaxies lying along the LOS within 10 or 20 virial radii from the cluster, for which the Hubble flow is insufficient to push them beyond 2.7σ or 3σ in the LOS velocity distribution of the intrinsic cluster members. Stacking 100 haloes from a cosmological simulation, Mamon et al. (2010) found that as many as 23 per cent of objects lying within the virial cone (or cylinder) lie within the virial sphere (after filtering out the interlopers with a local $2.7\,\sigma_{LOS}$ criterion). A distinctive approach is based on the friends-of-friends (FoF) algorithm applied to projected phase space, but this cluster finder cannot reach perfect membership: increasing the linking lengths improves completeness at the expense of reliability and vice versa, making it impossible for any combination of linking lengths to jointly achieve over 83 per cent completeness and reliability for clusters of estimated mass above $10^{14}\,M_{\odot}$ (see Q_{local} in fig. 9 of Duarte & Mamon 2014). Moreover, with optimal linking lengths, over half the FoF clusters of estimated mass $10^{14}\,M_{\odot}$ turn out to be secondary fragments of more massive clusters, and with the most optimal cluster finders as those of Yang et al. (2007) and MAGGIE (Duarte & Mamon 2015) this fraction remains as high as ∼15 per cent (fig. 10 of Duarte & Mamon 2015).

For most methods, membership assignment cannot be fully separated from cluster mass measurement. Arguably all algorithms for selecting cluster members involve either mass-dependent cuts in projected phase space, e.g. 3σ clipping, or mass-dependent scales of models assigning a probabilistic membership (a galaxy observed at fixed physical distance from the cluster centre is more likely a cluster member if its host cluster is more massive than a baseline value). Consequently, membership is typically refined in iterative steps based on trial estimates of cluster mass. In this sense, both the mass estimate and the membership classification constitute the outputs of a self-consistent algorithm processing cluster data.

Cluster membership is commonly treated as a binary feature classifying galaxies into cluster members or interlopers. An increasingly popular alternative is a probabilistic approach, i.e. where each galaxy is assigned a membership probability, enabling determination of mass profile parameters in a self-consistently Bayesian fashion (e.g. Wojtak et al. 2009; Mamon et al. 2013; Wojtak & Mamon 2013), as well as the detection of the gravitational redshift effect (Wojtak, Hansen & Hjorth 2011; Sadeh, Feng & Lahav 2015; Jimeno et al. 2017) and the spatial anisotropy of galaxy kinematics (Skielboe et al. 2012).

Despite its growing popularity (e.g. Rykoff et al. 2014), the probabilistic approach is only justified when there is an underlying model for the interloping galaxies. For stacked clusters, quasi-uniform distributions of interloper LOS velocities can be extracted from cosmological simulations (as proposed by Mamon et al. 2010). But for individual clusters, the LOS velocity distribution of interlopers

---

[3] The factor 2.7 was optimized by Mamon et al. (2010) on a stack of haloes from a cosmological simulation.

is the sum of many quasi-Gaussians (each one for a different cluster along the LOS of the considered one).[4]

Despite a rich variety of mass estimators and techniques for selecting cluster members, the literature lacks extensive studies on how imperfect membership degrades cluster mass estimates, how scale-dependent properties of galaxy clusters (e.g. mass-dependent amount of substructure; de Carvalho et al. 2017; Roberts & Parker 2017; Old et al. 2018) affect selection of cluster members, and what solutions can mitigate all these effects. With the growing role of galaxy-based cluster mass estimations in future cosmological surveys, it is timely to quantify the impact of imperfect membership on mass scaling relations in order to envision the strategies to minimize it. The present work is the first step in addressing these problems in a systematic way.

We make use of the Galaxy Cluster Mass Reconstruction Project (GCMRP) data (Old et al. 2014, 2015) to quantify the effect of imperfect membership on cluster mass measurements. The data include estimates of mass and galaxy membership for nearly 1000 mock low-redshift galaxy clusters analysed by 25 algorithms, together with the true masses and galaxy memberships of these clusters. With these data, we can quantify, for each method, how mass bias and scatter depend on membership incompleteness and contamination.

Our study is a continuation of the GCMRP. A comprehensive framework of the project with built-in procedures of data blinding enables us not only to compare individual mass estimation methods exploiting a wide range of possible models, scaling relations, and numerous algorithms for handling various steps of data processing, but also to study generic properties of cluster mass measurements in optical observations. In Old et al. (2015), we quantified performances of all 25 methods and showed that those based on richness or abundance matching return most precise cluster mass estimates, irrespective of intrinsic assumptions of mock galaxy catalogues. Exploiting information on dynamical substructure in our follow-up study (Old et al. 2018), we demonstrated that all methods systematically overestimate masses of clusters with significant substructure. This bias turned out to affect low-mass clusters more strongly than high-mass counterparts.

This paper is organized as follows. In Section 2, we describe the mock observations and the cluster mass reconstruction techniques. In Section 3, we estimate the level of imperfect membership in terms of contamination and incompleteness, and quantify its effects on cluster mass estimates returned by each method, explaining our adopted methodology of data analysis. We study the mass dependence of imperfect membership and the response of the cluster mass estimators in Section 4, quantifying how imperfect membership modifies the mass scaling relations. We summarize and conclude in Section 5.

2 DATA AND METHODS

We base our analysis on tables, generated by the GCMRP (Old et al. 2015), of estimated mass and galaxy membership for 967 clusters, obtained with 25 different algorithms, together with the true masses and memberships.

---

[4]Maximizing the benefits of the probabilistic approach for individual clusters can be achieved with machine learning algorithms, whose first applications to cluster mass estimates resulted in noticeable improvements (Ntampaka et al. 2015, 2016).

Two mock galaxy catalogues were generated for the 25 algorithms: one using a semi-analytical model (SAM), and the other based on a halo occupation distribution (HOD) approach. In this study, we use the outputs from all 25 mass reconstruction methods, but only for HOD mock observations, because the true 3D cluster membership assigned to galaxies in the SAM catalogue does not conform with the assumed virial overdensity of $200\rho_c$ defining both cluster masses in the two galaxy catalogues and 3D cluster membership in the HOD catalogue (see Section 3.1).

The mock catalogues submitted to the algorithms were developed in three steps: (1) extracting dark matter haloes from a cosmological dark matter simulation; (2) extracting galaxies from the haloes; and (3) building a mock galaxy catalogue. We describe below each step in more detail (see Old et al. 2014, 2015 for more details).

2.1 Cosmological simulation

The mock observations were generated using the Bolshoi dissipationless cosmological simulation based on a flat ΛCDM cosmological model with the matter density parameter $\Omega_m = 0.27$, the rms of the density fluctuations $\sigma_8 = 0.82$, the tilt of the primordial power spectrum $n = 0.95$ and the dimensionless Hubble constant $h = 0.7$. The simulation follows the evolution of 2048^3 dark matter particles of mass $1.35 \times 10^8\,h^{-1}\,M_{\odot}$ within a box of side length $250\,h^{-1}$ Mpc (Klypin, Trujillo-Gomez & Primack 2011). It was run with the ART adaptive refinement code with a force resolution of $1\,h^{-1}$ kpc. The final halo catalogues are complete down to circular velocity of 50 km s^{-1} (corresponding to $M_{200c} \approx 1.3 \times 10^{10}\,h^{-1}\,M_{\odot}$ with ~100 particles per halo).

Dark matter haloes were found using the ROCKSTAR algorithm (Behroozi, Wechsler & Wu 2013). The halo finder operates in full six-dimensional (6D) phase space, enabling it to resolve more effectively haloes with spatially aligned centres. It has been shown to recover halo properties with high accuracy and returns halo catalogues that are broadly consistent with other halo finders (Knebe et al. 2011). Halo masses were calculated using a spherical overdensity threshold fixed at 200 times that of the critical density at the considered redshift. These overdensities were estimated considering all particles and substructures contained in the halo.

2.2 3D galaxy catalogues

A 3D galaxy catalogue was generated using an updated HOD model described in Skibba et al. (2006) and Skibba & Sheth (2009). In this approach, dark matter haloes are populated with galaxies whose luminosities and colours are assigned so that the simulated galaxy population approximately reproduces the observed luminosity function, colour–magnitude distribution, luminosity- and colour-dependent two-point correlation function measured from the SDSS. All galaxy properties are also assumed to be fully determined by the parent halo mass. During the course of the GCMRP, several improvements regarding phase-space distribution of satellite galaxies were developed and implemented. All modifications account for a number of effects that are present in realistic groups or clusters of galaxies, but were neglected in the first phase of the project, such as non-central positions of brightest cluster galaxies, central galaxy velocity bias (Skibba et al. 2011), or difference between the concentrations of dark matter versus satellite galaxies (Wojtak & Mamon 2013), as well as those of red versus blue galaxies (Collister & Lahav 2005; Cava et al. 2017). However, the HOD has two important simplifications: (1) haloes are truncated at the virial radius, whereas the one-halo term of galaxy clusters extends to beyond 10 virial

Table 1. Summary of the cluster mass reconstruction methods.

Method	Initial galaxy selection	Mass estimation	Type of data required	Colour info	Reference
(1)	(2)	(3)	(4)	(5)	(6)
PCN	Phase space	Richness	Spectroscopy	No	Pearson et al. (2015)
PFN[a]	FoF	Richness	Spectroscopy	No	Pearson et al. (2015)
NUM	Phase space	Richness	Spectroscopy	No	Mamon et al. (in prep.)
RM1	Red sequence	Richness	Central spectra	Yes	Rykoff et al. (2014)
RM2[a]	Red sequence	Richness	Central spectra	Yes	Rykoff et al. (2014)
ESC	Phase space	Phase space	Spectroscopy	No	Gifford & Miller (2013)
MPO	Phase space	Phase space	Spectroscopy	Yes	Mamon et al. (2013)
MP1	Phase space	Phase space	Spectroscopy	No	Mamon et al. (2013)
RW	Phase space	Phase space	Spectroscopy	No	Wojtak et al. (2009)
TAR[a]	FoF	Phase space	Spectroscopy	No	Tempel et al. (2014)
PCO	Phase space	Radius	Spectroscopy	No	Pearson et al. (2015)
PFO[a]	FoF	Radius	Spectroscopy	No	Pearson et al. (2015)
PCR	Phase space	Radius	Spectroscopy	No	Pearson et al. (2015)
PFR[a]	FoF	Radius	Spectroscopy	No	Pearson et al. (2015)
MVM[a]	FoF	Abundance matching	Spectroscopy	No	Muñoz-Cuartas & Müller (2012)
AS1	Red sequence	Velocity dispersion	Spectroscopy	No	Saro et al. (2013)
AS2	Red sequence	Velocity dispersion	Spectroscopy	No	Saro et al. (2013)
AvL	Phase space	Velocity dispersion	Spectroscopy	No	Von Der Linden et al. (2007)
CLE	Phase space	Velocity dispersion	Spectroscopy	No	Mamon et al. (2013)
CLN	Phase space	Velocity dispersion	Spectroscopy	No	Mamon et al. (2013)
SG1	Phase space	Velocity dispersion	Spectroscopy	No	Sifón et al. (2013)
SG2	Phase space	Velocity dispersion	Spectroscopy	No	Sifón et al. (2013)
SG3	Phase space	Velocity dispersion	Spectroscopy	No	Lopes et al. (2009)
PCS	Phase space	Velocity dispersion	Spectroscopy	No	Pearson et al. (2015)
PFS[a]	FoF	Velocity dispersion	Spectroscopy	No	Pearson et al. (2015)

Note. Columns are (1): acronym of algorithm; (2): physical property for initial selection of cluster members; (3): physical property for mass estimation from the initial membership; (4): observed properties; (5): use of colour information; (6): reference. The colours indicate five main classes of cluster mass estimation methods. See Table A1 in the appendix for more details on each method.

