

Comparing different mass estimators for a large subsample of the *Planck*-ESZ clusters

L. Lovisari^{1,2}, S. Ettori^{1,3}, M. Sereno^{1,3}, G. Schellenberger², W. R. Forman², F. Andrade-Santos², and C. Jones²

¹ INAF – Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Piero Gobetti 93/3, 40129 Bologna, Italia
e-mail: lorenzo.lovisari@inaf.it

² Center for Astrophysics Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA

³ INFN, Sezione di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italia

Received 22 June 2020 / Accepted 6 October 2020

ABSTRACT

Context. Total mass is arguably the most fundamental property for cosmological studies with galaxy clusters. The individual cluster masses can be obtained with different methods, each with its own biases and limitations. Systematic differences in mass measurements can strongly impact the determination of the hydrostatic bias and of the mass-observable relations, key requirements of many cluster abundance studies.

Aims. We investigate the present differences in the mass estimates obtained through independent X-ray, weak-lensing, and dynamical studies using a large subsample of the *Planck*-ESZ clusters. We also discuss the implications for mass bias analyses.

Methods. After assessing the systematic differences in the X-ray-derived masses reported by distinct groups, we examine the mass estimates obtained with independent methods and quantify the differences as the mean ratio $1-b = M_{\text{HE}}/M_{\text{WL,dyn}}$, where HE refers to hydrostatic masses obtained from X-ray observations, WL refers to the results of weak-lensing measurements, and dyn refers to the mass estimates either from velocity dispersion or from the caustic technique. So defined, the $1-b$ parameter includes all possible astrophysical, observational, and methodological biases in one single value.

Results. Recent X-ray masses reported by independent groups show average differences smaller than $\sim 10\%$, posing a strong limit on the systematics that can be ascribed to the differences in the X-ray analysis when studying the hydrostatic bias. The mean ratio between our X-ray masses and the weak-lensing masses in the LC²-single catalog is $1-b = 0.74 \pm 0.06$, which corresponds to a mass bias of $26 \pm 6\%$, a value insufficient to reconcile the *Planck* cluster abundance and cosmic microwave background results. However, the mean mass ratios inferred from the WL masses of different projects vary by a large amount, with APEX-SZ showing a bias consistent with zero ($1-b = 1.02 \pm 0.12$), LoCuSS and CCCP/MENeCS showing a significant difference ($1-b = 0.76 \pm 0.09$ and $1-b = 0.77 \pm 0.10$, respectively), and WtG pointing to the largest deviation ($1-b = 0.61 \pm 0.12$), which would substantially reduce the tension between the *Planck* results. Because of small differences between our $M - Y_X$ relation and the one used by the *Planck* collaboration, our X-ray masses are on average 7% lower (4% at the same physical radius) than the *Planck* masses and can further reduce the required bias. At odds with the WL results, the dynamical mass measurements show better agreement with the X-ray hydrostatic masses, although there are significant differences when relaxed or disturbed clusters are used. However, the comparison is currently limited by the small sample sizes.

Conclusions. The systematic differences between total masses obtained with recent independent X-ray analyses are smaller than those found in previous studies. This shifts the focus to WL and dynamical studies for a better convergence of the level of mass bias. However, the ratios obtained using different mass estimators suggest that there are still systematics that are not accounted for in all the techniques used to recover cluster masses. This prevents the determination of firm constraints on the level of hydrostatic mass bias in galaxy clusters.

Key words. X-rays: galaxies: clusters – galaxies: clusters: general – galaxies: clusters: intracluster medium – gravitational lensing: weak

1. Introduction

Galaxy cluster abundances provide sensitive constraints on the cosmological parameters that govern the expansion history of the Universe (e.g., Vikhlinin et al. 2009; Allen et al. 2011; Kravtsov & Borgani 2012). Typical galaxy cluster number count experiments infer total masses from a mass-observable scaling relation, which in turn requires calibration of the mass-proxy bias and scatter (e.g., see Planck Collaboration XXIV 2016; Hilton et al. 2018; Bocquet et al. 2019, for Sunyaev-Zel'dovich-based experiments). Thus, to use clusters for cosmological studies, accurate total masses for large samples of clusters with well-understood selection criteria are required.

