Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **536,** 931–945 (2025) https://doi.org/10.1093/mnras/stae2593
Advance Access publication 2024 November 26

Copacabana: a probabilistic membership assignment method for galaxy clusters

J. H. Esteves [1,2★] M. E. S. Pereira [3] M. Soares-Santos [4] J. Annis,[5] B. Welch,[6,7] H.-Y. Wu,[8] A. Palmese,[9] A. Farahi [10] F. Andrade-Oliveira,[4] H. Lin,[5] J. García-Bellido,[11] M. Aguena,[12] O. Alves,[2] D. Bacon,[13] S. Bocquet,[14] D. Brooks,[15] A. Carnero Rosell,[12,16,17] J. Carretero,[18] M. Costanzi,[19,20,21] L. N. da Costa,[12] J. De Vicente,[22] P. Doel,[15] S. Everett,[23] B. Flaugher,[5] J. Frieman,[5,24] D. Gruen,[14] R. A. Gruendl,[25,26] G. Gutierrez,[5] S. R. Hinton,[27] D. L. Hollowood,[28] K. Honscheid,[29,30] D. J. James,[31] K. Kuehn,[32,33] C. Lidman,[34,35] M. Lima,[12,36] J. L. Marshall,[37] J. Mena-Fernández,[38] R. Miquel,[18,39] J. Myles,[40] R. L. C. Ogando,[41] A. Pieres,[12,41] A. A. Plazas Malagón,[42,43] A. K. Romer,[44] E. Sanchez,[22] D. Sanchez Cid,[22] B. Santiago,[12,45] M. Schubnell,[2] I. Sevilla-Noarbe,[22] M. Smith,[46] E. Suchyta,[47] M. E. C. Swanson,[26] N. Weaverdyck,[2,48] P. Wiseman,[46] and M. Yamamoto[49] (DES Collaboration)

Affiliations are listed at the end of the paper

Accepted 2024 November 12. Received 2024 October 23; in original form 2024 January 12

ABSTRACT

Cosmological analyses using galaxy clusters in optical/near-infrared photometric surveys require robust characterization of their galaxy content. Precisely determining which galaxies belong to a cluster is crucial. In this paper, we present the **CO**lor **P**robabilistic **A**ssignment of **C**lusters **A**nd **BA**yesia**N** **A**nalysis (Copacabana) algorithm. Copacabana computes membership probabilities for *all* galaxies within an aperture centred on the cluster using photometric redshifts, colours, and projected radial probability density functions. We use simulations to validate Copacabana and we show that it achieves up to 89 per cent membership accuracy with a mild dependence on photometric redshift uncertainties and choice of aperture size. We find that the precision of the photometric redshifts has the largest impact on the determination of the membership probabilities followed by the choice of the cluster aperture size. We also quantify how much these uncertainties in the membership probabilities affect the stellar mass–cluster mass scaling relation, a relation that directly impacts cosmology. Using the sum of the stellar masses weighted by membership probabilities ($\mu_\star$) as the observable, we find that Copacabana can reach an accuracy of 0.06 dex in the measurement of the scaling relation at low redshift for a Legacy Survey of Space and Time type survey. These results indicate the potential of Copacabana and $\mu_\star$ to be used in cosmological analyses of optically selected clusters in the future.

Key words: methods: data analysis – galaxies: clusters: general.

1 INTRODUCTION

Galaxy clusters have long been considered a promising astrophysical probes of dark energy (Albrecht et al. 2006; Allen, Evrard & Mantz 2011; Dodelson et al. 2016) as their abundance as a function of redshift and mass is sensitive to the growth rate of structures in the Universe. Clusters are complementary to geometry based probes such as type-Ia supernovae and can be used to test different dark energy models (Huterer 2023). The challenge in realizing this promise is to obtain a well-understood sample of clusters with unbiased mass measurements of a few per cent precision across a wide range of redshifts ($z \sim 0-1$) and masses ($M \sim 10^{13}-10^{15}\mathrm{M}_\odot$). In galaxy clusters, more than 80 per cent of the mass is in the form of dark matter while 5 per cent–15 per cent is diffuse hot gas, and only 1 per cent–5 per cent is in galaxies (Gonzalez, Zaritsky & Zabludoff 2007; Gonzalez et al. 2013; Laganá et al. 2013; Song et al. 2017; Pratt et al. 2019; Umetsu 2020). Apart from a few extremely massive clusters for which direct total mass measurements via gravitational lensing are possible, we rely on indirect scaling relations to infer cluster masses. For instance, the hot intracluster gas has two main observational signatures: a thermal bremsstrahlung (X-ray) emission and the Sunyaev–Zeldovich (SZ) effect. Although these signals correlate strongly with cluster mass, they are reliably detectable only for clusters at the high-mass end where the gas temperature and density are highest (e.g. Sarazin 1986; Bleem et al. 2020; Klein et al. 2022).

Large optical imaging surveys such as the Sloan Digital Sky Survey (SDSS; York et al. 2000) and the Dark Energy Survey (DES; The Dark Energy Survey Collaboration 2005) have produced samples of tens of thousands of clusters with masses $\lesssim 10^{14}\mathrm{M}_\odot$ (Rykoff et al. 2014; Rykoff, Rozo & DES Collaboration 2016). For these low-mass galaxy clusters, the galaxy content is crucial to unlocking their

★ E-mail: jesteves@fas.harvard.edu

potential for cosmology (Wu et al. 2021). Establishing observable quantities that correlate with cluster masses is a challenge. One such quantity is richness (λ), defined as the probability-weighted sum of red-sequence galaxies identified and selected by cluster finding algorithms such as redMaPPer (Rykoff et al. 2016). Richness is an empirical mass proxy optimized to find clusters and is correlated with the cluster mass (Rozo et al. 2009; Rykoff et al. 2012). This quantity relies on a linear colour–magnitude relation known as the red sequence.

It is essential to account for selection effects when measuring the mass of clusters, especially at the low-richness end. Significant efforts have been made to understand the systematic uncertainties associated with photometric cluster richness estimates. Projection effects, such as correlated and uncorrelated structures along the line of sight, can bias cosmological results if not accounted for (Costanzi et al. 2019; Myles et al. 2021). Additionally, it has been found that sample selection effects, such as the stacking of redMaPPer selected clusters, can introduce a bias of 20–60 per cent on the lensing profiles (Sunayama et al. 2020; Wu et al. 2022), making it one of the leading factors contributing to cluster mass systematic uncertainties.

Although red sequence is well studied, its formation time and quenching mechanisms are still debated (Butcher & Oemler 1984; Andreon et al. 2006; Cooper et al. 2007; De Lucia et al. 2007; Puddu et al. 2021). In particular, reproducing a precise colour relation for galaxy clusters is challenging (DeRose et al. 2019; Korytov et al. 2019), even in hydrodynamical simulations (Nelson et al. 2019; Kukstas et al. 2020). This lack of understanding poses a challenge to optical and NIR wavelength cluster cosmology programs that use red-galaxy counts as the mass proxy as systematic uncertainties associated with the mass-proxy scaling relations dominate the error budget (McClintock et al. 2019). One promising avenue to address this challenge is the development of a mass proxy that includes all of the galaxy content of the clusters. Such a mass proxy has a stronger theoretical foundation and, thus, can be simulated and studied more easily than the red sequence (e.g. Anbajagane et al. 2020).

Recently, alternative mass proxies have been studied, in particular, those derived from the stellar mass (e.g. Andreon 2012; Pereira et al. 2018; Bradshaw et al. 2020) and intracluster light (e.g. Huang et al. 2022; Golden-Marx et al. 2023). In Pereira et al. (2018) and Palmese et al. (2020), we designed a novel mass proxy around the stellar mass content of the galaxies in a cluster. Defined as the weighted sum of the stellar masses, $\mu_\star$ uses a more complete representation of the full population of the cluster galaxies than red-galaxy count methods. In Palmese et al. (2020), we used an X-ray sample of clusters to compare $\mu_\star$–mass observable and the λ method, finding that both present a similar scatter in total mass (see their fig. 6). In Pereira et al. (2020), we performed a detailed weak-lensing mass calibration of $\mu_\star$ and we found that the precision of the mass–$\mu_\star$–z scaling relation was comparable to the one obtained by McClintock et al. (2019) for a mass–richness–z relation. Those results indicate that $\mu_\star$ has the potential to become a competitive mass proxy for cluster cosmology.

In this study, we introduce a new methodology called **CO**lor **P**robabilistic **A**ssignment of **C**lusters **A**nd **BA**yesia**N** **A**nalysis (Copacabana). This method assigns probabilities for all galaxies in the cluster region, independent of the cluster finder selection. Copacabana continues to improve the methodology of previous $\mu_\star$-based papers (Pereira et al. 2018; Palmese et al. 2020).

Copacabana's membership assignment enables value-added information for cluster finders to be produced, even for those that do not rely on galaxy catalogues, such as SZ and X-ray. Moreover, Copacabana can be used to study the evolution and properties of galaxy clusters, such as their mass content and galaxy population. This algorithm is particularly useful for X-ray and SZ-selected samples, as their stellar-mass function and the baryon content of the Universe, can then be analysed (Leauthaud et al. 2012; Gonzalez et al. 2013; Kravtsov, Vikhlinin & Meshcheryakov 2018).

In this work, we apply Copacabana to improve and validate the estimates of $\mu_\star$. We use the BUZZARD DES Year 3 (Y3) simulations (DeRose et al. 2022) to validate our algorithm by quantifying the probability's impact on the scaling relation. In addition, we study the impact of the uncertainties in photometric redshifts of three large photometric surveys: SDSS, DES, and the Legacy Survey of Space and Time (LSST; LSST Science Collaboration 2009).

This paper is organized as follows: Section 2 details the membership assignment methodology based on the photometric redshift, colour, and projected radial probability distributions. Section 3 gives an overview of the simulated data set employed in our study, the validation of our algorithm is shown in Section 4, and conclusions are presented in Section 5. Throughout this manuscript, we use logarithm in base ten (log) and we adopt the cosmological parameter values: $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$, and $h = 0.7$.

2 FORMALISM

In this section, we outline the method used in the Copacabana algorithm[1] for assigning membership probabilities to galaxies for a given cluster field. The main motivation is to produce stellar mass estimations for cluster galaxies that only have photometric information. The method presented here improves and extends the algorithm used in Palmese et al. (2016) and Pereira et al. (2018, 2020). This work has been inspired by previous papers by George et al. (2011), Rykoff et al. (2014), and Castignani & Benoist (2016). Copacabana has two main differences relative to those papers: colour distribution and optimization of the cluster aperture. The red and blue galaxy populations are modelled simultaneously. For the cluster aperture, we estimate R_{200c} which is defined as the radius containing 200 times the critical the density of the universe (at the cluster redshift). R_{200c} is inferred from the galaxy distribution around each cluster as described in Section 2.5.

2.1 $\mu_\star$ estimator

A key motivation for the development of Copacabana is to improve measurements of the mass proxy $\mu_\star$. To accomplish that, we focus on the membership probabilities. $\mu_\star$ is defined in a probabilistic manner:

$$\mu_\star = \sum_i P_{{\rm mem},i}\, M_{\star,i} \quad \text{for} \quad R \leq R_{\rm aper}\ , \tag{1}$$

where $M_\star$ is the galaxy's stellar mass, $P_{\rm mem}$ is the membership probability (cf. equation 4). $R_{\rm aper}$ is the cluster radius aperture which is R_{200c}, computed with the method described in Section 2.5, if not defined otherwise. A galaxy's stellar mass is estimated from photometric data, assuming it is at the cluster redshift via the Bayesian Model Averaging (BMA) algorithm (Palmese et al. 2020).