[a]Method did not use our initial object target list but rather performed an independent cluster search and matched the cluster locations at the end of their analysis.

radii (Trevisan, Mamon & Stalder 2017), and (2) galaxies within the haloes are not in local dynamical equilibrium (some of the algorithms assume this equilibrium). For a complete description of all implemented improvements we refer the reader to Old et al. (2015).

2.3 Mock galaxy catalogue

The light cone was produced using online tools of the Theoretical Astrophysical Observatory (TAO; Bernyk et al. 2016). It subtends 60° by 60° on the sky and covers redshift range of $0 < z < 0.15$. The galaxy sample included in the cone is complete down to a minimum r-band luminosity of $M_r = -19 + 5\log h$ in the input galaxy catalogue. The input provided to the algorithms consisted of the full galaxy catalogue (sky position and redshift), as well as the cluster centres (sky positions and redshifts) given by the locations of the brightest cluster galaxies. In other words, the GCMRP assumes that clusters are previously detected and their centres are known.

2.4 Mass reconstruction methods

The 25 algorithms of cluster mass reconstruction exhaust nearly all possible ways of inferring cluster masses from galaxy data. Table 1 presents a brief overview of all basic characteristics. Following Old et al. (2015), we divide the methods into five broad categories with respect to what data features are effectively used by the mass estimators: methods based on cluster richness (richness), radial scale of galaxy overdensity (radius), velocity dispersion of galaxies (velocity dispersion), their distribution in projected phase space (phase space), and abundance matching (equating the cumulative galaxy

luminosity function with the cumulative theoretical halo mass function). The main motivation of introducing these categories is to check for any possible subtrends in the analysed effects.

The algorithms for selecting cluster members form a part of the whole data processing. In most cases, they are combined with the mass estimators through various iterative procedures aimed at refining both mass estimates and cluster membership. Every method begins with an initial selection of galaxies, performed in several possible ways: colour–magnitude space (targeting red sequence galaxies), projected phase space (cuts in clustercentric distances and/or velocities), or by applying a FoF grouping algorithm. Table 1 summarizes the initial galaxy selection adopted by each method. More detailed information regarding this matter is provided in Table A1 and in the main articles of the GCMRP: Old et al. (2014, 2015).

Since no method can recover the exact masses of clusters, one can think of each method's output as a recovered versus true cluster mass scaling relation or equivalently a scaling relation between mass bias and true mass.

3 IMPERFECT MEMBERSHIP

3.1 Definitions and raw results

We adopt a simple and intuitive definition of the cluster membership and assume that all galaxies within the virial sphere are regarded as cluster members. We consider the virial radius that of a sphere enclosing a density which is 200 times the critical density of the universe at the redshift of the cluster. We quantify the observa-

tional selection of cluster members in terms of contamination C and incompleteness I. The former measures the relative fraction of interlopers in the selected galaxy sample, whereas the latter measures the fraction of true cluster members that are missing in the sample of selected galaxies, i.e.

$$C = \frac{N_{\text{sel,non-mem}}}{N_{\text{sel}}}, \tag{1}$$

$$I = \frac{N_{\text{non-sel,mem}}}{N_{\text{mem}}}, \tag{2}$$

where N_{mem} is the number of true cluster members, N_{sel} is the number of selected galaxies, $N_{\text{sel, non-mem}}$ is the number of selected interlopers (non-members), and $N_{\text{non-sel, mem}}$ is the number of missing cluster members. Contamination and incompleteness take values between 0 and 1. For a perfect membership assignment with no interlopers and all true cluster members included, $C = 0$ and $I = 0$.

The black contours in Fig. 1 show the distributions of contamination and incompleteness levels in galaxy samples selected by the 25 algorithms. The extent of the contours in the figure demonstrates that a typical galaxy sample returned by virtually every method is contaminated and incomplete to some degree. Perfect membership defined within the 3D virial sphere is in practice attainable only for a very small fraction of clusters which are most likely isolated. It is also apparent that contamination and incompleteness vary substantially between individual clusters. As we shall demonstrate below, this has a strong impact both on the accuracy and precision of cluster mass measurements.

Although Fig. 1 demonstrates quite substantial differences between all 25 methods, it is possible to draw some general conclusions. First, it is clear that cluster membership assignment is more precise when galaxy velocities are considered (e.g. the large extent of the contours of the photometric RM1 and RM2 methods compared to the other ones). Secondly, despite quite significant differences in how algorithms handle the selection of cluster members, many methods return strikingly similar galaxy samples. For example, methods NUM, AvL, ESC, and RW show fairly similar contours of membership quality, despite their differences in galaxy selection: $2.7\sigma_v$ for NUM, $2\sigma_v$ for AvL, and escape velocity for ESC and RW. Despite the fact that the former two methods shrink velocity envelopes by 30–50per cent with respect to escape velocity, they do not return galaxy samples with noticeably smaller contamination. In fact, all methods, regardless of the employed amplitude of the maximum velocity profile, are similarly affected by a number of clusters with highly contaminated galaxy samples (see the long horizontal branches at high contamination and small incompleteness). This demonstrates that attempting to reduce contamination by narrowing down the initial velocity range does not guarantee an improvement in cluster membership and should thus be treated with caution.

3.2 Effects of imperfect membership on cluster mass estimation

All methods of cluster mass estimation rely on the assumption of self-similarity in cluster data. The different mass estimation algorithms differ in the data feature that is utilized (richness, radius, velocity dispersion, projected phase-space distribution, abundance matching, etc.) and how it scales with cluster mass (empirical relations or fundamental principles such as the virial theorem or the Jeans equation of local dynamical equilibrium). Some methods employ more sophisticated models that effectively provide higher

order corrections to the underlying recovered versus true cluster mass scaling relations. Marginalization over nuisance parameters of these models (e.g. the shape of the mass density profile for methods based on richness, the velocity anisotropy in dynamical models) is expected to provide more accurate mass determinations.

The apparent scatter in contamination and incompleteness of galaxy samples selected by each method, highlighted in Fig. 1, breaks the assumption of self-similarity in the input cluster data. This should affect cluster mass estimation in two ways. First, this will increase the scatter in the recovered versus true cluster mass scaling relation, for example if recovered mass depends on richness. Secondly, if either galaxy selection or the response of the mass estimator depends on true cluster mass, then one should also expect an alteration of the slope of the recovered versus true cluster mass scaling relation.

To begin with, we neglect a possible mass dependence of membership assignments and mass estimators, and focus on the global relationship between cluster mass accuracy and imperfect membership. The main effects of imperfect membership can then be assessed by comparing the mean difference between recovered (estimated) and true log masses, $\log_{10}(M_{\text{rec}}/M_{\text{true}})$, as a function of contamination and incompleteness. The results are shown in Fig. 1 as colour maps.

The colour gradient apparent in nearly all panels of Fig. 1 demonstrates that virtually all methods tend to overestimate (underestimate) cluster masses with higher contamination (incompleteness) of selected galaxy samples. In most cases, the maps display distinct lines of degeneracy along which an overestimation due to contamination is compensated by an underestimation caused by incompleteness. This feature is quite intuitive for richness-based methods where the mass proxy is simply proportional to the number of cluster members. For methods utilizing information on velocities, similar trends appear to be naturally expected too if we realize that galaxy selection operates effectively in the tails of the velocity distribution of cluster members: any contamination or incompleteness of galaxy samples in this velocity regime automatically leads to an overestimate or underestimate of the velocity dispersion and consequently the cluster mass.

Comparing the black contours and the colour maps in Fig. 1, we can see that in most cases the maximum of the accuracy (yellow) does not coincide with the peak in the membership distribution. This indicates that cluster mass estimates may be biased even though galaxy samples are characterized by contamination and incompleteness typical of a given method. The mass bias corresponding to the peak of the distribution (the innermost isocontour in the figure) varies between the methods with median and scatter of -0.05 and 0.16 dex.

Old et al. (2015) showed that most methods return a fraction (with mean of 3 per cent) of mass estimates that deviate from the true masses by as much as a factor of 10. Highlighting these catastrophic cases as circles in Fig. 1, we identify them with incidents of extreme contamination or incompleteness.

3.3 Impact on cluster mass estimation: the model

Fig. 1 demonstrates that the mean logarithmic differences between the recovered and true masses depend on contamination and incompleteness, i.e.