One surprising result from the *Planck* collaboration is that the σ_8 and Ω_M derived from Sunyaev-Zel'dovich (SZ, Sunyaev & Zeldovich 1972) cluster counts are inconsistent with

the values derived from the *Planck* cosmic microwave background (CMB) cosmology (see Planck Collaboration XXIV 2016). The tension would be removed if clusters were 60–85% more massive than estimates based on scaling relations. Although a redshift dependence could alleviate the tension (Sereno & Ettori 2017), this level of bias is much larger than expected from numerical simulations which suggest that X-ray-derived masses, based on hydrostatic equilibrium, are biased low by up to 30% (e.g., Piffaretti & Valdarnini 2008; Meneghetti et al. 2010; Rasia et al. 2012, 2014; Biffi et al. 2016; Ansarifard et al. 2020; Barnes et al. 2020). This level of bias is also not fully supported by observations, although different studies find quite different results hampered by the different mass and redshift ranges and related selection effects explored in each analysis, as well as the fact that analyses are performed with different methodologies. For instance, assuming that weak-lensing (WL) reconstruction

provides unbiased measurements of the true mass, von der Linden et al. (2014), with a sample of $N=22$ clusters with mass and redshift ranges of $M = [0 - 30] \times 10^{14} M_{\odot}$ and $z = [0.15 - 0.7]$, found a bias $1-b = M/M_{\text{true}} = 0.69 \pm 0.07$, which is consistent at the 1σ level with the bias required by Planck. Penna-Lima et al. (2017) ($N = 21, M = [3 - 11] \times 10^{14} M_{\odot}$ and $z = [0.19 - 0.89]$) found a similar result (i.e., $1-b = 0.73 \pm 0.10$). Herbonnet et al. (2020) ($N = 100, M = [0.5 - 20] \times 10^{14} M_{\odot}$ and $z = [0.05 - 0.55]$) and Medezinski et al. (2018) ($N = 5, M = [2 - 30] \times 10^{14} M_{\odot}$ and $z = [0.06 - 0.33]$) found a smaller bias (0.84 ± 0.04 and 0.80 ± 0.14 , respectively), while the results of Ettori et al. (2019) ($N = 13, M = [2 - 9] \times 10^{14} M_{\odot}$ and $z = [0.05 - 0.09]$), Smith et al. (2016) ($N = 50, M = [2 - 15] \times 10^{14} M_{\odot}$ and $z = [0.15 - 0.3]$), and Israel et al. (2014) ($N = 8, M = [1 - 5] \times 10^{14} M_{\odot}$ and $z = [0.4 - 0.8]$) point to a bias of only a few percent. Many other studies have investigated this subject and found a mass bias in the above-mentioned range (e.g., see Zhang et al. 2010; Mahdavi et al. 2013; Gruen et al. 2014; Hoekstra et al. 2015; Applegate et al. 2016; Miyatake et al. 2019). Thus, while assessing the level of mass bias in galaxy clusters is a fundamental step for their application to precision cosmology, the magnitude of the bias is still an open issue.

Under the assumption of hydrostatic equilibrium, the masses for individual clusters of galaxies can be derived from the distribution of their X-ray emitting gas. However, for most of the clusters, the hydrostatic equilibrium assumption is probably invalid, at least mildly, making hydrostatic mass determinations more prone to biases. The more disturbed the cluster is, the more underestimated the X-ray cluster mass is expected to be. In fact, the major source of bias is associated with bulk gas motions which act as additional pressure support against gravity (e.g., Rasia et al. 2006, 2012; Piffaretti & Valdarnini 2008; Lau et al. 2009; Biffi et al. 2016; Vazza et al. 2018). On top of that, the presence of temperature inhomogeneities and/or multi-phase gas can bias the estimate from X-ray instruments because of the higher weight of the colder gas components (e.g., see Mazzotta et al. 2004). Also, the bias is found to be higher in the outskirts because of the increasing contribution of gas motions and temperature inhomogeneities (e.g., Rasia et al. 2012).