2.2 Membership probabilities

For a given cluster photometric field, the galaxies present belong to only two classes: gravitationally bound systems and field galaxies,

[1] https://github.com/estevesjh/ccopa

i.e. in the background and foreground galaxies. For the bound systems, we define member galaxies as the population inside the R_{200c} defined by host halo mass. We make this distinction given that there are projected correlated structures, filaments, infalling groups, and galaxies. In simulations we know which are the correlated galaxies, in data we do not.

We adopt a Bayesian inference approach to estimate membership probabilities. Within this framework, the probability of a galaxy being a member of the cluster is in general:

$$\mathrm{P(member|data)} = \frac{\mathrm{P(data|member)P(member)}}{\mathrm{P(data)}}\ , \tag{2}$$

where 'data' represents the galaxy input variables: clustercentric distance, photo-z, and colour. The term P(data|member) is our likelihood distribution, described in detail in Section 2.2.1. The prior P(member) is defined as the ratio of the number of member galaxies and the total number of galaxies, i.e. $\mathrm{n_C}/(\mathrm{n_C} + \mathrm{n_F})$, where C and F indicate cluster and field, respectively.

2.2.1 *Cluster likelihood*

The cluster likelihood depends on the joint distribution $\mathrm{P(R,\ z_p,\ c|member)}$ of the galaxies cluster-radius position, photometric redshift, and colour, respectively. At first order, we can assume that these distributions are independent, thus we have:

$$\mathrm{P(R,\ z_p,\ c|member)} = \mathrm{P(R|member)P(z_p|member)P(c|member)}\ . \tag{3}$$

A potential improvement, not explored in this paper, would be to model the joint colour and radial probabilities since galaxies are redder at the cluster centre.

Combining equations (2) and (3), we can write:

$$\mathrm{P_{mem}(R,\ z_p,\ c)} = \frac{\mathrm{P(R,\ z_p,\ c|member)} \times \mathrm{P(member)}}{\mathrm{Q}}, \tag{4}$$

where Q comes from the law of total probability:

$$\mathrm{Q} = \mathrm{P(member)P(R,\ z_p,\ c|member)} + \mathrm{P(field)P(R,\ z_p,\ c|field)}\ . \tag{5}$$

In principle, the full membership probability with the three variables has more potential constraining power than probabilities using fewer variables. For some cases, though, it might be useful to separate the impact of the colour and photo-z variables. For example, we could study the impact of photo-z outliers in cluster galaxies or for blue brightest cluster galaxies (BCGs). For this reason, we also compute the probability for each model variable:

$$\mathrm{P_{mem}(R)} = \frac{\mathrm{P(R|member)} \times \mathrm{P(member)}}{\mathrm{P(member)P(R|member)} + \mathrm{P(R|field)P(field)}}\ , \tag{6}$$

$$\mathrm{P_{mem}(z_p)} = \frac{\mathrm{P(z_p|member)} \times \mathrm{P(member)}}{\mathrm{P(member)P(z_p|member)} + \mathrm{P(field)P(z_p|field)}}\ , \tag{7}$$

$$\mathrm{P_{mem}(c)} = \frac{\mathrm{P(c|member)} \times \mathrm{P(member)}}{\mathrm{P(member)P(c|member)} + \mathrm{P(field)P(c|field)}}\ . \tag{8}$$

In this formalism, the probabilities allow flexibility for the user to drop the selection in a given variable if needed. As an example, in the case of group galaxies, the assumption of a Navarro–Frenk–White (NFW) profile might not be applicable, so we might drop the radial probability and just restrict the cluster members to the aperture radius. In the next section, we present the definition of each probability.

2.2.2 *Membership probabilities for multiple cluster associations*

It is possible for a single galaxy to have non-zero membership probabilities for multiple clusters, particularly due to photometric redshift uncertainties, which allow for overlapping associations. These cases are not negligible, and we account for them in our analysis.

To handle galaxies associated with multiple clusters, we redefine the membership probabilities. Specifically, the sum of the probabilities across all clusters must always equal one. This is done by introducing the 'probability of the galaxy being taken' (Rykoff et al. 2012; Bellagamba et al. 2019). The updated membership probability for a galaxy in cluster i is given by

$$\tilde{P}_{\mathrm{mem,i}} = P_{\mathrm{taken,i}} \times P_{\mathrm{mem,i}}\ , \tag{9}$$

where i represents the cluster indices, sorted from the highest to lowest initial membership probabilities. The process is iterative, starting with the highest membership probability cluster (index $i = 0$), where the initial probability of being taken is $P_{\mathrm{taken,0}} = 1$. For subsequent clusters, the probability is updated as

$$P_{\mathrm{taken,i}} = 1 - \sum_{j=0}^{i-1} P_{\mathrm{mem,j}} \tag{10}$$

This method offers an advantage over the percolation scheme used in some algorithms (Rykoff et al. 2012), as it ranks clusters by the highest membership probability rather than by the step in which the cluster was detected. This prevents the most probable cluster associations from being undervalued due to the order of detection in percolation schemes, thereby preserving the integrity of the highest-probability associations.

This step is performed at the end of the Copacabana algorithm, once all membership probabilities have been calculated. It ensures that the total membership probability for a galaxy across all associated clusters is always less than or equal to one.

2.2.3 *Radial filter*

We assume that the cluster galaxy radial distribution is a projected NFW profile (Wright & Brainerd 2000), and a constant radial distribution for the background. In our case, the NFW profile density has the form:

$$\Sigma(R) = \begin{cases} \frac{2\rho_s R_s}{r^2-1}\left[1 - \frac{2}{\sqrt{r^2-1}}\arctan\sqrt{\frac{r-1}{r+1}}\right] & r > 1 \\ \frac{2\rho_s R_s}{3} & r = 1 \\ \frac{2\rho_s R_s}{r^2-1}\left[1 - \frac{2}{\sqrt{1-r^2}}\operatorname{arctanh}\sqrt{\frac{1-r}{r+1}}\right] & r > 1\ , \end{cases} \tag{11}$$

where $r = R/R_s$ is the dimensionless radial distance, ρ_s is the density scale parameter, and $R_s = R_{200c}/c_{200}$, where c_{200} is the concentration parameter, is a characteristic radius. To convert the surface mass density profile to a radial probability density function (PDF), we compute the normalization factor

$$\mathrm{Norm} = \int_0^{\mathrm{R_{200}}} 2\pi \mathrm{R'}\Sigma(\mathrm{R'})\mathrm{dR'} \tag{12}$$

such that

$$P(R|\mathrm{member}) = \Sigma(\mathrm{R,\ R_{200},\ c_{200}})/\mathrm{Norm}\ . \tag{13}$$

The NFW has two free parameters, the radius R_{200c} and the concentration c_{200}. We infer R_{200c} using a halo occupation distribution (HOD) model (see Section 2.5) and we set $c_{200} = 3.59$ as this was

shown to be a good fit for haloes in this mass range selected in the DES Science Verification data set (Hennig et al. 2017).

The field radial probability density is assumed to be a constant surface background density $\Sigma_{\rm field}(R) = n_{bkg}$ (Rykoff et al. 2016). As a result, the field radial density probability, P(R|field), is a constant value determined by the normalization $\int_0^{\rm R_{max}} 2\pi {\rm R}' \Sigma_{\rm field} {\rm dR}' = 1$.

2.2.4 Photometric redshift distribution

In the scenario where all members' redshifts are known, the cluster galaxy redshift distribution can be described by a normal distribution with mean $z_{\rm cls}$ and standard deviation $\sigma_{\rm rms}$ due to the galaxies' proper velocities. When using photometric redshifts, the combined distribution of all member galaxies' photometric redshifts generally results in a normal-like distribution for the cluster redshift (Castignani & Benoist 2016; Varga et al. 2019). However the uncertainty in the cluster redshift, $\sigma_{z,0}$, is significantly larger than the intrinsic scatter caused by proper velocities. For example, the typical redMaPPer cluster redshift error is $\sigma_{z,0} = 0.01(1 + z_{\rm cls})$ (Rykoff et al. 2016), whereas the spectroscopic redshift uncertainty due to proper velocities is much smaller, $\sigma_{\rm rms} \approx 0.001$.

The likelihood of a galaxy being a cluster member is computed as the convolution of the cluster's redshift distribution $\Pi_{\rm cls}(z)$ with the individual galaxy's photometric redshift distribution $\Pi_{\rm g}(z)$ (Castignani & Benoist 2016). In other words:

$$P(|z_g - z_c| \leq \sigma_z|{\rm member}) \propto \Pi_{\rm g}(z) * \Pi_{\rm cls}(z)\,. \tag{14}$$

In the regime where the individual galaxy photo-z distribution is normally distributed with $\Pi_{\rm g}(z) = \mathcal{N}(z_p, \delta z^2)$, the likelihood of this galaxy be at the cluster redshift is also a normal distribution with:

$$P\left(z_p|{\rm member}\right) = \frac{1}{\sqrt{2\pi\sigma^2_{\rm eff}}}\, {\rm e}^{-\frac{(z_{\rm cls}-z_p)^2}{2\sigma^2_{\rm eff}}} \tag{15}$$

with $\sigma^2_{\rm eff} = \sigma^2_z + \delta z^2$ assuming that the cluster distribution is also normal. Despite the simplicity of this model, it has been shown to be a robust estimator for membership likelihoods (Castignani & Benoist 2016).

For the field photometric redshift distribution, we use Gaussian Kernel Density Estimation (KDE). The KDE measures the photometric redshift distribution for field galaxies inside a ring[2] around the cluster centre:

$$P(z_p|{\rm field}) = {\rm KDE}(z_p, h). \tag{16}$$

The bandwidth, h, refers to the width of the Gaussian kernel and is a smoothing parameter of the KDE. A common choice is to use the 'Scott's rule of thumb' (Scott 1992):

$$h_{\rm Scott} = 3.49\frac{\sigma(\hat{z_p})}{n^{1/3}} \tag{17}$$

where $\hat{\sigma}(z_p)$ is the standard deviation of the field photometric redshifts input, and n is the total number of objects in the sample. The bandwidth is a more localized estimate of the density function for larger sample sizes given that $h \propto n^{1/3}$. Scott's rule assumptions include normal (or at least symmetric) and unimodal (single peak) distributions, which fairly make up the case for the background selection of redshift.

[2]Defined as an annulus centred on the cluster with inner and outer radii of 4 and 6 Mpc, respectively.

2.2.5 Colour distribution

Generalizing from Pereira et al. (2020), we add colour probabilities by using a colour distribution subtraction method. This method consists of decomposing the total colour distribution in two components, member and field, given by

$$P(c|{\rm total}) = {\rm f_{cls}P(c|member)} + {\rm f_{bkg}P(c|field)}\,, \tag{18}$$

where the $f_{\rm cls}$ and $f_{\rm bkg}$ are the fraction of cluster members and background galaxies in the cluster region within aperture area, $A_{\rm aper} = \pi\, R^2_{\rm aper}$.

We take a data-driven approach to model $P(c|{\rm member})$. We measure for each cluster region the distributions $P(c|{\rm total})$ and $P(c|{\rm field})$ by using a weighted KDE. The field colour distribution, $P(c|{\rm field})$, is measured on a ring around the cluster centre with an inner and outer radius of 4 and 6 Mpc, equally to our background subtraction region definition (see Section 2.4). Thus, the background fraction $f_{\rm bkg}$ is corrected by the ratio of the two areas, $f_{\rm bkg} = \left(N_{\rm ring}/N_{\rm total}\right) \times \left(A_{\rm aper}/A_{\rm ring}\right)$. Where $N_{\rm total}$ is the weighted sum of galaxies in the cluster region ($R \leq R_{\rm aper}$) and $N_{\rm ring}$ is the weighted sum of galaxies in the ring region, the sum is weighted by the photometric redshift probabilities (cf. equation 21).