$$\langle\log_{10}(M_{\text{rec}}/M_{\text{true}})\rangle = \mu(C, I). \tag{3}$$

The function $\mu(C, I)$ encapsulates our base model describing the primary effect of imperfect membership on cluster mass estimation.
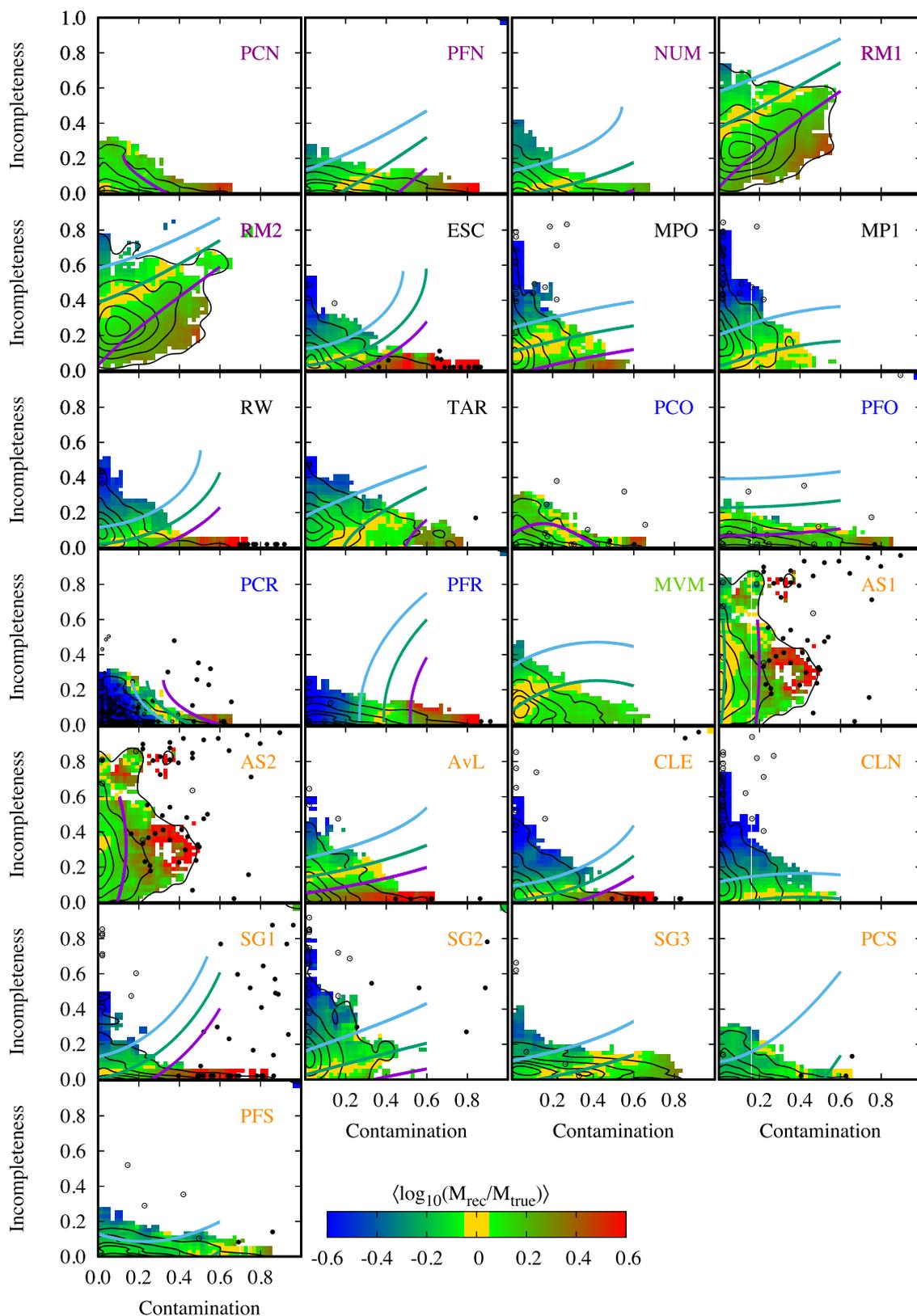
Figure 1. Mass bias and frequency as a function of contamination and incompleteness of galaxy samples selected by 25 different methods for nearly 1000 mock galaxy clusters. The black contours show the distribution of contamination and incompleteness (isodensity contours containing 10, 30, 50, 70, and 90 per cent of clusters). The colour maps show the mean mass bias, $\mu = \langle \log_{10}(M_{rec}/M_{true}) \rangle$, computed using a moving average method with a square-shaped 0.1×0.1 window function (yellow indicates $|\mu| < 0.05$). The filled and empty black circles show catastrophic cases with the estimated cluster masses, respectively, larger or smaller than the corresponding true cluster masses by a factor of 10. Pixels containing less than five galaxy clusters within the window were left blank. The coloured lines show the best-fitting mass bias model (equation 5) with purple, green, and cyan, respectively, corresponding to $\mu = 0.2, 0$, and -0.2 (all three present unless the range of μ does not include any of the three fixed values).

In the following, we make use of the output of each method to determine its empirical approximations.

Assuming a Gaussian distribution of variable $x = \log_{10}(M_{rec}/M_{true})$, the $\mu(C, I)$ of a given algorithm can be found by maximizing the following likelihood function:

$$L \propto \prod_i [(1 - w_c)G(x_i; \mu(C, I), \sigma) + w_c G(x_i; \mu(C, I), \sigma_c)], \quad (4)$$

where $G(x;; \mu, \sigma)$ is a Gaussian function of x with mean μ and variance σ^2 and where the product is over the 967 clusters analysed by the algorithm. The second term in the likelihood accounts for outliers whose relative fraction in the cluster sample is described by nuisance parameter w_c. We assume a flat distribution of outliers by fixing σ_c to a large value.

We optimize the parametric form of $\mu(C, I)$ by considering a series of truncated Taylor expansions about the mean contamination $\langle C \rangle$ and mean incompleteness $\langle I \rangle$. Employing the Bayesian information criterion (BIC) for model selection, we found that the following parametrization

$$\mu = \mu_0 + \mu_{C1}(C - \langle C \rangle) + \mu_{C2}(C - \langle C \rangle)^2$$
$$+\mu_{I1}(I - \langle I \rangle) + \mu_{I2}(I - \langle I \rangle)^2 \quad (5)$$

is sufficient to provide a satisfactory description of all data sets. The majority of the methods, including all higher ranking methods (see Old et al. 2015), do not support more complex models with cross or higher order terms (15 methods favour model 5 over a purely linear model with ΔBIC < -6, while only two methods favour the inclusion of the second-order cross term).

We determine best-fitting parameters and the corresponding confidence ranges using a Markov chain Monte Carlo (MCMC) technique based on the Hastings–Metropolis algorithm. We assume that the effective variance in the first Gaussian term of the likelihood (equation 4) consists of the intrinsic scatter and the contribution from shot (Poisson) noise, e.g.

$$\sigma^2 = \sigma_0^2 + \left(\frac{100}{N_{true}}\right)\sigma_1^2, \quad (6)$$

where N_{true} is the number of true cluster members, and where the irreducible scatter, σ_0, and the richness dependence of the scatter, σ_1, are both treated as additional free parameters in the MCMC analysis. As we shall see in Section 4.4, σ_0 and σ_1 are both strongly affected by imperfect membership.

For each method, it is possible to find a combination of contamination and incompleteness for which the mass overestimation due to contamination is fully compensated by its counterpart due to incompleteness. The cyan, green, and purple curved lines in Fig. 1 show the best-fitting models found for each method. The models are presented in the form of lines of constant μ with $\mu = 0, \pm 0.2$. Goodness of fit is addressed in more detail in Appendix B, where we show residuals for each method (see Fig. B1).

The linear terms of the model provide an accurate approximation in a narrow range of contamination and incompleteness about their mean values. Reducing the comparison between the methods to the level of linear coefficients μ_{C1} and μ_{I1}, as shown in Fig. 2. The figure shows that most mass estimators respond to changes in contamination in nearly the same way. Neglecting seven outliers lying outside of the $\pm 3\sigma$ range (AS1, AS2, PFR, PFS, PCO, PFO, and PCR), we find that the coefficients for the remaining methods can be described by a remarkably narrow distribution with mean of 0.50 and scatter 0.12. This distribution of μ_{C1} is consistent with the expectation of a simple, idealized richness-based mass estimator, where the recovered mass varies as the number of estimated cluster
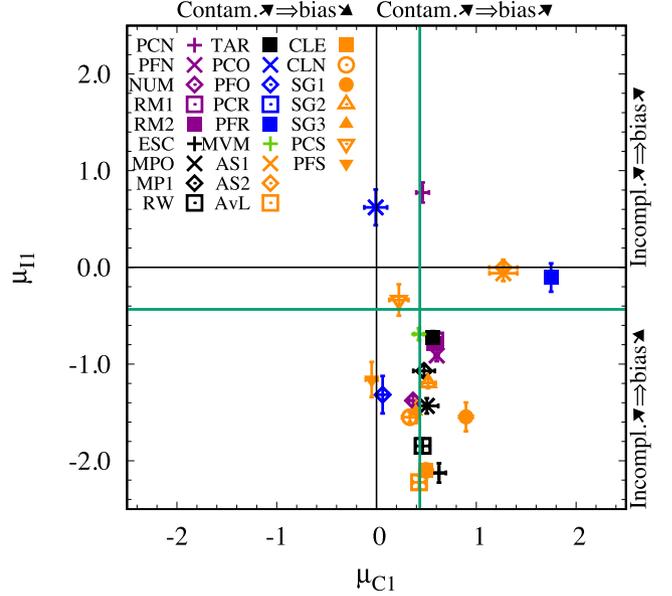

Figure 2. Increase of cluster mass biases of the 25 algorithms on the contamination and incompleteness of the galaxies selected by them in their mass analysis. The horizontal (vertical) axis shows the linear slope of the relationship between bias $\langle\log_{10}(M_{rec}/M_{true})\rangle$ and contamination (incompleteness) about its mean value, i.e. $\mu_{C1} = \delta\langle\log_{10}(M_{rec}/M_{true})\rangle/\delta C$ and $\mu_{I1} = \delta\langle\log_{10}(M_{rec}/M_{true})\rangle/\delta I$. These slopes are determined by fitting the model given by equation (5). The green vertical and horizontal lines indicate the coefficients expected for an idealized richness-based estimator employing the number of cluster members as a proxy for cluster mass. Method PCR lies out of the bounds of the plot with $\mu_{C1} \approx \mu_{I1} \approx 3$.

members, i.e. $M_{rec} \sim N_{sel, mem}$ and $M_{true} \sim N_{mem}$, leading to $\mu = \log_{10}(1 + C) \simeq C/(\ln 10) \simeq 0.43\,C$ (see the vertical green line in Fig. 2). This simple model appears to universally describe the effect of contamination on the mass estimation in three distinct groups of methods based on richness, velocity dispersion, and distribution in projected phase space. The seven outliers are among the methods with the lowest merit of the mass recovery accuracy (see Old et al. 2015).

The analogous simple, idealized, model for incompleteness, would lead to $\mu = \log_{10}(1 - I) \simeq -I/(\ln 10) \simeq -0.43\,I$. However, the nearly universal response of the mass estimators to changes in contamination does not have its analogy for incompleteness, as coefficient μ_{I1} ranges from -2.3 to 0 (neglecting three outliers with $\mu_{I1} > 0$). Most methods are characterized by coefficients $\mu_{I1} < -0.43$ (see the green horizontal line in Fig. 2). The lowest μ_{I1}, coefficients indicating the strongest dependence of the mass estimators on incompleteness, are found for methods based on velocity dispersion and on the distribution in projected phase space. This sensitivity of kinematical methods to incompleteness is not surprising: these methods select galaxies in velocity space; therefore, the missing cluster members are most likely to lie in the tails of the velocity distribution, in contrast to interlopers whose velocity distribution resemble quite closely the velocity distribution of true cluster members (Mamon et al. 2010). This in turn leads to a stronger effect on the velocity dispersion and the corresponding cluster mass estimate due to incompleteness compared to that due to contamination.
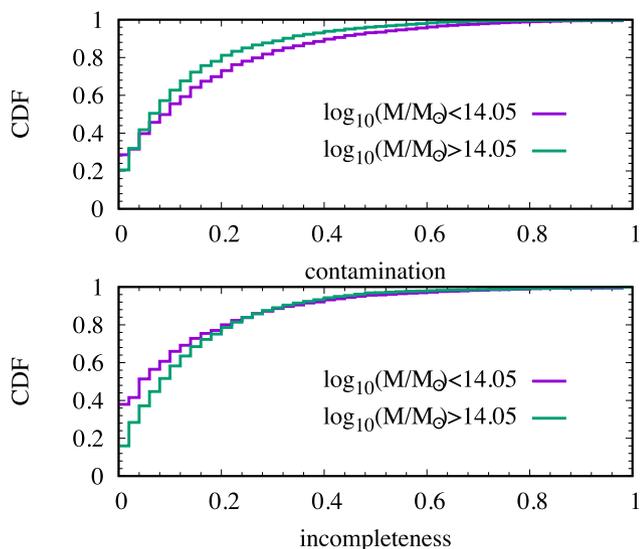
Figure 3. Cumulative distribution functions of contamination (top panel) and incompleteness (bottom panel) in the high-mass (cyan) and low-mass (purple) clusters (splitting the true cluster masses by the median value of $\log_{10}(M_{\rm true}/{\rm M_\odot}) = 14.05$). The probability combines results from all 25 methods and thus represents a global trend. Selected galaxy samples tend to be more incomplete for the most massive clusters and more contaminated for the least massive clusters.