At the moment, WL is considered the most reliable method to determine accurate masses because it measures the total mass directly, without relying on baryonic tracers. Hence, the lensing signal does not require any assumption about the dynamical state of the gravitating mass of the cluster. Motivated by the results of numerical simulations (e.g., Becker & Kravtsov 2011; Oguri & Hamana 2011; Rasia et al. 2012) that show that the average WL masses are biased low by less than 10%, many cluster mass studies used WL measurements to constrain cluster masses. Nonetheless, individual systems can suffer significant biases depending on the observer's viewing angle with respect to the cluster mass distribution and the presence of substructures (see, e.g., Meneghetti et al. 2010; Sereno et al. 2018, and Pratt et al. 2019 for more details). Moreover, the triaxial distribution of mass introduces a scatter of up to 30% in lensing masses for individual clusters. This quite large scatter implies that samples of a few dozens of clusters are required to constrain the normalization of scaling relations to better than 10% (e.g., Becker & Kravtsov 2011).

Techniques based on galaxy dynamics, such as the caustic technique (e.g., Diaferio & Geller 1997; Diaferio 1999) and methods using either the Jeans equation or the virial theorem (e.g., Biviano et al. 2013; Amodeo et al. 2017), can also be used to estimate total masses, although these methods are impeded

by the expensive observational requirements to obtain spectroscopic measurements of galaxies velocities. Similar to WL estimates, the total masses obtained with the caustic technique do not rely on the hypothesis of dynamical equilibrium which is instead required by the Jeans and virial estimators. In the literature, there are fewer caustic studies than those using WL measurements to compare hydrostatic and dynamical masses for relatively large samples. Maughan et al. (2016) found a small or zero value of the hydrostatic bias if the caustic masses are assumed to be equivalent to the true mass, and Rines et al. (2016) found a similar result comparing SZ mass estimates (calibrated with hydrostatic X-ray masses) with the caustic masses. On the contrary, Foëx et al. (2017) found that the dynamical masses are on average larger than the hydrostatic values with the ratio between the Jeans, virial theorem, and caustic estimators, with respect to the hydrostatic mass, being $1-b = 1.22 \pm 0.18$, $1-b = 1.51 \pm 0.26$, and $1-b = 1.32 \pm 0.18$, respectively.

Each of the above-mentioned approaches requires various hypotheses and suffers from different systematics (see, e.g., the review by Pratt et al. 2019 and references therein). Thus, an inter-comparison of the different results can provide useful information to better understand their respective systematics. In this paper, we examine the differences in such measurements for a large sample of Planck-selected clusters analyzed in X-rays in Lovisari et al. (2017, hereafter L17) and Lovisari et al. (2020, hereafter L20). The comparison will also shed light on the magnitude of the hydrostatic equilibrium mass bias, and whether the value $1-b = 0.58 \pm 0.04$, required to reconcile the Planck CMB with number counts (Planck Collaboration XXIV 2016, see also Salvati et al. 2019), is consistent with a significant underestimate of the X-ray-derived masses. In fact, the Planck team estimated cluster masses using the SZ signal Y_{SZ}^1 calibrated against hydrostatic masses. Therefore, a significant bias would help to reconcile the Planck results.

The paper is organized as follows. In Sect. 2, we introduce our sample and describe the data analysis. In Sect. 3, we present our results, and, in Sect. 4, we discuss our findings. A summary is provided in Sect. 5. In Appendix A we compare the biases computed with different estimators, and in Appendix B we show the comparison between the X-ray masses derived in different studies.