The cluster colour distribution is computed by subtracting the background colour distribution from the total colour distribution:

$$\begin{aligned} P(c|{\rm member}) = &\frac{1}{1-f_{\rm bkg}}{\rm KDE_{total}}(c|h_{\rm eff}, w) \\ &-\frac{f_{\rm bkg}}{1-f_{\rm bkg}}{\rm KDE_{bkg}}(c|h_{\rm eff}, w). \end{aligned} \tag{19}$$

with the weights w and the effective bandwidth $h_{\rm eff}$. The weights on the KDE are used to improve our colour distribution model by using photometric redshift probabilities. We empirically optimized the bandwidth to have more accurate colour probabilities because the colour distribution deviates from a normal distribution and does not satisfy Scott's rule assumption. We tested different bandwidth criteria, and the optimal result was $h_{\rm eff} = h_{\rm Scott}/10$. This can be explained by the colour distribution being oversmooth by Scott's bandwidth, given the colour distribution can have multiple peaks. Consequently, a smaller bandwidth characterizes the overall colour distribution better.

After the subtraction, we normalize $p(c|z, R)$ to unity. We choose one colour filter at a time to be our colour model:

$$c(z) = \begin{cases} (g-i) & \text{for } z \leq 0.35 \\ (r-z) & \text{for } z > 0.35, \end{cases} \tag{20}$$

this choice is imposed by the 4000 Å break and it is optimal for the DES filters. This colour evolution is valid up to $z \leq 0.75$ and for larger redshift ranges the effect is observed in redder filters, e.g. $(z-i)$ (Black & Evrard 2024, see their fig. 1). Compared with Pereira et al. (2020), adding colour probabilities, in general, improved the performance.

2.3 Cluster photometric redshift probabilities

We infer the numerical density of cluster galaxies in a probabilistic manner, given the large uncertainties on the galaxy photometric redshifts. We define the probability P_{z_0} of a galaxy being at the cluster redshift, $z_{\rm cls}$, as the integral of an individual galaxy's photometric redshift distribution ($\Pi(z)$):

$$P_{z_0} = \int_{z_-}^{z_+} \Pi(z)\, dz, \qquad z_\pm = z_{\rm cls} \pm 2 \times \sigma_{z,0}(1 + z_{\rm cls})\,, \tag{21}$$

where the limits of integration span over the average survey photometric redshift precision $\sigma_{z,0}$, for instance, the DES Y1 galaxies photo-z point estimates have a precision around $\sigma_{z,0} \approx 0.03$ (Gschwend et al. 2018; Aguena et al. 2021). An additional model choice would be to have an integration window dependent on magnitude, given fainter galaxies have larger uncertainties than brighter galaxies (Castignani & Benoist 2016). Because our study limits to simulated photometric redshift (cf. Section 3.1.2), we considered a fixed integration window.

A strong motivation for using individual galaxies PDFs, $\Pi(z)$, is because they carry more information than point estimates, especially for double-peak distributions where the mean and the standard deviation does not fully represent the measurement statistical information.

Computing the probability of galaxy be around the cluster redshift is the first step in our algorithm, and is used in the following steps: the background subtraction, the estimation of the radius R_{200c}, and the colour model.

2.4 Background subtraction

Background subtraction is an essential step for computing membership probabilities. There are two methods traditionally used: global and local subtraction. The global background subtraction method involves averaging over a large density field, assuming that the background density only depends on redshift (e.g. Rykoff et al. 2014). However, this approach underestimates the local density field because galaxy clusters reside in the most densely populated regions, which are more than 20 Mpc away from the mean density values of the large-scale structure. A more reliable method involves subtracting the local background as it probes the surrounding areas of each cluster. Our study calculates the background galaxy density locally in a ring around the cluster with inner and outer radii of 4 and 6 Mpc, respectively. The inner radii are always larger than R_{200c}, even for the most massive clusters for which R_{200c} is approximately 3 Mpc. Although scaling the radii with the cluster radius would be an optimal choice, we prefer to use fixed values since our R_{200c} estimation depends on the background density.

2.5 R_{200c} estimator: HOD Model

Galaxy clusters are objects that do not have well-defined edges in photometric images. Common cluster size definitions are based on the virial theorem or overdensities of the critical density. Cluster-finders have different definitions of cluster size when defining apertures since they do not know the mass beforehand. For instance, redMaPPer (Rykoff et al. 2014) assumes an aperture that scales with richness λ, and AMICO (Bellagamba et al. 2018) assumes a fixed aperture corresponding to a cluster with $M_{200,c} = 10^{13.5}$ M$_\odot$. In this work, we introduce a new cluster aperture estimator based on an HOD model, which is independent of our mass proxy.

Our aperture estimator uses the galaxy number density profile of an HOD model. A given HOD model provides the number of halo galaxies as a function of mass which allows us to convert the number density profile to a mass density profile. Assuming spherical symmetry, we can calculate the mass density field, $\rho_h(R)$. We can make a rough estimation of R_{200c} by interpolating the mass density profile as a function of radii. By definition $\rho_h(R_{200c}) \equiv 200\rho_{\rm crit}$, where the mass density profile is 200 times the critical density, we have our aperture estimation, $R^{\rm HOD}_{200c}$. For this work, we adopt the HOD model of Tinker et al. (2012). The model consists of a relation between the number of central ($N_{\rm cen}$) and satellite galaxies ($N_{\rm sat}$) inside a halo of given mass ($M_{200,c}$) and below a given luminosity threshold.

Our estimator of R_{200c} is computed by finding the radius value where the halo mass density profiles ρ_h is two hundred times the critical density $\rho_{\rm crit}$, i.e.

$$\rho_h(R_{200c}) - 200\rho_{\rm crit} = 0 \tag{22}$$

To infer the halo mass density profile, we assume a relation of the number of galaxies inside a halo and its mass:

$$\rho_h(< R) = M_h(N_h(< R))/\left(\frac{4\pi}{3}R^3\right) \tag{23}$$

We take advantage that an HOD model describes the average number of central ($\langle N_{\rm cen}\rangle_M$) and satellites ($\langle N_{\rm sat}\rangle_M$) associated within a halo for a given halo mass M. The total number of galaxies ($\langle N_{\rm cen}\rangle_M + \langle N_{\rm sat}\rangle_M$) associated within the halo can be written as (Tinker et al. 2012)

$$\langle N_h\rangle_M = \langle N_{\rm cen}\rangle_M \times \left[1 + \left(\frac{M}{M_{\rm sat}}\right)^{\alpha_{\rm sat}} \exp\left(-\frac{M_{\rm cut}}{M}\right)\right], \tag{24}$$

where $\alpha_{\rm sat}$ is the slope at high halo masses, with an exponential cut-off at halo masses below $M_{\rm cut}$, and $M_{\rm sat}$ is the characteristic halo mass for satellites. With $\langle N_{\rm cen}\rangle_M$ given by

$$\langle N_{\rm cen}\rangle_M = \frac{1}{2}\left[1 + {\rm erf}\left(\frac{\log M - \log M_{\rm min}}{\sigma_{\log M}}\right)\right], \tag{25}$$

with the $M_{\rm min}$ representing the halo mass at which the probability of containing a central galaxy is 50 per cent, and $\sigma_{\log M}$ accounts for the scatter in halo mass at a fixed luminosity of the galaxy population.

By taking the inverse of the relation equation (24), we can build a numerical relation of the total number of galaxies with the halo mass density profile $\rho_h(R)$. For a photometric data set, we have a noise estimator $\langle\tilde{N}_h\rangle$ of the actual $\langle N_h\rangle$ since we do not have membership information. Thus, by summing all field and halo galaxies and subtracting the number of field galaxies, the $\langle\tilde{N}_h\rangle$ is

$$\langle\tilde{N}_h\rangle(< R) = N_{h+f}(< R) - N_f = \sum_i P_{z_0,i} - \Sigma_{bkg}\pi R^2 \tag{26}$$

where Σ_{bkg} is a surface background density computed in the ring of 3 and 4 Mpc inner and outer size, respectively. This measurement is done by selecting all galaxies within $|z - z_{cls}| \leq 2\sigma_{z_0}(1 + z_{cls})$.

In addition, on the HOD model there are five free parameters, cf. equations (25) and (24). The best-fitting values for these parameters (given in table 4 of Tinker et al. 2012) were derived from the SDSS data set, using the maxBCG cluster sample (Koester et al. 2007), with an absolute magnitude cut $M_r^{0.1} \leq -19.5$. We use these values as a reference since they are close to the $0.2L_\star$ magnitude cut applied here (Rykoff et al. 2012).

2.6 Stellar mass estimation: BMA

The code we call BMA[3] is a Bayesian model averaging code (see e.g. Taylor et al. 2011) applied to the output of a stellar population synthesis code, and developed into a pipeline (Palmese et al. 2020). We use the stellar population synthesis code FSPS (Conroy, Gunn & White 2009; Conroy & Gunn 2010) to evaluate a five-dimensional space of quantities, resulting in 24 models. We choose the models evaluated at the cluster redshift for a given galaxy with apparent magnitude, colours, and photo-z. Then the likelihood of each model

[3]https://github.com/apalmese/BMAStellarMasses

given the galaxy magnitudes, colours, and errors is computed. The properties of interest from the models, e.g. stellar mass, are then computed as the likelihood-weighted sum over all models, a Bayesian model average. The code was validated on the Millennium simulations (Springel et al. 2005) and on the COSMOS data set (Laigle et al. 2016). See Palmese et al. (2020) for a full description of the BMA methodology.

3 VALIDATION SET-UP

3.1 Data

To validate Copacabana, we use the BUZZARD v2.0 simulations (DeRose et al. 2019), meant to correspond with the DES Y3 area. The data set consists of synthetic dark matter simulations with galaxy information added by the AddGals (Adding Density Dependent GAlaxies to Lightcone Simulations) algorithm (Wechsler et al. 2022). This procedure places galaxies on to the dark matter-only simulation, weighted by local dark matter density, matching the observed luminosity function and luminosity-dependent two-point correlation function.

For context, we briefly describe how the galaxies are pasted on to the dark matter particles. First, AddGals creates a catalogue of galaxies based on the luminosity function, $\phi(M_r)$, performing subhalo abundance matching between a small high-resolution N-body simulation and the observed SDSS luminosity function in the r band. The algorithm calibrates a model of the luminosity–density–redshift, $P(R_\delta|M_r, z)$ relation for the central and the non-central galaxies on the high-resolution N-body simulations, where R_δ is the radius enclosing $1.3 \times 10^{13}\, h^{-1}\,\mathrm{M}_\odot$. These relations once calibrated on high-resolutions simulations are then used to assign galaxies to resolved haloes or dark matter particles in a large light-cone simulation with a lower resolution. For instance, for BUZZARD, AddGals was trained on one high-resolution simulation of box size $400\,h^{-1}$ Mpc. Then, the luminosity–density–redshift relations were applied to three larger lower resolution simulations of box sizes 1.05, 2.6, and $4.0\,h^{-1}$ Gpc, respectively (DeRose et al. 2019).

The colours are assigned using a spectral energy distribution (SED), chosen such that the simulation matches the SED distribution (at fixed luminosity and galaxy density) measured in the SDSS data. The AddGals modelling scheme was chosen such that it predicts the clustering in SDSS to high precision (e.g. Conroy, Wechsler & Kravtsov 2006; Reddick et al. 2013; Lehmann et al. 2017).