4 MASS-DEPENDENT EFFECTS OF IMPERFECT MEMBERSHIP ON MASS ESTIMATION

Imperfect membership may give rise to a mass-dependent effect on cluster mass estimation in two different ways. First, the algorithms selecting cluster members may be scale dependent and return galaxy samples with contamination and incompleteness that depend on cluster mass. Second, there is no guarantee that the same level of contamination or incompleteness affects cluster mass measurements in the same way regardless of the actual cluster mass. This kind of mass dependence may occur due to the presence of an exponential cut-off in the mass function that breaks the self-similarity between how small galaxy groups perturb mass measurements of massive galaxy clusters and vice versa. In the following subsections, we seek to identify the extent to which these two aspects of mass-dependent imperfect membership underlie the performance of all methods.

4.1 Mass-dependent galaxy selection

To compare the effects of cluster mass on the relation between mass bias and membership quality, we compare contamination and incompleteness levels in two subsamples of, respectively, low and high true cluster mass, splitting at the median mass of $\log_{10}(M_{\rm true}/{\rm M_\odot}) = 14.05$. Fig. 3 shows the cumulative probability of contamination and incompleteness in the two groups of clusters. The calculation combines the data from all methods and thus the results demonstrate a global trend common to all 25 techniques (25×500 data points used for calculating each cumulative distribution). The distribution of contamination (incompleteness) in the two samples of galaxy clusters appears to be significantly different. The maximum difference between the cumulative probability distributions is 0.09 for contamination and 0.16 for incompleteness, while the upper limit required for rejection of the null hypothesis at level $p = 0.001$ of the Kolmogorov–Smirnov test is 0.025. Galaxy sam-
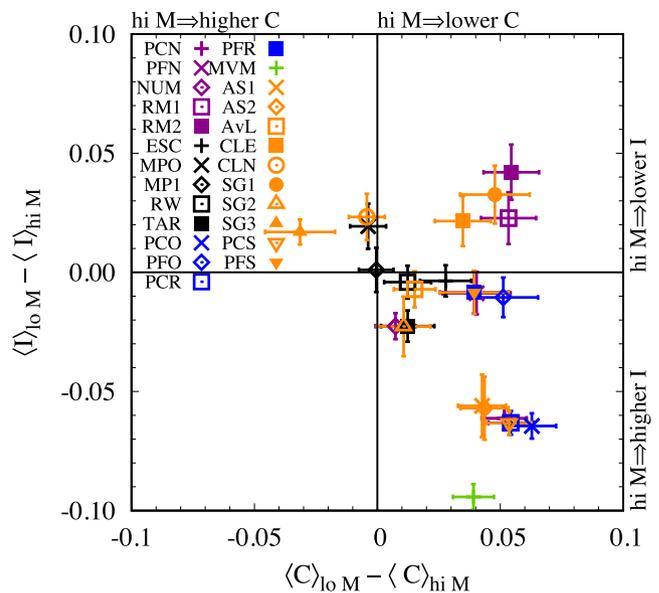

Figure 4. Differences in low versus high-mass (lo M, hi M) cluster contamination (C) and incompleteness (I), where the cluster masses are split by the median true mass of $\log_{10}(M_{\rm true}/{\rm M_\odot}) = 14.05$. Most methods return more contaminated samples for less massive clusters (21 methods) and more incomplete samples for more massive ones (17 methods).

ples selected from high-mass clusters tend to be less contaminated, but more incomplete.

The global trend shown in Fig. 3 reflects in large part the behaviour of each method analysed separately. This is demonstrated in Fig. 4 that shows the differences between the two groups of clusters in terms of the mean contamination and incompleteness calculated for each method (with errors estimated from bootstrapping). Except for four methods (MPO, MP1, CLN, and SG3), the mean contamination for the least massive clusters is clearly larger than that for the most massive ones at confidence levels ranging from 1σ up to 5σ. In agreement with Fig. 3, one also notices that galaxy samples selected from more massive clusters appear to be more incomplete, although the number of exceptions increases here to eight out of 25 methods. Among all five groups of the methods, those based on projected phase-space analysis appear to minimize the dependence of galaxy selection on cluster mass.

The mass dependence of cluster membership can also be demonstrated by using the galaxy sample selection function that is defined as the ratio of completeness to purity, where completeness $= 1 -$ incompleteness and purity $= 1 -$ contamination (by analogy to the cluster selection function; see e.g. Soares-Santos et al. 2011). Fig. C1 in the appendix shows the selection functions for all 25 methods compared to mean biases of cluster mass estimates. The strongest sensitivity of the mass function to cluster mass is unsurprisingly revealed by the methods with reversed dependences of contamination and incompleteness on cluster mass (bottom right-hand corner in Fig. 4), such as MVM, PCO, PCR, PCS, AS1, and AS2. As expected from the trends shown in Fig. 4, these methods are characterized by the selection functions that decrease with increasing cluster mass. Fig. C1 also points to another aspect of interconnections between membership and cluster mass measurements. The contours of constant bias appear not to coincide precisely with the contours of constant selection function (see e.g. NUM, PCO, and MVM as extreme examples). This implies that some methods may return cluster mass estimates with mass-dependent bias even

though selection of cluster members remains strictly independent of cluster mass. We quantify this effect in a more rigorous way in the following section.

4.2 Mass-dependent effects on mass estimation

We quantify whether the impact of imperfect membership on the mass bias depends on the cluster mass by analysing the output data of all methods using the following generalization of our base model (5):

$$
\begin{aligned}
\mu = {} & \mu_0 + (\alpha_0 - 1)\log_{10}(M_{\text{true}}/M_0) \\
& + (M_{\text{true}}/M_0)^{\alpha_C}\left[\mu_{C1}(C - \langle C \rangle) + \mu_{C2}(C - \langle C \rangle)^2\right] \\
& + (M_{\text{true}}/M_0)^{\alpha_I}\left[\mu_{I1}(I - \langle I \rangle) + \mu_{I2}(I - \langle I \rangle)^2\right].
\end{aligned} \tag{7}
$$

The three new mass-dependent terms account for a potential dependence of three different components of the base model on cluster mass. The first new term (proportional to $\alpha_0 - 1$) describes an alteration of the slope of the scaling relation between the recovered and true cluster mass. When ignoring all terms dependent on contamination and incompleteness, the $M_{\text{rec}} \propto M_{\text{true}}^{\alpha_0}$ scaling relation absorbs all effects of imperfect membership. As measured by Old et al. (2015), the slope α_0 in this case varies quite substantially between the methods with mean and scatter of 0.97 and 0.19. In our approach, α_0 is determined simultaneously with all parameters of the model accounting for imperfect membership. Comparing its values to those from Old et al. (2015) will demonstrate if and how contamination and incompleteness modifies the $M_{\text{rec}} - M_{\text{true}}$ scaling relation. The second and third new terms describe a mass-dependent response of mass estimators to contamination and incompleteness. We choose a power-law ansatz in order to avoid a sign change for the expressions in square brackets and minimize degeneracies with parameters of the base model. The signs of the slopes distinguish between whether the effect of contamination or incompleteness is stronger in more massive clusters (positive signs) or less massive ones (negative signs). The pivot mass M_0 in all three new terms is set at the median mass of the whole cluster sample, i.e. $\log_{10}M_{\text{true}} = 14.05$. We find best-fitting parameters following the same approach as outlined in Section 3.2. Table D1 in the appendix shows the results for each method and compares to those obtained for a simplistic model neglecting any dependence on contamination or incompleteness, i.e. $\mu_{C1} = \mu_{I1} = \mu_{C2} = \mu_{I2} = 0$. The model corrected for imperfect membership fits the mass biases much better. Even with the six extra parameters, it results in a much smaller BIC, with $\Delta\text{BIC} \ll -10$ for all methods, indicating very strong evidence for this more complex model.

Fig. 5 shows the constraints on the mass-dependent effects of contamination and incompleteness, α_C and α_I, respectively, on mass bias, as determined for each method using equations (4), (6), and (7). We find that there is a clear excess of methods with $\alpha_C < 0$, i.e. methods for which cluster mass overestimation due to contamination appears to be stronger for less massive clusters. On the other hand, the distribution of α_I does not reveal any significant asymmetry and thus any generic trend in mass dependence of bias due to incompleteness.

4.3 Effects of imperfect membership on the slope of mass scaling relations

The slope of the $M_{\text{rec}} - M_{\text{true}}$ scaling relation remains independent of contamination and incompleteness if both the selection of cluster members and the response of the underlying mass estimator to imperfect membership are independent of cluster mass. Our study
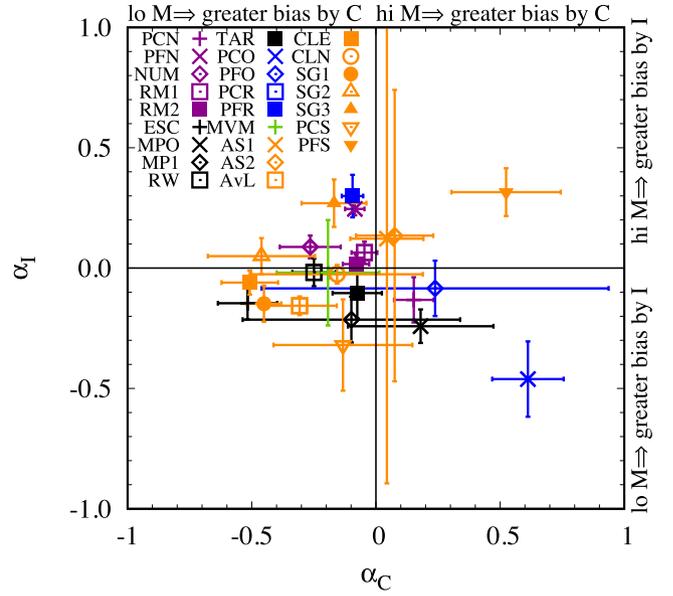
Figure 5. Effects of true cluster mass on the variation of mass bias with contamination and incompleteness. The mass dependence is assumed to be a power-law function of true cluster mass with slopes α_C and α_I corresponding to the effect of contamination and incompleteness, respectively (equation 7). There is a clear of excess methods for which cluster mass overestimation due to contamination tend to be stronger for low-mass (lo M) clusters. Method PCR lies out of the bounds of the plot with $\alpha_C = -0.7 \pm 0.1$ and $\alpha_I = -1.5 \pm 0.3$.

demonstrates that this condition is not satisfied in general. Therefore, one may expect that the effective slope of the mass scaling relations can be modified to some extent by how different methods select cluster members. We quantify this effect by comparing the slopes measured in two modes: ignoring any dependence of $\log_{10}(M_{\text{rec}}/M_{\text{true}})$ on membership ($\mu_{C1} = \mu_{C2} = \mu_{I1} = \mu_{I2} = 0$ and free α_0, σ_0, σ_1) and including a complete (generalized) model describing effects of imperfect membership, as given by equation (7). The effective slopes measured in the first mode (the same as measured and discussed in Old et al. 2015) depend both on intrinsic properties of the mass estimators and imperfect membership (contamination and incompleteness) of selected galaxy samples, while the slopes measured in the second mode reflect primarily performance of the mass estimators. Therefore, comparing the slopes from the two modes is expected to extract a genuine effect of imperfect membership on the slopes of the mass scaling relations.