Throughout this paper, we assume a flat Λ CDM cosmology with $\Omega_m = 0.3$ and $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$. Uncertainties are at the 68% confidence level. Log is always base 10.

2. Data and analysis

2.1. Sample

The Planck Early Sunyaev-Zel'dovich (ESZ, Planck Collaboration VIII 2011) sample consists of 188 massive clusters, which were selected by imposing a signal-to-noise ratio threshold of 6 and a Galactic cut $|b| > 14^\circ$ on the catalog of SZ detections from the first ten months of the Planck survey. Although SZ-selected samples suffer from Malmquist bias and are not fully mass-selected, they are considered representative of the underlying cluster population. In fact, simulations have shown that SZ quantities do not strongly depend on the dynamical state of the clusters (e.g., Motl et al. 2005), with only a modest effect due to mergers (e.g., Battaglia et al. 2012).

¹ The Y_{SZ} parameter corresponds to the Compton y -profile integrated within a sphere of radius R_{500} , the radius within which the mean overdensity of a cluster is 500 times the critical density of the Universe at the cluster redshift.

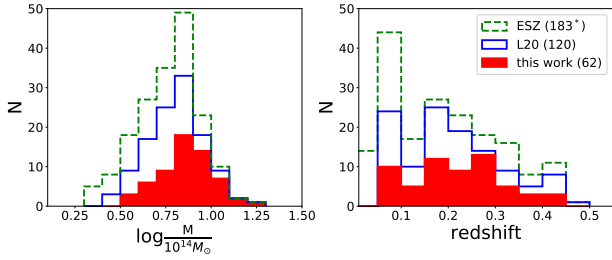


Fig. 1. Distribution of the cluster masses (*left panel*) and redshifts (*right panel*) for the ESZ clusters (dashed green), the L20 sample (solid blue), and the clusters with weak-lensing masses (filled red). Numbers in the legend represent the sample size.

This is supported observationally by the tight correlation between SZ signal and mass (e.g., [Planck Collaboration XI 2011](#)). Moreover, the morphology of the source has little impact on the detection procedure in the *Planck* survey, as shown via MonteCarlo simulations ([Planck Collaboration XXVII 2016](#)).

L17 and L20 analyzed the *XMM-Newton* data to determine the morphological and global cluster properties for a representative subsample of 120 ESZ clusters, selected to ensure that R_{500} is completely covered by *XMM-Newton* observations (either single or multiple pointings). A single pipeline was used in the analysis allowing a uniform processing of the data, eliminating one source of uncertainties in the comparison with the masses estimated from other methods.

For all these 120 clusters, we know the dynamical state and both X-ray and SZ masses (from L20 and [Planck Collaboration XXVII 2016](#), respectively). Instead of the SZ masses estimated in [Planck Collaboration VIII \(2011\)](#), we preferred to use the more recent and more accurate PSZ2 masses reported in [Planck Collaboration XXVII \(2016\)](#) obtained from the 29 month full *Planck* mission data. We note that five ESZ clusters are not included in the PSZ2 catalog because they fall into the PSZ2 point source mask, but only three of these overlap with our subsample of 120 clusters. For these clusters we use the PSZ1 masses.

For WL masses, we used the compilation obtained in the CoMaLit framework (see, e.g., [Sereno & Ettori 2015](#)). The Literature Catalogs of Weak Lensing Clusters (LC², [Sereno 2015](#)) is publicly available². LC²-single (v. 3.9) is a meta-catalog with 806 unique entries. When multiple analyses per cluster are available, preference is given to studies exploiting the deeper observations and multiband optical coverage for optimal galaxy background selection. There are 62 clusters in common with the sample analyzed in L20. The distribution of mass and redshift for these clusters with respect to the ESZ and L20 samples is shown in Fig. 1.