As describe above, the galaxies in BUZZARD, unlike many cosmological simulations, are not placed using an HOD prescription. None the less, the HOD for haloes with $M \geq 5 \times 10^{13}\,\mathrm{M}_\odot$ on Buzzard follows the Tinker et al. (2012) model closely for an absolute magnitude selection of $M_r^{0.1} \leq -19.5$, as we can see in Fig. 1. Below this halo mass, the occupation is higher in the simulations compared to the HOD model. In addition, Zacharegkas et al. (2022), found a quantitative agreement between an HOD model and the Buzzard redMaGic galaxies.

3.1.1 Sample selection

For an unbiased assessment of the performance of our code, we have selected haloes uniformly across the bins of logarithmic mass and redshift. Our objective is to maintain an equal halo count within each bin, except for the highest mass bins where a lack of clusters is observed. This approach prevents our assessment from being biased by the low end of the halo distribution.

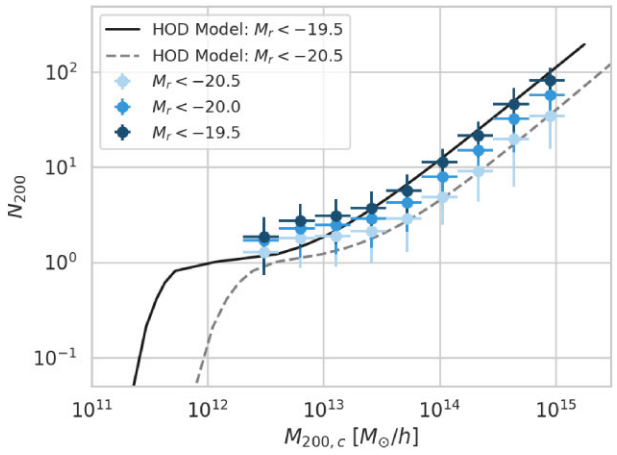

Figure 1. Number of galaxies inside $R_{200,c}$ (N_{200}) as a function of the halo mass $M_{200,c}$, for three absolute magnitude cuts in the r band. The Buzzard galaxy distribution follows the Tinker et al. (2012) model closely (solid line) when both apply the same magnitude selection of $M_r^{0.1} \leq -19.5$ mag (points with error bars).

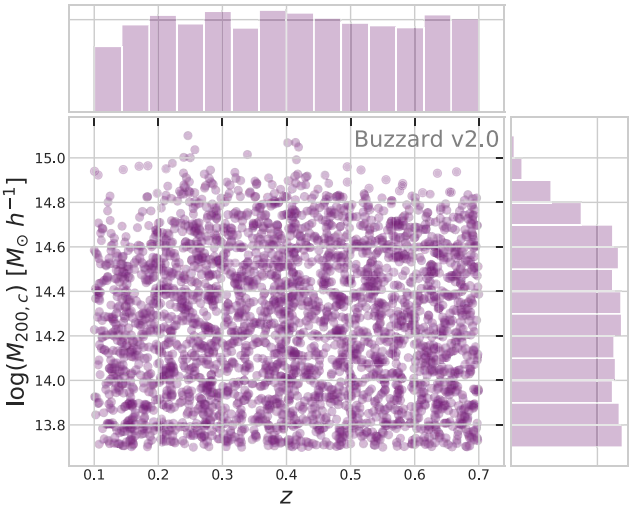

Figure 2. Uniform selection of 2200 Buzzard v2.0 haloes on a halo mass–redshift grid. The upper and right panels are the redshift and halo mass distributions, respectively.

The data chosen by our selection is presented in Fig. 2, where 2200 haloes are plotted in $(z, \log M_{200,c})$ space with histograms on the x- and y-axis. The limits of the sample are $z \in [0.1, 0.65]$ and $(\log M_{200,c}) > 13.5\ \mathrm{M}_\odot/h$. The choice of the redshift range follows the DES cluster cosmology analysis (Abbott et al. 2020). The halo mass threshold is similar to that adopted by other cluster finder algorithms (e.g. Castignani & Benoist 2016; Bellagamba et al. 2019). Overall, there is a uniform selection, except for the highest mass bins, where there are not enough systems.

3.1.2 Simulated photo-z

To validate the Copacabana algorithm with respect to photo-z, we add offsets to the simulated galaxy redshifts. In detail, we draw a random offset following:

$$z_p = \mathcal{N}\left(z, \sigma_{z,0}(1+z)\right)\ , \tag{27}$$

where $\sigma_{z,0}$ is the photo-z precision that correspondents to a typical photo-z error.

We simulate three different levels of uncertainty: $\sigma_{z,0} = 0.01, 0.03,$ and 0.05 as an ideal, a realistic, and a pessimistic case, respectively. These choices mimic three different surveys: LSST (LSST Science Collaboration 2009), DES (Gschwend et al. 2018; Aguena et al. 2021), and SDSS (Carliles et al. 2010), respectively.

In the context of clusters, the main differences between simulated Gaussian photo-z's with real data photo-z's are the bias and the presence of outliers. For instance, Aguena et al. (2021) using the WaZP cluster catalogue studied the differences of redMaPPer cluster redshift with the ones derived from the DNF photo-z algorithm in the DES Y1 (Gschwend et al. 2018). They quantified a redshift bias that is less than $0.003 \times (1 + z)$. For future applications of Copacabana on data, a description of the bias between the photo-z sample employed and the cluster redshift must be taken into account as a bias on the cluster photo-z distribution.

3.2 Validation metrics

In this section, we present a set of definitions and metrics to assess the performance of the Copacabana algorithm.

3.2.1 *Assessing $\mu_\star$ precision*

To validate our $\mu_\star$ estimation, we compare it with the simulation cluster member stellar masses. For this purpose, we define $\mu_{\star,\rm true}$,

$$\mu_{\star,\rm true} = \sum_{i \in \rm members} M_{\star,i} \quad \text{with} \quad R \leq R_{\rm aper} \,, \tag{28}$$

as the sum of the cluster members' stellar masses, where the 'true' members are defined as the galaxies inside the three-dimensional R_{200c} distance from the cluster centre. In other words, the line-of-sight infall galaxies are not considered, nor the gravitational status of the galaxy.

At very low $\mu_{\star,\rm true}$, below $10^{12}\,{\rm M}_\odot$, our data has a large tail. For this reason, a robust metric is adopted the scaled median absolute deviation (MAD):

$$\sigma_{\rm MAD}(\log(x)) = 1.48 \times \text{Median}\left(|\log(x) - \text{Median}\left(\log(x)\right)|\right) \,, \tag{29}$$

where $x \equiv \mu_\star/\mu_{\star,\rm true}$. Note that if $\log(x)$ follows a normal distribution $\sigma_{\rm MAD}$ is equal to the standard deviation.

It is important to stress that our assessment is primarily on our estimator due to membership probabilities and we do not evaluate uncertainty due to stellar mass estimates.

3.2.2 *Assessing the accuracy of R_{200c}*

To validate our estimates of R_{200c}, we use the current value from the simulation. $R_{200c,\rm true}$ was retrieved from the Buzzard truth table. We evaluate equation (29) with the ratio of true versus measured R_{200c}, i.e. $x \equiv R_{200c}/R_{200c,\rm true}$.

3.2.3 *Completeness and purity*

The membership probabilities play a role in thresholds that distinguish the classes in the framework of classifying members and non-members of a given galaxy cluster. To assess the performance, we use metrics commonly used in statistical classification problems, purity (P) and completeness (C).These metrics rely on true positive (TP), false positive (FP), and false negative (FN) predictions:

$$P = \frac{TP}{TP + FP} \quad \text{and} \quad C = \frac{TP}{TP + FN};, \tag{30}$$

TP represents correct positive predictions, while FP and FN refer to incorrect positive and negative predictions, respectively. Purity indicates the proportion of positives that are cluster members. Completeness measures the fraction of true members that were successfully identified among all the selected galaxies.

The overall accuracy of a classifier can be evaluated by

$$\text{accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \,, \tag{31}$$

where TN is the true negatives, i.e. correctly identified field galaxies.

3.2.4 *$\mu_\star$–cluster mass scaling relation*

For photometric surveys, one of the main requirements for tight constraints on cosmological parameters is a mass proxy that predicts the cluster mass with significant accuracy and is robust against systematic effects. Here, we assess the possible impact of the membership probabilities on deriving cosmological results using the relation between the weighted stellar mass and the total cluster mass, which we will refer to as the $\mu_\star$–cluster mass scaling relation.

In simulations, the $\mu_\star$–cluster mass scaling relation is accessible since the halo mass is known. The probability of a given halo of mass $M_{200,c}$ to have a $\mu_\star$ value is generally modelled by a lognormal relation with mean:

$$\left\langle \log(\mu_\star) | M_{200,c} \right\rangle = \alpha + \beta \log(M_{200,c}/M_p) \,, \tag{32}$$

where σ is the $\mu_\star$-cluster mass intrinsic scatter. Here, α is the intercept, β is the slope, and $M_p = 10^{15.5}{\rm M}_\odot$ is the pivot mass. The inference of the model parameters is made by employing a hierarchical Bayesian algorithm (`linmix`; Kelly 2007). The `linmix` algorithm allows us to include the error on the y-dependent variable, in our case, $\mu_\star$.

In general, the scaling relation evolves with redshift. Accordingly, the fit is conducted in small redshift bins where the redshift evolution can be neglected.

4 RESULTS

In this section, we examine the performance of Copacabana using the BUZZARD simulation.

4.1 Uncertainty in $\mu_\star$ estimations

We run Copacabana on the BUZZARD v2.0 simulation using the three values of $\sigma_{z,0}$ presented in Section 3.1.2. We employ the photometric stellar masses computed by BMA at the cluster redshift (Section 2.6). The stellar mass $\mu_\star$ is computed with $R_{\rm aper} = R_{200c}$ estimated using the HOD model presented in Section 2.5. The membership probabilities are expected to depend mainly on the photo-z uncertainty.

Fig. 3 shows the estimated $\mu_\star$ versus $\mu_{\star,\rm true}$ within R_{200c} for three values of $\sigma_{z,0}$. The $\mu_\star$ values for all halo mass regimes follow the $\mu_{\star,\rm true}$ values closely. The $\mu_\star$ errors are roughly within 0.20 dex $\sqrt{\mu_{\star,p}/\mu_\star}$, where $\mu_{\star,\rm p} = 10^{12.22}\,{\rm M}_\odot$ is the sample mean. Uncertainty in the position of the galaxies along the line of sight adds significant noise, especially for the low-mass haloes as they have fewer galaxies. As a result, the quality of the photometric redshifts

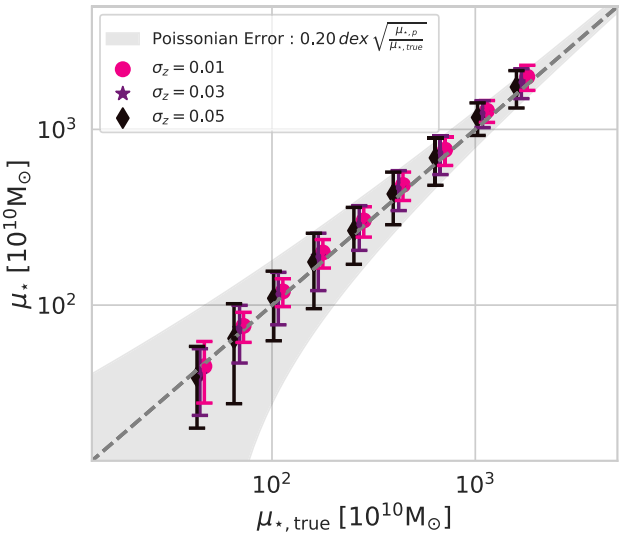

Figure 3. Copacabana predicted $\mu_\star$ as a function of $\mu_{\star,\mathrm{true}}$ (the sum of the stellar masses of the true cluster members within R_{200c}) for different values of $\sigma_{z,0}$, the uncertainty in the galaxy photometric redshifts. The binned points were slightly shifted for better visualization. The estimated values closely follow the one-to-one relation (dotted line). For comparison, we also show the range corresponding to a Poissonian error of 0.20 dex $\sqrt{\mu_{\star,p}/\mu_\star}$ where $\mu_{\star,\mathrm{p}}$ is the sample mean of $10^{12.22}$ M$_\odot$. Note that our result is robust well below the threshold of 10^{12} M$_\odot$ used in our previous work (Pereira et al. 2020).