Fig. 6 shows the constraints on slope α_0 measured in the two modes described above. The results clearly demonstrate that imperfect membership gives rise to a flattening of the $M_{\text{rec}} - M_{\text{true}}$ scaling relations for most methods. We find that the relationship between the two slopes is well approximated by a linear model with slope 1.01 ± 0.11, intercept 0.084 ± 0.019, and scatter 0.081 ± 0.017 (see the dashed green line). This simple model implies that a typical reduction of the slope α_0 due to imperfect membership is equal to 0.084 and is independent of the slope found by neglecting the effects of imperfect membership.

4.4 Effects of imperfect membership on the scatter of mass scaling relations

We follow the same approach as described in the previous subsection to study the effects of imperfect membership on the scatter in the $M_{\text{rec}} - M_{\text{true}}$ scaling relations. We measure both free parameters
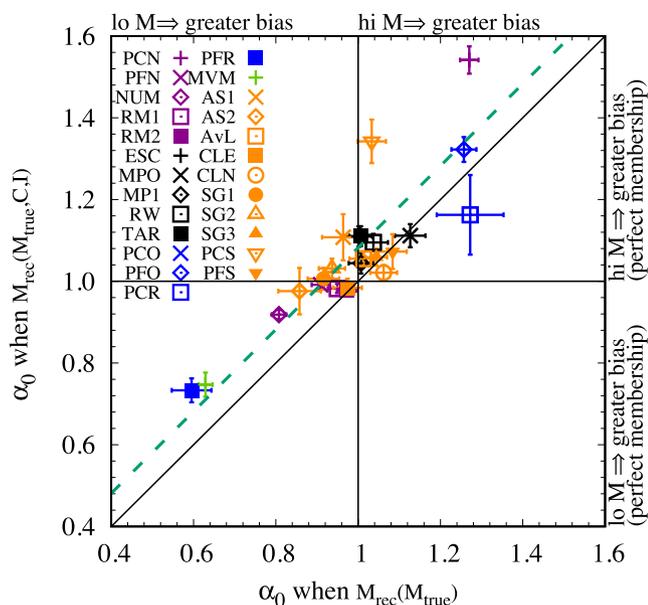
Figure 6. Effects of imperfect membership on the slope of the mass scaling relation $M_{\rm rec} \sim M_{\rm true}^{\alpha_0}$. The slope is measured for each method in two cases: neglecting any prior information on membership (horizontal axis) and including our full model accounting for effects of contamination and incompleteness on cluster mass estimates (equation 7, vertical axis). The dashed line shows a best-fitting linear model that signifies that imperfect membership tends to flatten the $M_{\rm rec}-M_{\rm true}$ relation (greater bias for low-mass – 'lo M' – clusters) reducing its slope on average by 0.084.

of the effective scatter given by equation (6) in two modes: ignoring any dependence on membership ($\mu_{C1} = \mu_{C2} = \mu_{I1} = \mu_{I2} = 0$ and free α_0, σ_0, σ_1) and including a complete (generalized) model describing effects of imperfect membership, as described by equation (7). Comparing values of both σ_0 and σ_1 determined in these two modes quantifies the contribution of imperfect membership to the total scatter in the mass scaling relations.

As demonstrated in Fig. 7, imperfect membership increases both the intrinsic and Poisson-like scatter substantially. Applying corrections due to imperfect membership can potentially bring down the Poisson-like scatter to theoretically expected levels. This is illustrated in the bottom panel, which shows theoretical lower limits of the Poisson-like scatter for two basic mass estimators based on richness or velocity dispersion (magenta and orange symbols, respectively), i.e. $M_{\rm true} \propto N_{\rm mem}$ and $M_{\rm true} \propto \sigma_v^3$. The former is the prediction for richness-based methods, assuming $\alpha_0 = 1$ as is roughly the case for most methods, as seen in Fig. 6, leading to $\sigma_1 = 1/(10 \ln 10) \simeq 0.043$. The latter is the prediction for methods utilizing kinematics (velocity dispersion or distribution in projected phase space), and given that the uncertainty on the standard deviation for a Gaussian distribution is $\epsilon(\sigma) = \sigma/\sqrt{2N}$, the scatter on $\mu \propto \log_{10} \sigma_v^3$ is $3/(\sqrt{2N} \ln 10)$, leading to $\sigma_1 = 0.3/(\sqrt{2N} \ln 10) \simeq 0.092$. This assumes that any velocity bias of galaxies tracing the gravitational potential is independent of mass (as found by Munari et al. 2013, but disputed by Old, Gray & Pearce 2013). All richness-based methods except PCN have their Poisson scatter term, σ_1, as low as the theoretical limit of $0.1/(\ln 10)$. Nearly all the velocity dispersion and projected phase-space methods (except AS1 and AS2; orange and black symbols, respectively) have their Poisson scatter term fairly close to the theoretical limit of $0.3/(\sqrt{2N} \ln 10)$.
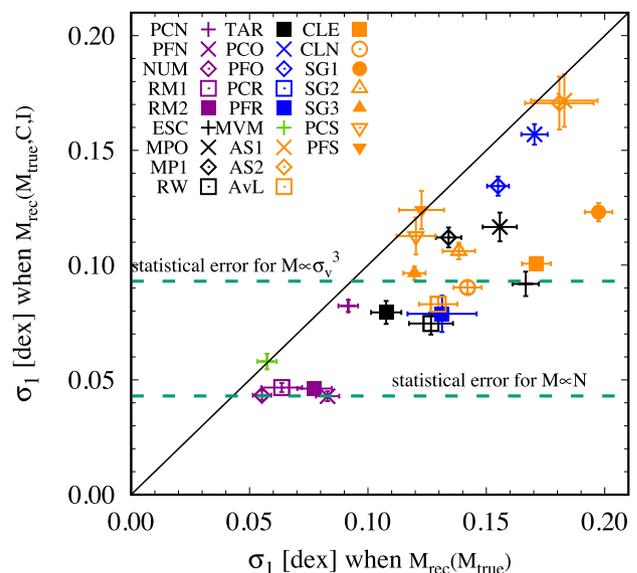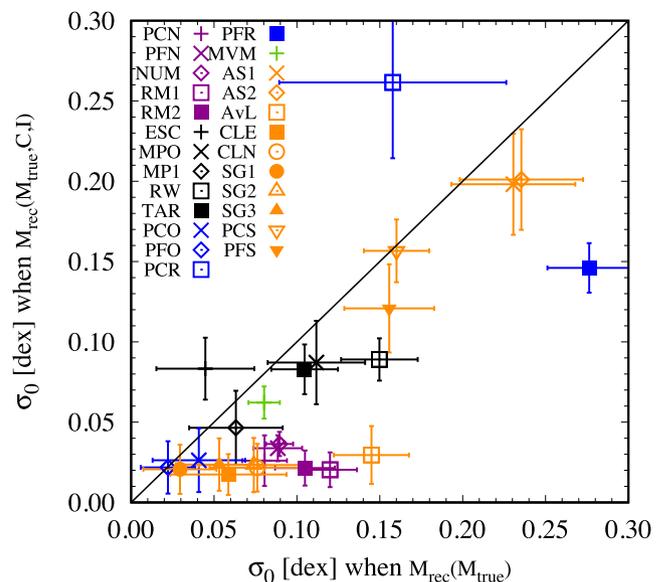

Figure 7. Effects of imperfect membership on the Poisson-like scatter (top panel) and the irreducible scatter (bottom panel), as defined by equation (6), in the $M_{\rm rec}-M_{\rm true}$ scaling relations. The two scatter parameters are measured in two modes: neglecting any prior information on membership (horizontal axis) and including a full model accounting for effects of contamination and incompleteness on the cluster mass estimates (vertical axis). The dashed lines in the bottom panel indicate two characteristic levels of the Poisson-like scatter for richness-based and velocity dispersion-based methods. The results demonstrate that imperfect membership increases substantially both the intrinsic and Poisson-like scatter.

Imperfect membership also appears to be a primary source of the intrinsic (irreducible) scatter, σ_0. For 12 methods, the intrinsic scatter becomes negligible ($\sigma_0 < 0.03$) when mass estimates are corrected for imperfect membership. But several methods (AS1, AS2, PCR, PCS, PFR, and PFS) have an important intrinsic scatter ($\sigma_0 > 0.1$) after correction for imperfect membership. Except for PCS, these methods appear to be substantially dominated by intrinsic errors of the mass estimators and are amongst those with the lowest ranks of the mass recovery accuracy and the rms difference between the recovered and true log cluster mass larger than

0.45 dex, compared to the rms of ∼0.20 dex for the best methods (see table 2 in Old et al. 2015).

5 SUMMARY AND CONCLUSIONS

We used mock observations of galaxy clusters in the optical band (photometric and spectroscopic data) to study the impact of imperfect membership in selected galaxy samples, as quantified by contamination and incompleteness, on cluster mass estimates based on 25 different methods employing various techniques of galaxy selection and dynamical mass estimation. The mock catalogue was generated using an HOD approach with some improvements to emulate more realistically observations of galaxy clusters in terms of the colour distribution of galaxies, the phase-space distribution of satellites, miscentring of brightest cluster galaxies, and several other properties (see Old et al. 2015). Although each of the 25 methods considered in this study has its own peculiarities, the methods can be grouped into four broad categories with respect to what part of the available data is utilized to select cluster members and estimate cluster masses: methods based on richness, galaxy positions, velocity dispersion, and projected phase-space distributions.