Given the relatively large sample, we also investigated whether or not the X-ray/SZ masses of disturbed clusters are significantly more biased than the masses derived for relaxed clusters. The dynamical state of each cluster was determined by using two morphological parameters: one sensitive to the core-properties (i.e., surface brightness concentration) and one sensitive to the large-scale inhomogeneities (i.e., centroid-shift). In L20 (see also [Rasia et al. 2013](#)), these two parameters have been combined to define a new general parameter M (hereafter M_{par}), which allows us to use a single value to rank the clusters from the most relaxed to the most disturbed in the sample of interest. In the following, we refer to relaxed and disturbed cluster when M_{par} is greater than and smaller than zero, respectively. This

classification provides a simple way to divide a sample between the most relaxed and most disturbed clusters but does not necessarily provide an absolute reference.

2.2. Estimation of the hydrostatic bias

We estimated the mass ratio by fitting the data with LIRA which allows normalization, slope, and scatters (and relative uncertainties) to be fitted simultaneously³. In particular, for each subsample of WL (or dynamical) masses we linearly fit our data as

$$\log\left(\frac{M_{\text{HE}}}{6 \times 10^{14} M_{\odot}}\right) = \alpha + \beta \log\left(\frac{M_{\text{true}}}{6 \times 10^{14} M_{\odot}}\right) \pm \sigma_{\text{HE}|\text{true}}, \quad (1)$$

$$\log(M_{\text{true}}) = \log(M_{\text{WL}}) \pm \sigma_{\text{WL}|\text{true}}$$

where α quantifies the mass ratio (i.e., $1-b = M_{\text{HE}}/M_{\text{WL}} = 10^{\alpha}$), β is fixed to 1 (unless otherwise specified), and the M_{HE} and M_{WL} masses are evaluated at their respective R_{500} (unless otherwise stated). M_{true} represents the unbiased cluster mass. The intrinsic scatters $\sigma_{\text{HE}|\text{true}}$ and $\sigma_{\text{WL}|\text{true}}$ are assumed to be uncorrelated. As described in [Sereno \(2016\)](#), for each set of parameters, the linear regression is performed calling the function *lira*, whose output are Markov chains produced with a Gibbs sampler. The median values of α , $\sigma_{\text{HE}|\text{true}}$, $\sigma_{\text{WL}|\text{true}}$ and their 1σ uncertainties are taken from the distribution of 80 000 Markov chains. For comparison, in Appendix A we provide also the bias estimated by computing the geometric mean and the median.

2.3. X-ray total masses

Under the assumption of spherical symmetry, one can solve the hydrostatic equilibrium equation and recover the total mass with only two ingredients: gas density and temperature profiles. For the ESZ sample, the gas density profiles were recovered from the best-fit parameters of a double β -model, while, for the temperature profiles, we used the 3D model by [Vikhlinin et al. \(2006\)](#) projected along the line of sight to fit the observed data points following the implementation presented in [Schellenberger & Reiprich \(2017a\)](#).

In L20, the X-ray total cluster masses have been obtained by assuming a Navarro-Frenk-White (NFW, [Navarro et al. 1996, 1997](#)) model for the mass profile with the relation from [Bhattacharya et al. \(2013\)](#) between the Dark Matter concentration and radius as a prior. However, we also estimated the masses by directly solving the equation of hydrostatic equilibrium without any prior on the form of the gravitational potential. Following the nomenclature by [Ettori et al. \(2013\)](#) (see also [Pratt et al. 2019](#)), we refer to these masses as backward (hereafter M_{HE}) and forward (hereafter M_{for}), respectively, each measured at their R_{500} .

There is a good correlation between M_{HE} and M_{for} (see Fig. 2), with a one-to-one relation and small intrinsic scatter with respect to the fitted relation (i.e., $4 \pm 1\%$). However, the M_{for} masses are on average $\sim 9\%$ lower than M_{HE} masses, almost independently of the dynamical state of the clusters. Instead, the mass difference slightly decreases as a function of the X-ray data quality. For instance, clusters for which the last temperature bin is at a radius $< 0.6R_{500}$, within $0.6-0.8R_{500}$, or $> 0.8R_{500}$ have an average offset of 13%, 9%, and 7%, respectively. Thus, the smaller the extrapolation required, the better the two estimates

³ LIRA (LInear Regression in Astronomy) is based on a Bayesian method (see [Sereno 2016](#) for more details) and is available from the R archive network at <https://cran.r-project.org/web/packages/lira/index.html>.