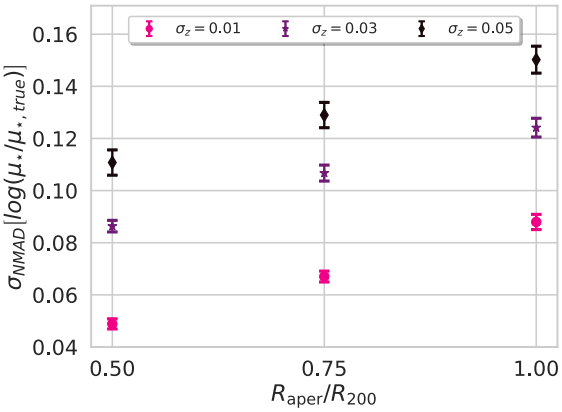

Figure 4. Copacabana σ_{MAD} error of the log $\mu_\star/\mu_{\star,\mathrm{true}}$ as a function of the cluster aperture radius for three different photo-z precisions. The σ_{MAD} error is always below 0.16 dex, and it varies significantly with the photo-z quality (different markers). In particular, for an LSST-like photo-z (rounded markers), σ_{MAD} is 0.09 dex. The precision can be improved considerably by defining a smaller cluster aperture since the density is higher in the cluster centre. The error bars were estimated by bootstrapping the cluster sample a thousand times.

substantially impacts the $\mu_\star$ measurements and particularly does at the low-mass end. Nevertheless, a sample with accurate and precise photometric redshifts can accurately predict $\mu_\star$ down to the lowest bin, 5×10^{11}M$_\odot$.

4.1.1 *Photometric redshift uncertainties and cluster apertures*

In this section, we quantify how the uncertainty in $\mu_\star$ depends on $\sigma_{z,0}$ and the size of the apertures used to estimate $\mu_\star$. Here, we vary the cluster aperture given that the membership selection has a significant dependence with radius (Lopes & Ribeiro 2020).

In Fig. 4, we show the σ_{MAD} (defined in equation 29) as a function of the cluster aperture and photo-z uncertainty. The accuracy of $\mu_\star$ correlates linearly with the uncertainty in the photometric redshifts. For instance, an improvement of a factor 5 (1.7) in $\sigma_{z,0}$ reduces the $\mu_\star$ error to 0.15 dex (0.09 dex) when compared to $\sigma_{z,0} = 0.05$. This improvement in the photometric redshifts has a big impact, especially for low-mass haloes. This result follows from the fact that the uncertainty along the line of sight is the major source of galaxy membership contamination. For example, a redshift error of $0.01 \times (1 + z)$ translates into a physical length of ~ 40 Mpc which is $\sim 10 - 50 \times R_{200,\mathrm{c}}$.

A second source of error is the cluster aperture, which has a non-negligible effect on the uncertainty in $\mu_\star$. As shown in Fig. 4, smaller cluster apertures decrease the uncertainties in $\mu_\star$. This effect is driven by the higher galaxy density in the core. Lopes & Ribeiro (2020) showed that the local density is a very good indicator of membership galaxies in their well-characterized cluster sample.

The choice of the cluster aperture is important for studying cluster cosmology. Ideally, we would expect that $\mu_\star$ has the highest correlation with the halo mass when computed at R_{200c} (Rykoff et al. 2012; Bradshaw et al. 2020; Huang et al. 2020). In Section 4.4, we discuss further the choice of the cluster aperture in terms of optimizing the scatter of the $\mu_\star$–$M_{200,\mathrm{c}}$ relation.

4.1.2 *Stellar mass uncertainty*

Thus far, the $\mu_\star$ uncertainty was computed without taking into account the errors on stellar mass. The BMA stellar-mass error is around 0.2 dex (Palmese et al. 2020), validated using the stellar masses computed with multiband data in 16 filters from UV to infrared of the COSMOS deep field (Laigle et al. 2016). The BMA stellar-mass errors are comparable to the uncertainties induced by the photo-z errors, see Fig. 4. Therefore, they have a significant impact. To quantify this impact, we add random noise normally distributed to the estimated stellar masses. We assume the typical BMA error of 0.20 dex. This assumption should set an upper bound on the uncertainty in $\mu_\star$. Because $\mu_\star$ is dominated by high stellar mass (bright) galaxies that have lower mass uncertainties.

The additional scatter on $\mu_\star$ is 0.07 ± 0.01 dex for the three photo-z samples. This result is equivalent to adding the stellar mass error in quadrature, $\sigma^2_{\mathrm{MAD}} + \sigma^2_{\mathrm{BMA}}$. The implication of this additional error for the DES-like photo-z case is that the $\mu_\star$ error is at the same level as an SDSS-like photo-z. In future work, it would be important to reduce the uncertainty on stellar masses to reduce the uncertainty on $\mu_\star$.

4.2 Precision of R_{200c} estimations

The new R_{200c} estimator is based on an HOD model, which is sensitive to the relation between the number of galaxies and the halo mass. We found that this approach results in R_{200c} estimates that are biased low. To account for this bias, we introduce a calibration factor which can be computed as the mean ratio of our predictions and the actual R_{200c} values in the simulations:

$$\eta_{\mathrm{HOD}} = \frac{\mathrm{R_{200c}}}{\mathrm{R_{200c,true}}} \tag{33}$$

We derived a calibration factor η_{HOD} (equation 33) of 0.63 ± 0.11 for the same HOD luminosity cut. We use a larger data set, all

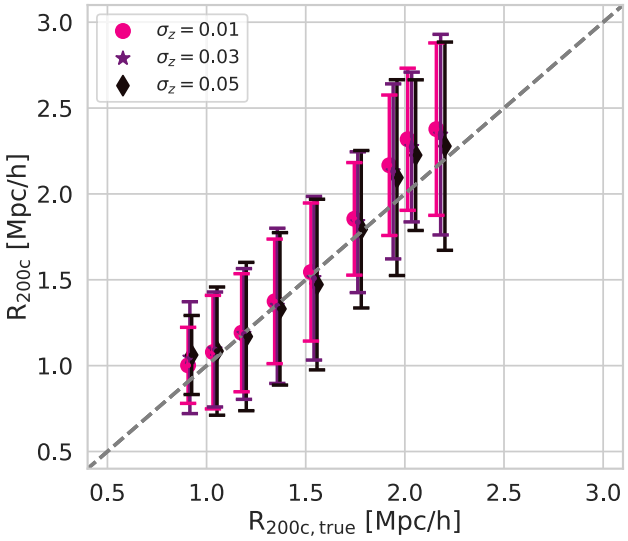

Figure 5. The Copacabana-predicted R_{200c} based on an HOD model versus the true radius ($R_{200c,true}$) for three values $\sigma_{z,0}$. The values have been corrected using a calibration factor, $\eta_{HOD} = 0.63 \pm 0.11$. The scatter, around 30 per cent, is mainly due to intrinsic errors in the HOD method, with minimal impact from photometric redshift uncertainty.

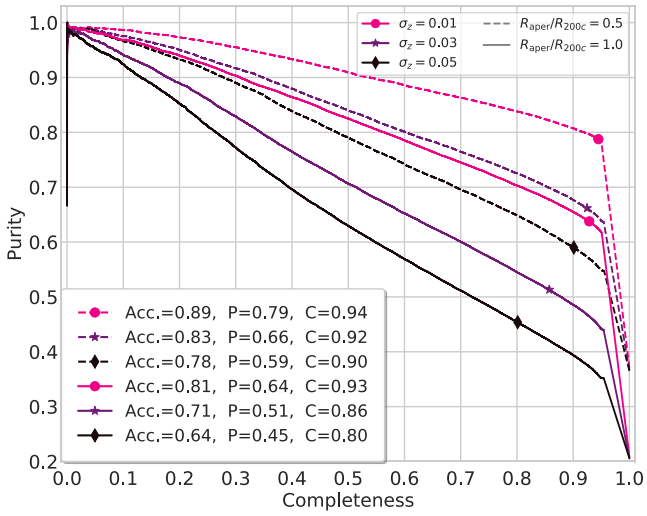

Figure 6. Purity (P) versus completeness (C) for different photo-z samples. The purity of the membership selection is dependent on the accuracy of the photometric redshifts. For instance, with an optimistic accuracy of $\sigma_{z,0} = 0.01$, the purity and completeness reach an optimal value of 79 per cent and 94 per cent, respectively. The DNF photo-z algorithm has an accuracy that is similar to our realistic $\sigma_{z,0} = 0.03$ scenario. However, it has lower completeness due to outliers, which are not present in the artificial Gaussian photo-z sample.

Buzzard haloes with $M_{200,c} > 5 \times 10^{13}\,M_\odot$ for this computation. This calibration factor is independent of redshift, given we did not find significant variations with redshift.

This intrinsic bias can be related to the HOD of the simulation (see Fig. 1) or the galaxy density field at small scales. Our analysis shows that this data set closely follows our HOD model, which suggests that the observed bias is more likely connected to the small-scale galaxy density field in the simulation. This is likely because the AddGals method does not include an explicit prescription for placing galaxies

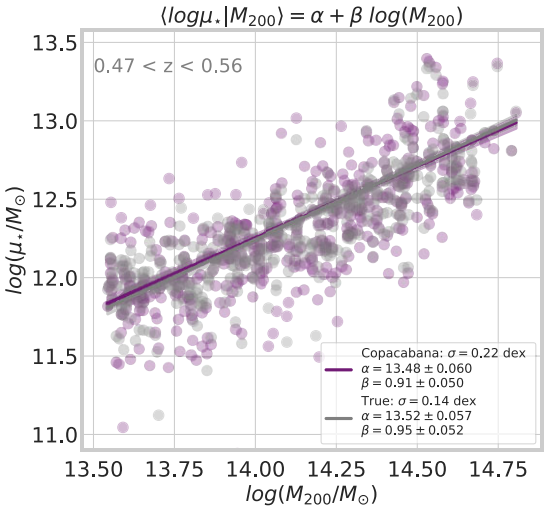

Figure 7. Scaling relation, $\mu_\star$ versus $M_{200,c}$. The DES-like photo-z, $\sigma_{z,0} = 0.03$, has a slope of β=0.91±0.05 close to the actual galaxie's stellar mass relation, $\mu_{\star,true}$, slope. The fitted linear relation using a Bayesian regression linear method (linmix) is shown in solid lines, and the 68 per cent confidence level is shown in shaded bands. The density of points in this relation includes around 370 clusters uniformly distributed in mass, which differs from what might be expected in a typical survey selection.

inside large haloes and has not been extensively tested for galaxy cluster science.