We demonstrated that contamination and incompleteness give rise to, respectively, overestimation and underestimation of the measured cluster masses. This general rule holds for nearly all methods and for all four categories of mass estimators. For each method, it is possible to find a combination of contamination and incompleteness for which the mass overestimation due to contamination is fully compensated by its counterpart due to incompleteness (green lines in Fig. 1). The mass estimation accuracy given by $\langle \log_{10}(M_{rec}/M_{true}) \rangle$ evaluated for the mean contamination and incompleteness does not vanish and thus sets an irreducible bias for each mass estimator. Using a linear model for describing the impact of imperfect membership about the mean contamination and incompleteness, we find that all methods exhibit (Fig. 2) the same dependence on contamination with $\log_{10}(M_{rec}/M_{true}) \sim (0.50 \pm 0.12)C$, consistent with $M_{rec} \sim (1 + C)$ expected for a simple, idealized richness-based method with $M_{rec} \sim N_{mem}$. On the other hand, the analogous dependence on incompleteness is not universal, with $\log_{10}(M_{rec}/M_{true})$ in the range of $-2.0\,I$ to $-0.5\,I$, with the highest sensitivity to incompleteness for methods based on kinematics (velocity dispersion and projected phase-space analysis).

Imperfect membership modifies the M_{rec}–M_{true} scaling relation in two respects. The primary effect, demonstrated also in other studies (see e.g. Saro et al. 2013), is an increase of scatter (Fig. 7). We found that this affects both the intrinsic scatter and the Poisson-like scatter scaling with $1/N_{mem}^{1/2}$, where N_{mem} is the number of true cluster members. Secondly, due to a mass-dependent selection of cluster members (Fig. 3) and a mass-dependent response of the cluster mass estimators to imperfect membership (Fig. 5), the M_{rec}–M_{true} relation becomes flatter than that based on the assumption of fully self-similar samples of galaxies selected as cluster members (Fig. 6). This flattening arises from both a higher contamination of galaxy samples (Fig. 3) and a higher sensitivity of cluster mass estimators to contamination (Fig. 4) for less massive systems.

Fig. 8 schematically illustrates how imperfect membership affects the M_{rec}–M_{true} scaling relation. The quantitative description of effects of imperfect membership on the mass scaling relations is based on a specific choice of the prior distribution of cluster masses with a complete sampling of the underlying mass function at $\log_{10}M_{true} \gtrsim 14$ and a smooth cut-off at low masses (see the histogram in Fig. 8). We expect that exact values of parameters describing the M_{rec}–M_{true} relation and their response to imperfect
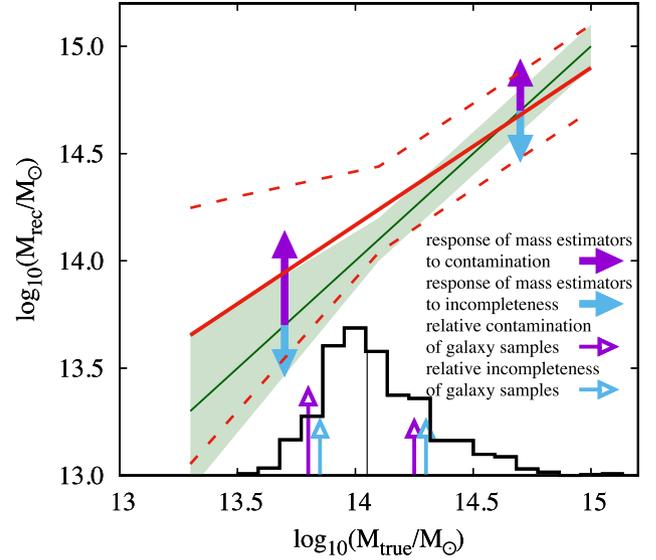

Figure 8. Illustration of the effects of imperfect membership on the mass scaling relation. The shaded region represents the scaling relation unaffected by imperfect membership, with the width rendering scatter dominated by shot noise at low masses and intrinsic scatter at high masses. The thick solid and dashed lines show the alteration of the relation due to typical contamination and incompleteness of selected galaxy samples. Imperfect membership flattens the relation and increases both types of scatter. The arrows show the relative impact of two main effects: the response of the mass estimator to imperfect membership and the typical levels of contamination/incompleteness in selected galaxy samples. The apparent asymmetry of the two effects at different mass scales is responsible for flattening of the M_{rec}–M_{true} relation. The black histogram shows the distribution of true cluster masses in the catalogue, with the median mass indicated by the thin black vertical line.

membership may change when providing cluster samples that are more complete at lower masses. Including more systems at low masses, however, can only enhance the effects illustrated in Fig. 8, in particular the flattening of the mass scaling relation and the increase of the Poisson-like scatter.

Our results show that improvement in assigning cluster membership can substantially reduce the scatter in the mass scaling relations for virtually all methods. As demonstrated in Fig. 7, accounting for imperfect membership turns the intrinsic scatter into a subdominant contribution for most techniques. The exact values of the intrinsic scatter may be underestimated due to some simplifying assumptions adopted in the HOD approach to generating galaxy catalogues. We expect that an additional contribution to the intrinsic scatter may arise from effects that are not accounted for in the HOD catalogue, e.g. intrinsic shapes of galaxy clusters, the orbital anisotropy, or realistic substructure. The impact of these effects on the effective scatter can be quantified using mock observations based on more realistic models describing connections between galaxy properties and the underlying dark matter haloes. For example, Old et al. (2018) utilized mock observations (a part of the GCMRP; Old et al. 2015) generated with the Semi-Analytic Galaxy Evolution (SAGE) galaxy formation model (Croton et al. 2006) and measured an average difference of 0.054 dex between log masses of galaxy clusters with or without substructures. Assuming a Gaussian distribution of their effect on mass estimates, dynamical substructures are then a source of a 0.03 dex intrinsic scatter in the mass scaling relations based on galaxies.

The mass scaling relations studied in our work are of particular importance for inferring cosmological constraints from the clus-
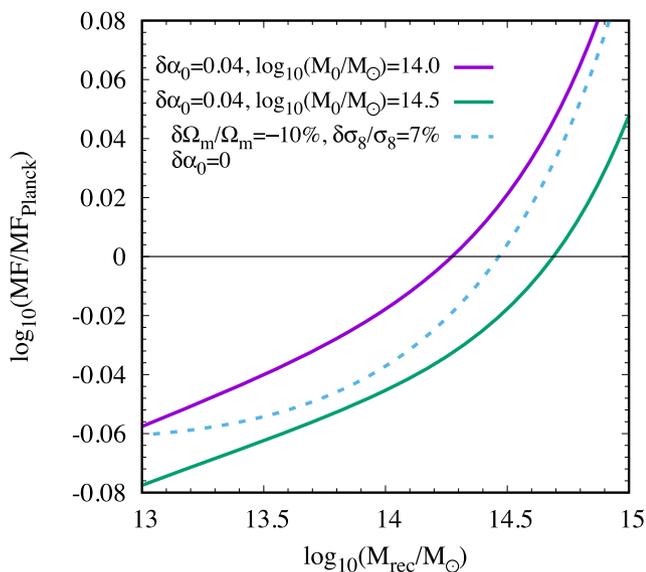
Figure 9. The impact of neglecting the flattening of the $M_{rec}-M_{true}$ relation due to mass-dependent imperfect membership on the observationally measured mass function (MF). It is assumed that the $M_{rec} \propto M_{true}^{\alpha_0}$ relation is determined in a very narrow mass range about M_0, leading to a bias $\delta\alpha_0$ when extrapolating the relation to lower and higher masses. The dashed line shows the mass function for modifications of cosmological parameters that approximately compensate the effects of mass-dependent imperfect membership. The mass functions are calculated relative to a fiducial model given by the *Planck* cosmology (MF_{Planck}).

ter abundance measured in upcoming cosmological surveys such as *Euclid* or LSST. From this point of view, our results shed light on some aspects of observational strategies for cluster cosmology. First of all, imperfect membership appears to be the main and for some methods the only source of intrinsic scatter in the mass scaling relation. Therefore, it is clear that further development of most methods for cluster mass estimation shall prioritize improvement of algorithms assigning cluster membership. Secondly, our study shows that the effective slope of the mass scaling relation does not only reflect the intrinsic performance of the mass estimator, but is also affected by mass-dependent effects of imperfect membership. Bearing this in mind, it becomes clear that a self-consistent comparison between observational and simulation-based mass scaling relations requires that both mock and real cluster data are nearly the same in terms of the mass range and they are analysed using the same algorithms for assigning membership and estimating cluster masses.

Uncorrected flattening of the $M_{rec}-M_{true}$ relation due to mass-dependent effects of imperfect membership may distort the mass function reproduced from observations and consequently bias the measurement of cosmological parameters. Fig. 9 shows the magnitude of this effect in a simplified scenario where we assume that the $M_{rec}-M_{true}$ relation is determined in a very narrow mass range about M_0 and its extrapolation to low/high masses results in adopting too high a slope for the actual $M_{rec}-M_{true}$ relation valid in the wider mass range. Using only a half of the mean increase of α_0 shown in Fig. 6, we find that the observationally reconstructed mass function is suppressed at low masses and amplified at high masses by $\sim$0.05 dex (12 per cent) compared to the fiducial model given by a fitting function from Tinker et al. (2008) and the *Planck* cosmology (Planck Collaboration XIII 2016). As shown by the dashed line, this in turn can degrade the accuracy of cosmological inference with Ω_m

biased down by $\sim$10 per cent and σ_8 biased up by $\sim$7 per cent. Interestingly, a comparable bias due to unaccounted projection effects for richness-based methods to measure cosmological parameters from the cluster mass function was demonstrated by Costanzi et al. (2018). We think that our estimated biases shall be regarded as upper limits, since realistic calibrations of the $M_{rec}-M_{true}$ relation can be performed in a mass range comparable to that used for measuring the mass function, alleviating the effects of mass-dependent imperfect membership. However, the magnitude of the obtained biases in cosmological parameters demonstrates the importance of accounting for imperfect membership on all mass scales in a robust way, especially when cluster cosmology from future surveys such as *Euclid* is expected to reach a sub-per cent precision (Sartoris et al. 2016). Another complication may arise from a possible redshift dependence of imperfect membership. This is not considered in our work, but it is definitely worth studying using high-redshift mock observations of galaxy clusters.

Although our study is solely based on the HOD mock galaxy catalogue of the GCMRP, we confirm that all main effects unveiled by the HOD data are also readily visible in the twin galaxy catalogue produced by the SAM model (Old et al. 2014). In particular, we find that the $M_{rec} \propto M_{true}^{\alpha_0}$ relation is affected by a comparable flattening due to a net effect of mass-dependent galaxy selection and sensitivity of the mass estimators to imperfect membership, with a mean reduction of 0.075 ± 0.032 in α_0. Consistency between results obtained from the two mock galaxy catalogues and the fact of using a wide variety of cluster mass reconstruction methods corroborate the generality of the conclusions reached in this study.