Our estimator predicts R_{200c} for the halo masses probed in this study, as shown in Fig. 5, with a scatter of approximately 30 per cent. Despite the relatively large scatter, the predicted values show a strong correlation with $R_{200c,true}$ for both low- and high-mass clusters, making our estimator a reliable probe of cluster size. The scatter seen in Fig. 5 is primarily driven by the intrinsic error of the HOD method. Because the photometric redshift uncertainty has minimal impact on the predicted values. Similarly, the precision of the estimated radius is independent of cluster mass, in contrast to what we observed with $\mu_\star$.

The significant uncertainties in our estimates of R_{200c} relate to the uncertainty of the HOD model itself. For instance, the HOD model halo mass scatter is 0.204 dex. Also, optical mass proxies, in general, have similar or higher intrinsic errors, e.g. the BCG stellar-mass proxy intrinsic scatter is 0.20 (Behroozi, Conroy & Wechsler 2010; Tinker et al. 2012). In the future, if there are precise stellar mass measurements, the stellar-to-halo mass can be incorporated into our methodology (Behroozi et al. 2010).

4.3 Completeness and purity of the members list

In Fig. 6, we plot the purity versus completeness for different values of $\sigma_{z,0}$, the photometric redshift uncertainty. The figure is constructed by varying the cluster membership probability threshold. The optimal choice (coloured points) is the one that maximizes the product of both quantities.

Overall, Copacabana performs well for different photo-z samples when compared with other classifiers (George et al. 2011; Castignani & Benoist 2016; Lopes & Ribeiro 2020). The product of purity (P) and completeness (C) is maximized a $P = 64$ per cent and $C = 93$ per cent in an optimistic scenario ($\sigma_{z,0} = 0.01$), and values of $P = 45$ per cent and $C = 80$ per cent in the worst scenario ($\sigma_{z,0} = 0.05$). It is important to note that the completeness is not higher than 96 per cent. The 2σ photo-z threshold we use translates

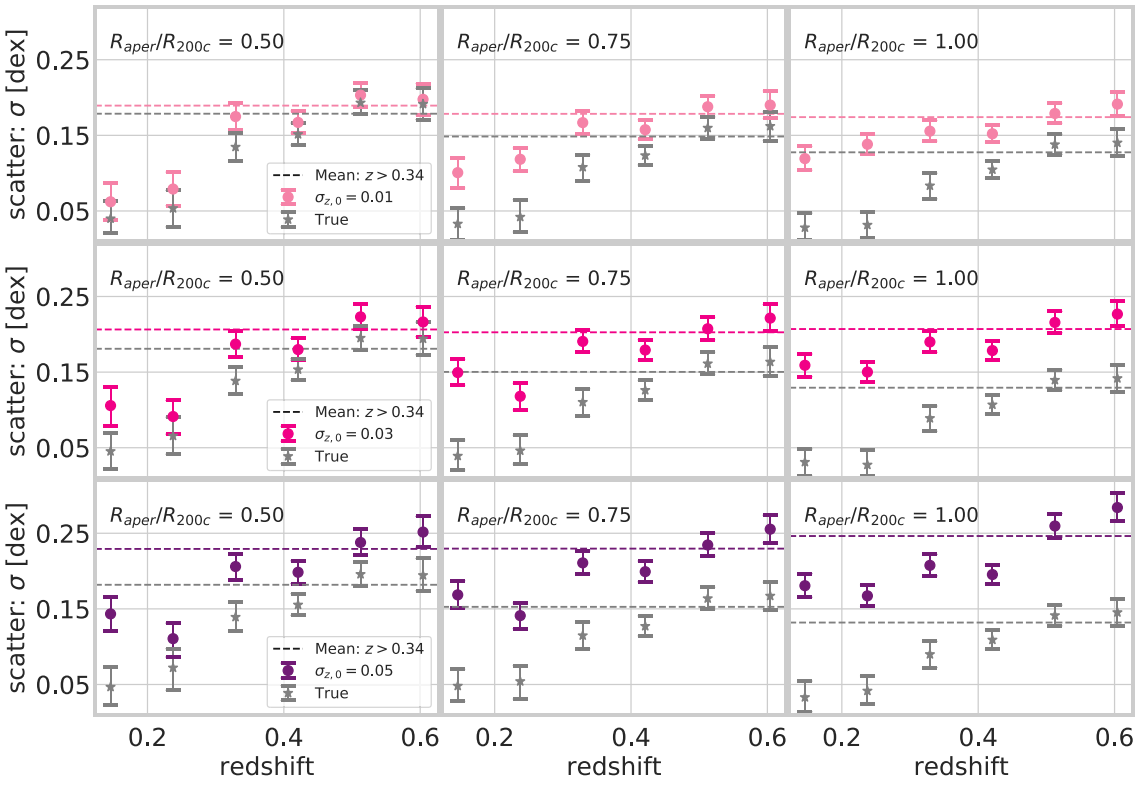

Figure 8. Scatter versus redshift for two stellar mass estimations $\mu_{\star,\rm true}$ (True, star markers) and $\mu_\star$(Copacabana, circle points). Each row displays the results for a given photo-z sample and each column for a given cluster aperture. Down in columns, the scatter increases as the uncertainty in the photometric redshifts increases, while the scatter for the true relation is fixed. The dashed lines show the mean values of the scatter in each sub-plot for visual comparison with the mean scatter of the true relation (dashed lines). The stars points are the scatter of the true relation, and therefore do not change between rows. Across a row, the scatter in the true relation decreases.

into ~ 5 per cent of galaxies having $P_{\rm mem} = 0$. This threshold avoids outliers that might be present in the colour distributions.

The membership accuracy is higher for the smaller $0.5 \times R_{200c}$ aperture (see the dashed lines in Fig. 6) and an accuracy of 89 per cent is achieved in the best scenario. For a given science case, for instance, for studies of the red sequence, Copacabana can provide an excellent membership selection without relying on previous knowledge of the red sequence.

4.4 $\mu_\star - M_{200,c}$ scaling relation

An example of the $\mu_\star - M_{200,\rm c}$ scaling relations is shown in Fig. 7, using $\mu_\star$ and $\mu_{\star,\rm true}$ as the mass proxy for clusters in the lowest redshift bin, $0.47 < z < 0.56$, and using DES-like photo-z accuracies, $\sigma_{z,0} = 0.03$. The actual relation, $\mu_{\star,\rm true} - M_{200,\rm c}$ (grey line) is consistent (within 2σ) of with the purple line that Copacabana estimated. We note some small differences between the two curves. For example, the intrinsic scatter is larger, and the slope is shallower for the purple line. These two differences hint at how membership probabilities bias our results of the actual scaling relation. In the following section, we present and discuss the impact of the quality of the photometric redshifts and size cluster aperture on the fitted parameters.

4.4.1 *Fitted parameters*

For cosmological parameter estimation, the scatter at a fixed $\mu_\star$ is the important quantity describing the halo mass function. At first order, the scatter at fixed mass proxy ($\sigma_{\log M|\log\mu_\star}$) can be written as (e.g. Evrard et al. 2014)

$$\sigma_{\log M|\log\mu_\star} = \sigma_{\log M_{200,\rm c}}/\beta \ . \tag{34}$$

A steeper slope results in a lower $\sigma_{\log M|\log\mu_\star}$ just as much as a smaller intrinsic scatter σ. For this reason, we focus on the slope β and the scatter σ since they are the important parameters for cluster cosmological analysis.

The fitted scatter and slope values as a function of redshift are shown in Figs 8 and 9, respectively. The panel displays three different photo-z precisions across rows and three different cluster apertures across columns. There is an overall shift of the Copacabana from the true values, indicated by the mean values (dashed lines). The gap, i.e. the additional shift, increases with a poorer photo-z precision and a larger cluster aperture.

The remarkable precision achieved by using all the cluster population, blue and red galaxies, in photometric data demonstrates the power of the Copacabana probabilities. We have obtained an intrinsic scatter of 0.06 dex for $z < 0.32$ and 0.19 dex $0.32 < z < 0.65$ for our mass proxy $\mu_\star$ in a DES-like photo-z sample. In particular, the fitted scatter does not change significantly with the cluster aperture. In Fig. 8, the mean values are about the same across the columns for any photo-z. While the difference between dashed lines is larger, the actual intrinsic scatter (grey stars) decreases with the cluster aperture, which counterbalances the noise added by a larger aperture. The outcome, stressed by the mean value, is that the intrinsic scatter does not depend significantly on the cluster aperture. In contrast, intrinsic

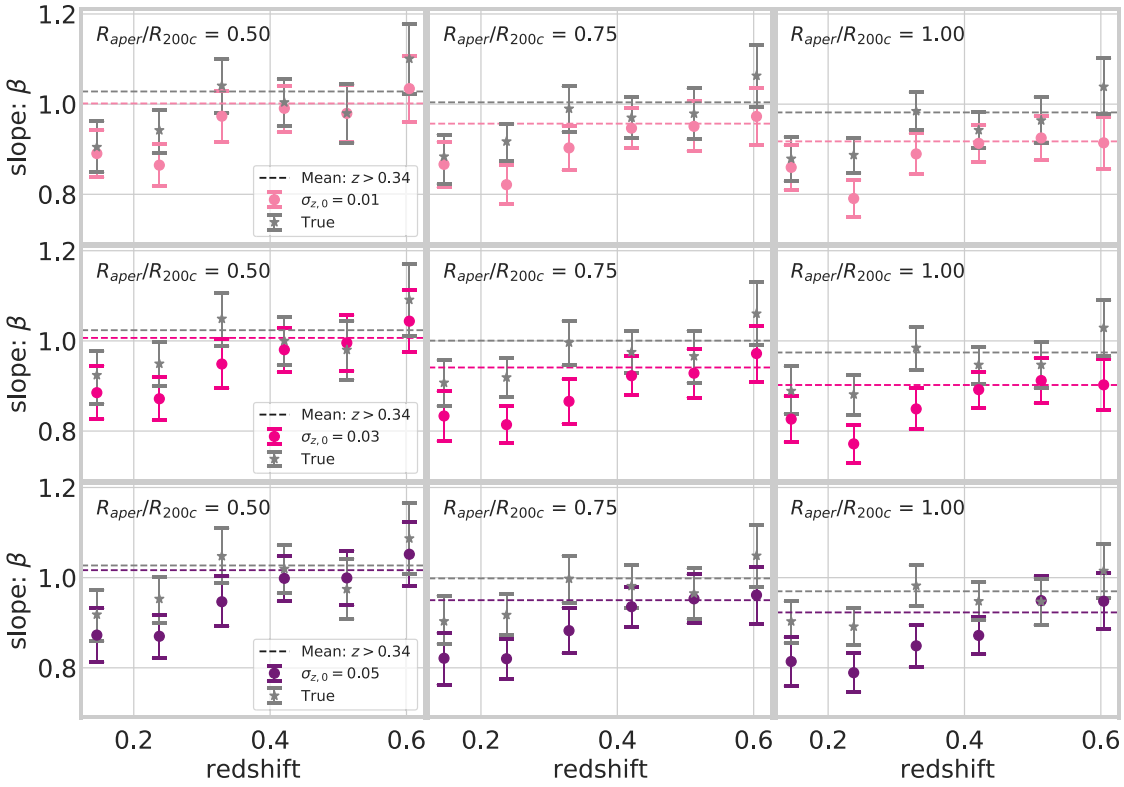

Figure 9. As in Fig. 8 but for the slope of the scaling relation. A larger cluster aperture increases the contribution from background galaxies and produces a shallower slope. As a result, the $0.5 \times R_{200c}$ aperture has the steepest slope, and it is the only case where the slope derived by Copacabana agrees with the true one irrespective of the photo-z sample used.

scatter did depend on aperture when considering the $\mu_\star$ precision in Section 4.1.1.

Regarding the slope, Fig. 9 demonstrates that a decrease in the accuracy of membership probability tends to result in a shallower slope. As discussed earlier in Section 4.1.1, larger photo-z precision and a larger aperture size can reduce the accuracy of the membership probability. This effect is most noticeable in the bottom left panel of Fig. 9, where the most significant discrepancy with the actual value is seen due to the combination of low photo-z precision and a large aperture. When we consider a specific survey scenario (represented by a given row), the mean slope (colourful dashed line) tends to be shallower when the cluster aperture is larger. This trend is reduced with better photo-z accuracy.

The simultaneous change on the slope and the scatter imply that the scatter at fixed $\mu_\star$, $\sigma_{\log M|\log\mu_\star}$, is affected by the uncertainty on the membership probabilities. Using equation (34), we can infer $\sigma_{\log M|\log\mu_\star}$ and quantify the impact of our methodology on the scaling relation parameters. Fig. 10 shows these results displayed similarly to that of Fig. 8.

A smaller cluster aperture improves the $\mu_\star$ measurement by introducing less noise to the intrinsic scatter of the observable–mass relation. However, for an optimistic photo-z sample such as LSST with $\sigma_{z,0} = 0.01$, using scatter computed at the R_{200c} is a feasible option, as suggested by the observed trend in the mean scatter value. Conversely, when dealing with an SDSS-like photo-z sample, where $\sigma_{z,0} = 0.05$, opting for a smaller cluster aperture may be the optimal approach to reduce the noise introduced by uncertainty in the redshift.

Regarding the redshift evolution of the fitted parameters, we do not see a smooth evolution with redshift as we would expect. Instead, there is a jump around redshift 0.3. This feature is also in the true underlying relation. DeRose et al. (2019) showed that the switch on the simulation resolution at $z = 0.32$ impacts the matter power spectrum of the BUZZARD data. This change in resolution might explain the redshift discontinuity seen in this work.

4.4.2 $\mu_\star$ *as a low scatter mass proxy*

Let us consider the case where all the true members inside R_{200c} are known and focus on the actual $\mu_{\star,\rm true}$– $M_{200,c}$ relation, i.e. without photo-z errors. Here, we can see the ideal stellar-mass-based mass proxy. The grey points in Fig. 8 show the scatter, and the $\mu_{\star,\rm true}$–$M_{200,c}$ relation exhibits a low scatter. At the lowest redshift bins, σ is about 0.05 dex, a comparable value to the intracluster medium, e.g. $Y_{\rm SZ}$, $M_{\rm gas}$ (Kravtsov, Vikhlinin & Nagai 2006; Mulroy et al. 2019; Bleem et al. 2020), and redshift-based mass proxies, e.g. velocity dispersions (Ruel et al. 2014). At higher redshifts $z > 0.32$, the scatter increases, though we believe this is a simulation effect as the switch of box resolution of the BUZZARD simulation occurs at $z = 0.33$.

When compared with other mass proxies from simulations, the typical scatter for the number of satellites lies between $0.15 - 0.20$ dex (Anbajagane et al. 2020), for X-ray luminosity is 0.114 dex and, gas mass is 0.037 dex, respectively (Ho et al. 2023).

The $\mu_\star$–M_{200c} scaling relation from the Buzzard simulation aligns well with the stellar–halo mass relation (SHMR) derived from several high-resolution hydrodynamical simulations, including BAHAMAS + MACSIS, TNG300 from the IllustrisTNG suite, and Magneticum Pathfinder (Anbajagane et al. 2020). It is important to note that while the stellar mass in our study is based on photometric measurements,

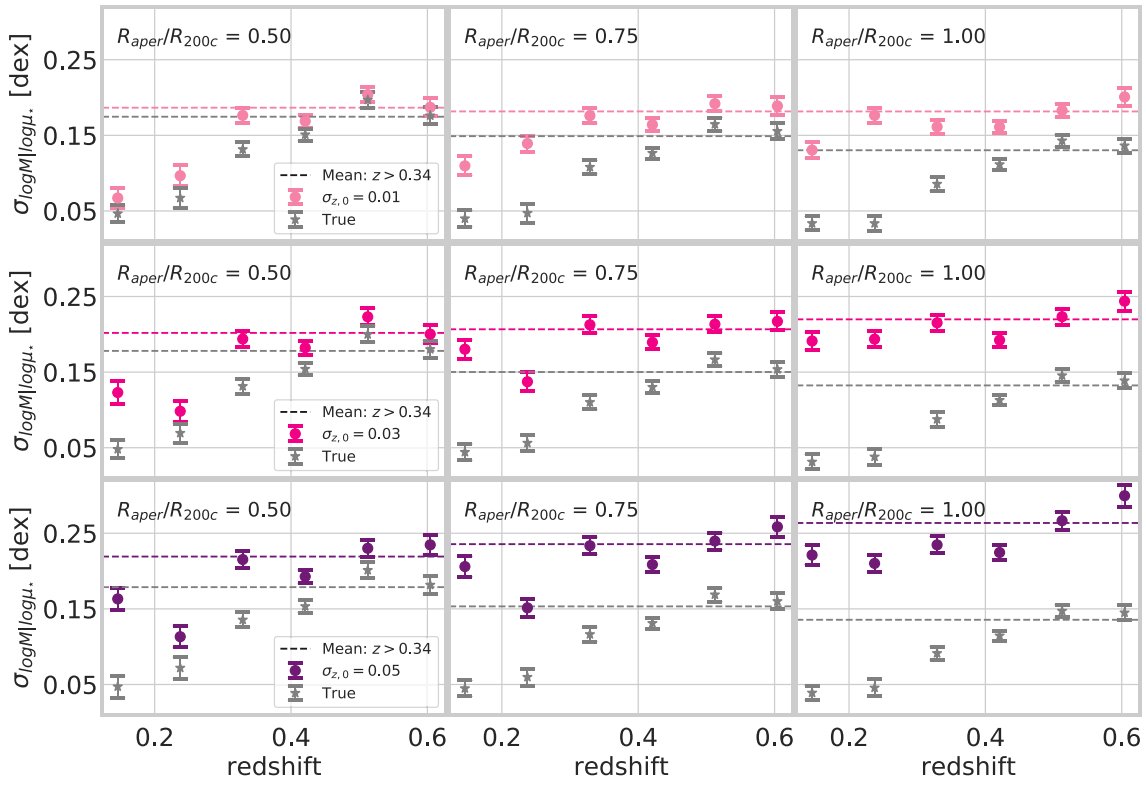

Figure 10. Scatter at fixed observable relation versus the redshift for two independent variables $\mu_{\star,\mathrm{true}}$(True) and $\mu_\star$(Copacabana). Each row displays a result for a given photo-z sample, and each column for a given cluster aperture. The photo-z precision impacts the scatter of the recovered $\mu_\star$–$M_{200,c}$ scaling relation for any given cluster aperture. The scatter at fixed $\mu_\star$ has a trend with the cluster aperture. The gap between the two dashed lines increases across the rows and columns. Note: the stars points are fixed between rows.

the hydrodynamical simulations measure stellar particles directly. Despite these differences, the comparison reveals that the intrinsic scatter of the SHMR is consistently below 0.05 dex (in base-10 log scale) at $z = 0$ for high-mass haloes, a value comparable to our lowest redshift results in Fig. 10.

In terms of the slope of the SHMR, there is some variation across the simulations, ranging from 0.85 to 1.0. In contrast, the slope of the $\mu_\star$–M_{200c} relation is slightly lower at low redshifts, varying between 0.8 and 0.9. However, at $z > 0.32$, the slope aligns well with that of the hydrodynamical simulations, as shown in Fig. 9. Overall, the agreement between the hydrodynamical simulations and the BUZZARD results suggests that the $\mu_\star$ scaling relation is a reliable proxy for the total stellar mass scaling relation.

5 CONCLUSIONS

This work presents the Copacabana algorithm, which assigns membership probabilities to galaxies in a given cluster field. We validated the algorithm using the Buzzard simulation. In particular, the stellar mass of cluster galaxies weighted by the membership probabilities, $\mu_\star$, was found to have a steep cluster mass scaling relation with an intrinsic scatter of 0.06 dex at low redshift for photometric redshifts that have photo-z uncertainties comparable to that expected in LSST data. In addition, we show that our methodology could precisely recover the scaling relation between $\mu_\star$ and cluster mass, indicating that $\mu_\star$ can be a competitive mass proxy for optically selected clusters in future cosmological analysis.

(i) **Performance:** we show the uncertainty in $\mu_\star$ is mainly affected by the quality of the photometric redshifts, followed by the cluster aperture. In the best case, photometric redshifts with LSST-like accuracies ($\sigma_{z,0} = 0.01$), we reported a $\mu_\star$ uncertainty of 0.09 dex. A smaller aperture, for instance, $0.5 \times R_{200c}$, leads to a smaller uncertainty ~ 0.05 dex.

(ii) **Cluster size:** We present a new method to measure cluster size, R_{200c}, with only photometric data. The procedure is based on the HOD relation. We inferred a precision of 30 per cent in the context of the Buzzard simulations. The accuracy of our estimate does not depend on the quality of the photometric redshifts or the halo mass.

(iii) **Impact on halo mass estimations:** We quantify how the $\mu_\star$ uncertainty propagates to estimates of the halo mass. Specifically, we study the parameters of the scaling relation $\mu_\star - M_{200c}$ with a focus on the slope and the intrinsic scatter. The photo-z uncertainty is the main parameter in affecting the scaling relation parameters. For instance, in the LSST-like scenario, we recover the parameters with no significant difference compared to the true relation. In contrast, in the scenario of large photo-z uncertainties and large apertures, there was a significant impact on recovering the true parameters.

(iv) **How the cluster aperture impacts the halo mass:** While the accuracy of the photometric redshifts is survey-dependent, the aperture size can be adjusted to suit the scientific objectives. For example, a smaller cluster aperture can significantly improve the membership probabilities. However, that improvement does not necessarily translate into gains in predicting the halo mass. For instance, we do not find a substantial improvement in the scatter of the scaling relation, and the precision gained by a smaller aperture is counterbalanced by the larger intrinsic scatter in $\mu_\star - M_{200c}$, which is minimal at R_{200c}. On the other hand, we find that the aperture size significantly affects the slope. In particular, at an aperture of

$0.5 \times R_{200c}$, the recovered slope is very close to the true slope and is insensitive to the range of photometric redshift uncertainties that one encounters in modern photometric surveys. While at R_{200c}, the recovered slope can be up to 20 per cent shallower.

(v) **Membership probability performance:** We report our galaxy member selection in terms of Purity (P) and Completeness (C). In our best scenario, the accuracy was 81 per cent with P and C of 64 per cent and 93 per cent; when we consider a smaller aperture ($0.5 \times R_{200c}$), these values were 89 per cent, 79 per cent, and 94 per cent, respectively. The membership probabilities substantially improved with a smaller cluster aperture.

(vi) **The impact of stellar masses errors:** The typical uncertainty in stellar mass is approximately 0.20 dex (Palmese et al. 2020) which we found to be larger than the uncertainty associated with membership probabilities. This stellar mass error introduces additional scatter in the scaling relation $\mu_\star - M_{200c}$, contributing with an additional uncertainty of 0.07 ± 0.01 dex to $\mu_\star$. This emphasizes the significant role that stellar mass uncertainties play in the overall error budget of $\mu_\star$, highlighting the need to minimize this source of noise for more precise cluster mass estimates.

In sum, Copacabana is a powerful tool to predict galaxies' total stellar mass content in clusters. In future work, we will apply the stellar mass-proxy $\mu_\star$ in a cosmological analysis using optical data.