ACKNOWLEDGEMENTS

REFERENCES

Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013, ApJ, 762, 109
Bernyk M. et al., 2016, ApJS, 223, 9
Bocquet S. et al., 2015, ApJ, 799, 214
Cataneo M. et al., 2015, Phys. Rev. D, 92, 044009
Cava A. et al., 2017, A&A, 606, A108
Collister A. A., Lahav O., 2005, MNRAS, 361, 415
Costanzi M. et al., 2018, preprint (arXiv:1807.07072)
Croton D. J. et al., 2006, MNRAS, 365, 11
de Carvalho R. R., Ribeiro A. L. B., Stalder D. H., Rosa R. R., Costa A. P.,
    Moura T. C., 2017, AJ, 154, 96
den Hartog R., Katgert P., 1996, MNRAS, 279, 349
Diaferio A., 1999, MNRAS, 309, 610
Duarte M., Mamon G. A., 2014, MNRAS, 440, 1763
Duarte M., Mamon G. A., 2015, MNRAS, 453, 3848

Fadda D., Girardi M., Giuricin G., Mardirossian F., Mezzetti M., 1996, ApJ, 473, 670

Gifford D., Miller C. J., 2013, ApJ, 768, L32

Gladders M. D., Yee H. K. C., 2000, AJ, 120, 2148

Jimeno P., Broadhurst T., Lazkoz R., Angulo R., Diego J.-M., Umetsu K., Chu M.-c., 2017, MNRAS, 466, 2658

Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ, 740, 102

Knebe A. et al., 2011, MNRAS, 415, 2293

Lopes P. A. A., de Carvalho R. R., Kohl-Moreira J. L., Jones C., 2009, MNRAS, 392, 135

Mamon G. A., Biviano A., Murante G., 2010, A&A, 520, A30

Mamon G. A., Biviano A., Boué G., 2013, MNRAS, 429, 3079

Mantz A. B., Allen S. W., Morris R. G., Rapetti D. A., Applegate D. E., Kelly P. L., von der Linden A., Schmidt R. W., 2014, MNRAS, 440, 2077

Mantz A. B. et al., 2015, MNRAS, 446, 2205

Munari E., Biviano A., Borgani S., Murante G., Fabjan D., 2013, MNRAS, 430, 2638

Munari E., Biviano A., Mamon G. A., 2014, A&A, 566, A68

Muñoz-Cuartas J. C., Müller V., 2012, MNRAS, 423, 1583

Ntampaka M., Trac H., Sutherland D. J., Battaglia N., Póczos B., Schneider J., 2015, ApJ, 803, 50

Ntampaka M., Trac H., Sutherland D. J., Fromenteau S., Póczos B., Schneider J., 2016, ApJ, 831, 135

Old L., Gray M. E., Pearce F. R., 2013, MNRAS, 434, 2606

Old L. et al., 2014, MNRAS, 441, 1513

Old L. et al., 2015, MNRAS, 449, 1897

Old L. et al., 2018, MNRAS, 475, 853

Pearson R. J., Ponman T. J., Norberg P., Robotham A. S. G., Farr W. M., 2015, MNRAS, 449, 3082

Planck Collaboration XIII, 2016, A&A, 594, A13

Roberts I. D., Parker L. C., 2017, MNRAS, 467, 3268

Rozo E. et al., 2010, ApJ, 708, 645

Rykoff E. S. et al., 2014, ApJ, 785, 104

Sadeh I., Feng L. L., Lahav O., 2015, Phys. Rev. Lett., 114, 071103

Saro A., Mohr J. J., Bazin G., Dolag K., 2013, ApJ, 772, 47

Sartoris B. et al., 2016, MNRAS, 459, 1764

Sifón C. et al., 2013, ApJ, 772, 25

Skibba R. A., Sheth R. K., 2009, MNRAS, 392, 1080

Skibba R., Sheth R. K., Connolly A. J., Scranton R., 2006, MNRAS, 369, 68

Skibba R. A., van den Bosch F. C., Yang X., More S., Mo H., Fontanot F., 2011, MNRAS, 410, 417

Skielboe A., Wojtak R., Pedersen K., Rozo E., Rykoff E. S., 2012, ApJ, 758, L16

Soares-Santos M. et al., 2011, ApJ, 727, 45

Sunyaev R. A., Zeldovich Y. B., 1970, Ap&SS, 7, 3

Tempel E. et al., 2014, A&A, 566, A1

The L. S., White S. D. M., 1986, AJ, 92, 1248

Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709

Trevisan M., Mamon G. A., Stalder D. H., 2017, MNRAS, 471, L47

Von Der Linden A., Best P. N., Kauffmann G., White S. D. M., 2007, MNRAS, 379, 867

Wojtak R., Łokas E. L., 2010, MNRAS, 408, 2442

Wojtak R., Mamon G. A., 2013, MNRAS, 428, 2407

Wojtak R., Łokas E. L., Mamon G. A., Gottlöber S., Prada F., Moles M., 2007, A&A, 466, 437

Wojtak R., Łokas E. L., Mamon G. A., Gottlöber S., 2009, MNRAS, 399, 812

Wojtak R., Hansen S. H., Hjorth J., 2011, Nature, 477, 567

Yahil A., Vidal N. V., 1977, ApJ, 214, 347

Yang X., Mo H. J., van den Bosch F. C., Pasquali A., Li C., Barden M., 2007, ApJ, 671, 153

APPENDIX A: PROPERTIES OF THE MASS RECONSTRUCTION METHODS

Table A1. Detailed description of the member galaxy selection process for all methods.

| Methods | Member galaxy selection methodology | | |
	Initial galaxy selection	Membership refinement	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	Friends of friends (FoF)	No	No
NUM	Within 3 Mpc, $4000\,\mathrm{km\,s^{-1}}$	(1) Estimate R_{200c} from the relationship between R_{200c} and richness deduced from CLE; (2) select galaxies within R_{200c} and with $|v| < 2.7\,\sigma_{\mathrm{los}}^{\mathrm{NFW}}(R)$	Same as CLE
RM1	Red sequence	Red sequence	Probabilistic
RM2	Red sequence	Red sequence	Probabilistic
ESC	Within preliminary R_{200c} estimate and $3500\,\mathrm{km\,s^{-1}}$	Gapper technique	Removed by Gapper technique
MPO	Input from CLN	(1) Calculate R_{200c}, R_ρ, R_{red}, R_{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_{200c} and $|2\Phi(R)|^{1/2}$, where R_{200c} 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
MVM	FoF (ellipsoidal search range, centre of most luminous galaxy)	Increasing mass limits, then FoF, loops until closure condition	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_{200c} and σ_v by σ-clipping	Implicit with σ-clipping
CLE	Within 3 Mpc, $4000\,\mathrm{km\,s^{-1}}$	(1) Estimate R_{200c} from the aperture velocity dispersion; (2) select galaxies within R_{200c} and with $|v| < 2.7\,\sigma_{\mathrm{los}}^{\mathrm{NFW}}(R)$; (3) iterate steps (1) and (2) until convergence	Obvious interlopers are removed by velocity gap technique, then further treated by iterative local $2.7\,\sigma(R)$ clipping
CLN	Input from NUM	Same as CLE	Same as CLE
SG1	Within $4000\,\mathrm{km\,s^{-1}}$	(1) Measure σ_{gal}, estimate M_{200c} and R_{200c}; (2) select galaxies within R_{200c}; (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 σ_{gal}, estimate M_{200c} and R_{200c}; (2) select galaxies within R_{200c}; (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
SG3	Within $2.5\,h^{-1}$ Mpc and $4000\,\mathrm{km\,s^{-1}}$. Velocity distribution symmeterized	Measure σ_{gal}, correct for velocity errors, then estimate M_{200c} and R_{200c} and apply the surface pressure term correction (The & White 1986)	Shifting Gapper with minimum bin size of $420\,h^{-1}$ kpc and 15 galaxies
PCS	Input from PCN	Input from PCN	Same as PCN
PFS	Input from PFN	Input from PFN	No

Note. The colour of the acronym for each method colour corresponds to the main galaxy population property used to perform mass estimation richness (magenta), radii (blue), velocity dispersion (red), projected phase space (black), or abundance matching (green). 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.

APPENDIX B: GOODNESS OF FIT
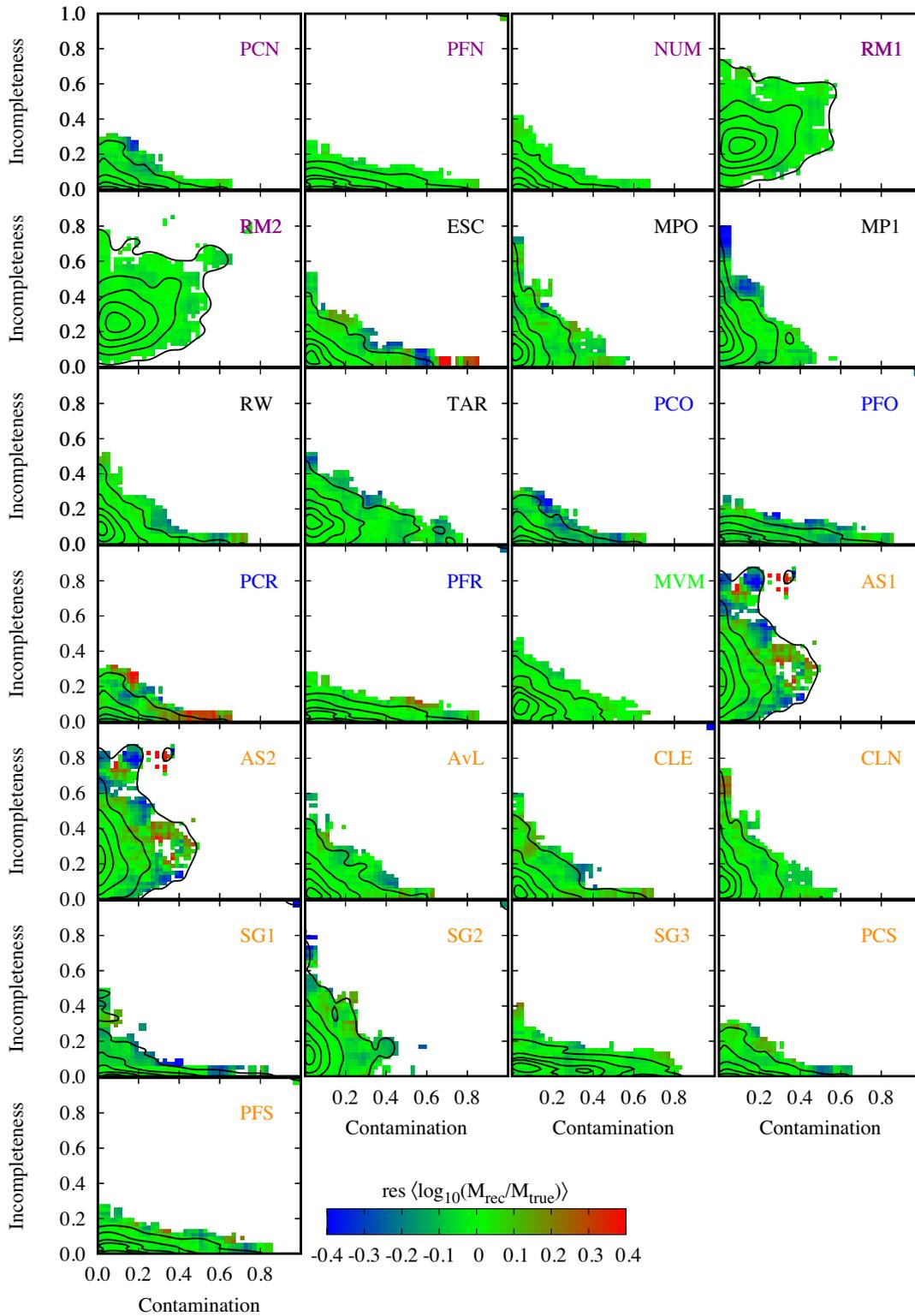
Figure B1. Residual maps demonstrating robustness of the best-fitting models, as given by equation (5), in recovering the dependence of the mass estimate accuracy of different methods on contamination and incompleteness of selected galaxy samples (see Fig. 1).