ACKNOWLEDGEMENTS

JHE and MSS acknowledge the support by the U.S. Department of Energy (DOE) under grant No. DE-SC0019193. MESP acknowledges support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2121 ``Quantum Universe" – 390833306.

This research uses resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory.

Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft, and the Collaborating Institutions in the Dark Energy Survey.

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, NSF's NOIRLab, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, Texas A&M University, and the OzDES Membership Consortium.

Based in part on observations at Cerro Tololo Inter-American Observatory at NSF's NOIRLab (NOIRLab Prop. ID 2012B-0001; PI: J. Frieman), which is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

The DES data management system is supported by the National Science Foundation under grant numbers AST-1138766 and AST-1536171. The DES participants from Spanish institutions are partially supported by MICINN under grants ESP2017-89838, PGC2018-094773, PGC2018-102021, SEV-2016-0588, SEV-2016-0597, and MDM-2015-0509, some of which include ERDF funds from the European Union. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya. Research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013) including ERC grant agreements 240672, 291329, and 306478. We acknowledge support from the Brazilian Instituto Nacional de Ciência e Tecnologia (INCT) do e-Universo (CNPq grant 465376/2014-2).

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

DATA AVAILABILITY

The data used in this work are available upon request.

REFERENCES

Abbott T. M. C. et al., 2020, Phys. Rev. D, 102, 023509
Aguena M. et al., 2021, MNRAS, 502, 4435
Albrecht A. et al., 2006, preprint (arXiv:astro-ph/0609591)
Allen S. W., Evrard A. E., Mantz A. B., 2011, ARA&A, 49, 409
Anbajagane D., Evrard A. E., Farahi A., Barnes D. J., Dolag K., McCarthy I. G., Nelson D., Pillepich A., 2020, MNRAS, 495, 686
Andreon S., 2012, A&A, 548, A83
Andreon S., Quintana H., Tajer M., Galaz G., Surdej J., 2006, MNRAS, 365, 915
Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379
Bellagamba F., Roncarelli M., Maturi M., Moscardini L., 2018, MNRAS, 473, 5221
Bellagamba F. et al., 2019, MNRAS, 484, 1598
Black W. K., Evrard A. E., 2024, Open J. Astrophys., 7, 25
Bleem L. E., Bocquet S., Stalder B., Gladders M. D., Ade P. A. R., Allen S. W., 2020, ApJS, 247, 25
Bradshaw C., Leauthaud A., Hearin A., Huang S., Behroozi P., 2020, MNRAS, 493, 337
Butcher H., Oemler A. J., 1984, ApJ, 285, 426
Carliles S., Budavári T., Heinis S., Priebe C., Szalay A. S., 2010, ApJ, 712, 511
Castignani G., Benoist C., 2016, A&A, 595, A111
Conroy C., Gunn J. E., 2010, ApJ, 712, 833
Conroy C., Wechsler R. H., Kravtsov A. V., 2006, ApJ, 647, 201
Conroy C., Gunn J. E., White M., 2009, ApJ, 699, 486
Cooper M. C. et al., 2007, MNRAS, 376, 1445
Costanzi M. et al., 2019, MNRAS, 482, 490
De Lucia G. et al., 2007, MNRAS, 374, 809
DeRose J. et al., 2019, preprint (arXiv:1901.02401)

DeRose J. et al., 2022, Phys. Rev. D, 105, 123520
Dodelson S., Heitmann K., Hirata C., Honscheid K., Roodman A., Seljak U., Slosar A., Trodden M., 2016, preprint (arXiv:1604.07626)
Evrard A. E., Arnault P., Huterer D., Farahi A., 2014, MNRAS, 441, 3562
George M. R. et al., 2011, ApJ, 742, 125
Golden-Marx J. B. et al., 2023, MNRAS, 521, 478
Gonzalez A. H., Zaritsky D., Zabludoff A. I., 2007, ApJ, 666, 147
Gonzalez A. H., Sivanandam S., Zabludoff A. I., Zaritsky D., 2013, ApJ, 778, 14
Gschwend J. et al., 2018, Astron. Comput., 25, 58
Hennig C. et al., 2017, MNRAS, 467, 4015
Ho M., Soltis J., Farahi A., Nagai D., Evrard A., Ntampaka M., 2023, MNRAS, 524, 3289
Huang S. et al., 2020, MNRAS, 492, 3685
Huang S. et al., 2022, MNRAS, 515, 4722
Huterer D., 2023, A&AR, 31, 2
Kelly B. C., 2007, ApJ, 665, 1489
Klein M., Oguri M., Mohr J. J., Grandis S., Ghirardini V., Liu T., Liu A., Bulbul E., 2022, A&A, 661, A4
Koester B. P. et al., 2007, ApJ, 660, 239
Korytov D. et al., 2019, ApJS, 245, 26
Kravtsov A. V., Vikhlinin A., Nagai D., 2006, ApJ, 650, 128
Kravtsov A. V., Vikhlinin A. A., Meshcheryakov A. V., 2018, Astron. Lett., 44, 8
Kukstas E., McCarthy I. G., Baldry I. K., Font A. S., 2020, MNRAS, 496, 2241
LSST Science Collaboration, 2009, preprint (arXiv:0912.0201)
Laganá T. F., Martinet N., Durret F., Lima Neto G. B., Maughan B., Zhang Y. Y., 2013, A&A, 555, A66
Laigle C. et al., 2016, ApJS, 224, 24
Leauthaud A. et al., 2012, ApJ, 744, 159
Lehmann B. V., Mao Y.-Y., Becker M. R., Skillman S. W., Wechsler R. H., 2017, ApJ, 834, 37
Lopes P. A. A., Ribeiro A. L. B., 2020, MNRAS, 493, 3429
McClintock T. et al., 2019, MNRAS, 482, 1352
Mulroy S. L. et al., 2019, MNRAS, 484, 60
Myles J. et al., 2021, MNRAS, 505, 33
Nelson D. et al., 2019, Comput. Astrophys. Cosmol., 6, 2
Palmese A., Lahav O. et al., 2016, MNRAS, 463, 1486
Palmese A. et al., 2020, MNRAS, 493, 4591
Pereira M. E. S. et al., 2018, MNRAS, 474, 1361
Pereira M. E. S., Palmese A., Soares-Santos M., DES Collaboration, 2020, MNRAS, 498, 5450
Pratt G. W., Arnaud M., Biviano A., Eckert D., Ettori S., Nagai D., Okabe N., Reiprich T. H., 2019, Space Sci. Rev., 215, 25
Puddu E. et al., 2021, A&A, 645, A9
Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, ApJ, 771, 30
Rozo E. et al., 2009, ApJ, 699, 768
Ruel J. et al., 2014, ApJ, 792, 45
Rykoff E. S. et al., 2012, ApJ, 746, 178
Rykoff E. S. et al., 2014, ApJ, 785, 104
Rykoff E. S., Rozo E., DES Collaboration, 2016, ApJS, 224, 1
Sarazin C. L., 1986, Rev. Mod. Phys., 58, 1
Scott D. W., 1992, Multivariate Density Estimation: Theory, Practice and Visualization. John Wiley & Sons, Inc., New York
Song H., Hwang H. S., Park C., Tamura T., 2017, ApJ, 842, 88
Springel V. et al., 2005, Nature, 435, 629
Sunayama T. et al., 2020, MNRAS, 496, 4468
Taylor E. N. et al., 2011, MNRAS, 418, 1587
The Dark Energy Survey Collaboration, 2005, preprint (astro-ph/0510346)
Tinker J. L. et al., 2012, ApJ, 745, 16
Umetsu K., 2020, A&AR, 28, 7
Varga T. N. et al., 2019, MNRAS, 489, 2511
Wechsler R. H., DeRose J., Busha M. T., Becker M. R., Rykoff E., Evrard A., 2022, ApJ, 931, 145
Wright C. O., Brainerd T. G., 2000, ApJ, 534, 34
Wu H.-Y., Weinberg D. H., Salcedo A. N., Wibking B. D., 2021, ApJ, 910, 28
Wu H.-Y. et al., 2022, MNRAS, 515, 4471
York D. G. et al., 2000, AJ, 120, 1579
Zacharegkas G. et al., 2022, MNRAS, 509, 3119

[1]*Department of Physics, Harvard University, Cambridge, MA 02138, USA*
[2]*Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA*
[3]*Hamburger Sternwarte, Universität Hamburg, Gojenbergsweg 112, D-21029 Hamburg, Germany*
[4]*Department of Physics, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland*
[5]*Fermi National Accelerator Laboratory, P. O. Box 500, Batavia, IL 60510, USA*
[6]*Observational Cosmology Lab, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA*
[7]*Department of Astronomy, University of Maryland, College Park, MD 20742, USA*
[8]*Department of Physics, Boise State University, Boise, ID 83725, USA*
[9]*Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15312, USA*
[10]*Departments of Statistics and Data Science, University of Texas at Austin, Austin, TX 78757, USA*
[11]*Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, E-28049 Madrid, Spain*
[12]*Laboratório Interinstitucional de e-Astronomia – LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ 20921-400, Brazil*
[13]*Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK*
[14]*Faculty of Physics, University Observatory, Ludwig-Maximilians-Universität, Scheinerstr 1, D-81679 Munich, Germany*
[15]*Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, UK*
[16]*Instituto de Astrofisica de Canarias, E-38205 La Laguna, Tenerife, Spain*
[17]*Departamento de Astrofísica, Universidad de La Laguna, E-38206 La Laguna, Tenerife, Spain*
[18]*Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, E-08193 Bellaterra (Barcelona), Spain*
[19]*Astronomy Unit, Department of Physics, University of Trieste, via Tiepolo 11, I-34131 Trieste, Italy*
[20]*INAF-Osservatorio Astronomico di Trieste, via G. B. Tiepolo 11, I-34143 Trieste, Italy*
[21]*Institute for Fundamental Physics of the Universe, Via Beirut 2, I-34014 Trieste, Italy*
[22]*Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid E-28040, Spain*
[23]*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr, Pasadena, CA 91109, USA*
[24]*Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA*
[25]*Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA*
[26]*Center for Astrophysical Surveys, National Center for Supercomputing Applications, 1205 West Clark St, Urbana, IL 61801, USA*
[27]*School of Mathematics and Physics, University of Queensland, Brisbane, QLD 4072, Australia*
[28]*University of California, Santa Cruz, Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA*
[29]*Department of Physics, The Ohio State University, Columbus, OH 43210, USA*
[30]*Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA*
[31]*Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA*

[32]*Australian Astronomical Optics, Macquarie University, North Ryde, NSW 2113, Australia*
[33]*Lowell Observatory, 1400 Mars Hill Rd, Flagstaff, AZ 86001, USA*
[34]*Centre for Gravitational Astrophysics, College of Science, The Australian National University, ACT 2601, Australia*
[35]*The Research School of Astronomy and Astrophysics, Australian National University, ACT 2601, Australia*
[36]*Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, CP 66318, São Paulo, SP 05314-970, Brazil*
[37]*George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, and Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA*
[38]*LPSC Grenoble-53, Avenue des Martyrs, F-38026 Grenoble, France*
[39]*Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain*
[40]*Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA*
[41]*Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ 20921-400, Brazil*
[42]*Kavli Institute for Particle Astrophysics & Cosmology, Stanford University, P. O. Box 2450, Stanford, CA 94305, USA*
[43]*SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*
[44]*Department of Physics and Astronomy, Pevensey Building, University of Sussex, Brighton BN1 9QH, UK*
[45]*Instituto de Física, UFRGS, Caixa Postal 15051, Porto Alegre, RS 91501-970, Brazil*
[46]*School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK*
[47]*Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA*
[48]*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*
[49]*Department of Physics, Duke University Durham, NC 27708, USA*

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