APPENDIX C: GALAXY SAMPLE SELECTION FUNCTION
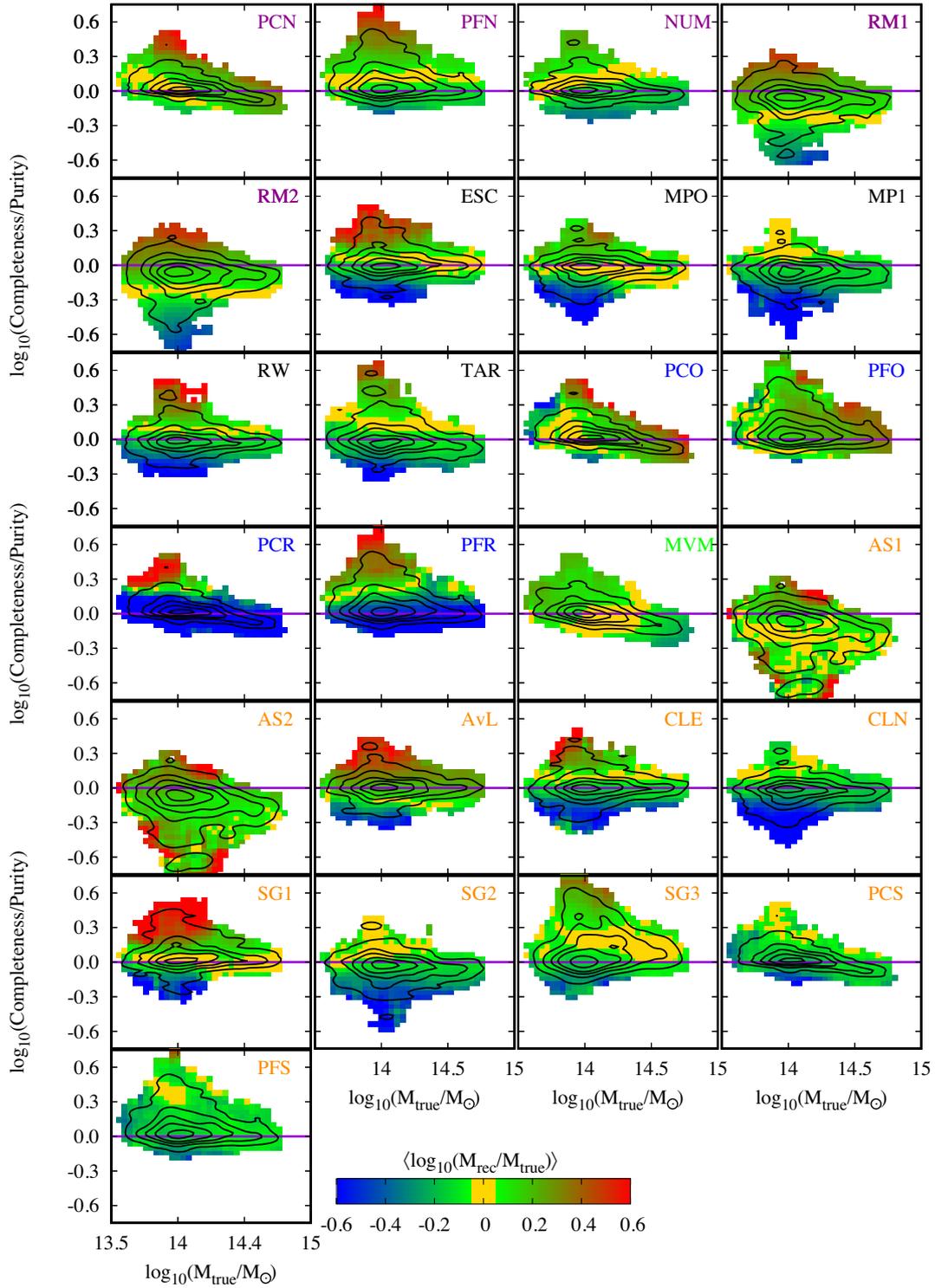
Figure C1. Mass dependence of the galaxy sample selection function for all 25 methods of cluster mass estimation. The selection function is defined as the ratio of completeness to purity, where completeness = 1 − incompleteness and purity = 1 − contamination. The black contours show the distribution of the selection function evaluated at every individual measurement and the true cluster mass. The colour maps show the mean mass bias. Both the maps and contours were computed using the same technique as in Fig. 1.

APPENDIX D: TABLE OF RESULTS

Table D1. Best-fitting parameters of a model describing dependence of cluster mass estimates on true cluster mass, contamination, and incompleteness, as given by equation (7). 'Simplistic' columns show results for a restricted model with no dependence on contamination and incompleteness ($\mu_{C1} = \mu_{I1} = \mu_{C2} = \mu_{I2} = 0$), mean contamination ($\langle C \rangle$), and mean incompleteness ($\langle I \rangle$). Parameter α_0 is the slope of the $M_{\rm rec}-M_{\rm true}$ scaling relation, parameters μ_{ij} are the coefficients of the linear ($j = 1$) and quadratic ($j = 2$) terms in contamination ($i = C$) or incompleteness ($i = I$), while parameters σ_0 and σ_1 are, respectively, the intrinsic scatter and the Poisson-like scatter for $N_{\rm true} = 100$ cluster members (see equation 6). The model accounting for imperfect membership is strongly favoured for every method with $\Delta {\rm BIC} \ll -10$.

Method	$\langle C \rangle$	$\langle I \rangle$	Simplistic μ_0	σ_0	σ_1	α_0	Corrected for imperfectness μ_0	μ_{C1}	μ_{C2}	μ_{I1}	μ_{I2}	σ_0	σ_1	α_0
PCN	0.13	0.06	0.10	0.08	0.09	1.27	0.08	0.63	0.80	-0.65	-1.00	0.03	0.08	1.54
PFN	0.22	0.06	-0.01	0.09	0.08	0.91	-0.02	0.61	0.51	-0.79	-0.79	0.03	0.04	0.99
NUM	0.11	0.06	-0.06	0.09	0.06	0.81	-0.07	0.38	0.10	-1.22	1.29	0.04	0.04	0.92
RM1	0.20	0.33	0.13	0.12	0.06	0.95	0.14	0.58	0.36	-0.75	-0.79	0.02	0.05	0.98
RM2	0.19	0.34	0.12	0.10	0.08	0.97	0.13	0.57	0.50	-0.79	-0.86	0.02	0.05	0.98
ESC	0.13	0.09	-0.03	0.04	0.17	1.00	-0.09	0.66	1.18	-2.17	2.50	0.08	0.09	1.11
MPO	0.08	0.15	-0.03	0.11	0.16	1.12	-0.04	0.37	0.03	-1.50	0.01	0.09	0.12	1.11
MP1	0.08	0.20	-0.16	0.06	0.13	1.01	-0.17	0.41	-0.34	-1.11	0.22	0.05	0.11	1.04
RW	0.11	0.11	-0.11	0.15	0.13	1.04	-0.16	0.48	1.30	-1.92	2.20	0.09	0.07	1.09
TAR	0.14	0.14	-0.11	0.10	0.11	1.01	-0.11	0.58	0.23	-0.87	-1.30	0.08	0.08	1.11
PCO	0.14	0.06	0.12	0.04	0.17	1.37	0.08	0.26	1.41	-1.35	-1.34	0.03	0.16	1.65
PFO	0.22	0.06	0.19	0.02	0.15	1.26	0.17	0.08	0.12	-1.84	0.77	0.02	0.13	1.32
PCR	0.13	0.06	-0.74	0.16	0.37	1.27	-0.59	3.48	-3.18	2.54	-4.77	0.26	0.28	1.16
PFR	0.22	0.06	-0.31	0.28	0.13	0.59	-0.26	1.77	-0.59	0.19	-1.51	0.15	0.08	0.73
MVM	0.13	0.11	0.06	0.08	0.06	0.63	0.07	0.40	-0.65	-0.19	-0.87	0.06	0.06	0.75
AS1	0.11	0.31	0.09	0.23	0.18	0.97	0.09	1.24	-0.18	-0.10	0.32	0.20	0.17	1.11
AS2	0.11	0.31	0.18	0.24	0.18	0.86	0.16	1.17	-0.18	-0.01	0.52	0.20	0.17	0.98
AvL	0.11	0.10	0.16	0.15	0.13	1.01	0.13	0.46	0.18	-2.31	1.93	0.03	0.08	1.05
CLE	0.13	0.11	-0.11	0.06	0.17	0.97	-0.18	0.58	0.70	-2.11	1.82	0.02	0.10	0.98
CLN	0.08	0.15	-0.24	0.08	0.14	1.06	-0.24	0.29	-0.49	-1.55	-0.59	0.02	0.09	1.02
SG1	0.18	0.08	0.07	0.03	0.20	0.92	-0.01	0.97	1.19	-1.66	0.50	0.02	0.12	1.01
SG2	0.10	0.20	-0.15	0.07	0.14	0.94	-0.20	0.58	-0.13	-1.18	1.16	0.02	0.11	1.03
SG3	0.25	0.07	-0.05	0.05	0.12	1.04	-0.08	0.39	0.33	-1.36	1.00	0.02	0.10	1.06
PCS	0.13	0.06	-0.17	0.16	0.12	1.03	-0.21	0.32	0.73	-1.47	1.19	0.16	0.11	1.34
PFS	0.22	0.06	-0.15	0.16	0.12	1.08	-0.19	-0.04	0.96	-1.16	-0.42	0.12	0.12	1.07

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