² <http://pico.oabo.inaf.it/~sereno/CoMaLit/LC2/>

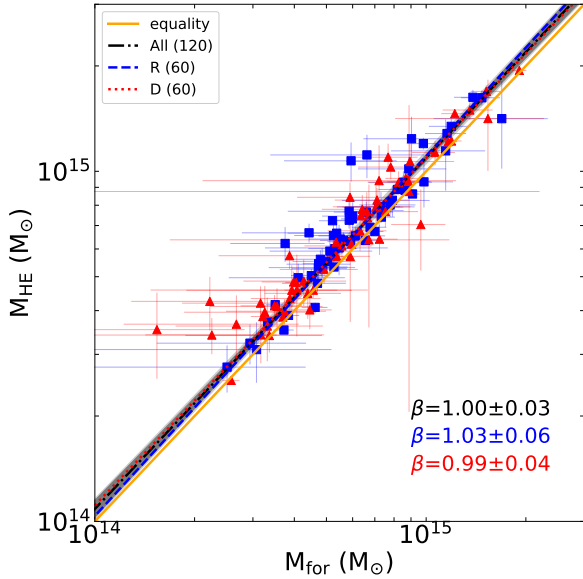


Fig. 2. Comparison between the total masses obtained with a forward approach (M_{for}) and a backward approach (M_{HE}). Relaxed (R) and disturbed (D) clusters are displayed as blue squares and red triangles, respectively. Numbers in the legend represent the sample size. The β values represent the fitted slope for disturbed (red), relaxed (blue), and all (black) clusters, respectively.

agree. Low redshift clusters (i.e., $z < 0.2$) show a marginally smaller offset (7% on average) than high redshift (i.e., $z > 0.2$) clusters (11%). This is mainly due to higher quality data of the low redshift clusters that allow measurements of the temperature to a larger fraction of R_{500} .

Previous studies showed that X-ray masses obtained with different analyses can be quite discordant. Indeed, this is a source of noise in the calculation of the level of hydrostatic bias. Since most of our clusters have been previously studied by other authors, here we provide a comparison of the mass estimates, evaluated at their respective R_{500} and in Appendix B we show the plot comparing the L20 total masses with those from literature.

There are 48 clusters that overlap with the sample used by the *Planck* Collaboration XI (2011) to investigate the relation between the integrated Compton parameter and several X-ray-derived properties. On average, our masses are $4 \pm 2\%$ lower than their masses estimated using the $M - Y_X$ relation⁴ by Arnaud et al. (2010). There is a mild, not statistically significant, dependence on the dynamical state: for the relaxed clusters, our mass estimates are in agreement (i.e., offset of $2 \pm 2\%$) with *Planck* Collaboration XI (2011), while for disturbed clusters the average offset is slightly larger (i.e., offset of $7 \pm 3\%$).

There are 18 clusters in common with Martino et al. (2014), who analyzed 50 clusters from the LoCuSS (Local Cluster Substructure Survey, e.g., Zhang et al. 2008) sample. Of these 18 clusters, Martino et al. (2014) provide 17 (14) total masses obtained with *Chandra* (*XMM-Newton*) data. The mass profiles in Martino et al. (2014) were derived with a forward approach using a powerlaw model for the temperature profile and the parametric form by Vikhlinin et al. (2006) for the gas density profile. On average, our masses are $7 \pm 10\%$ ($10 \pm 8\%$) lower than what they reported.

⁴ The $Y_X = M_{\text{gas}} \times kT$ parameter measures the total gas energy content and is the X-ray counterpart of the Sunyaev-Zel'dovich signal Y_{SZ} .

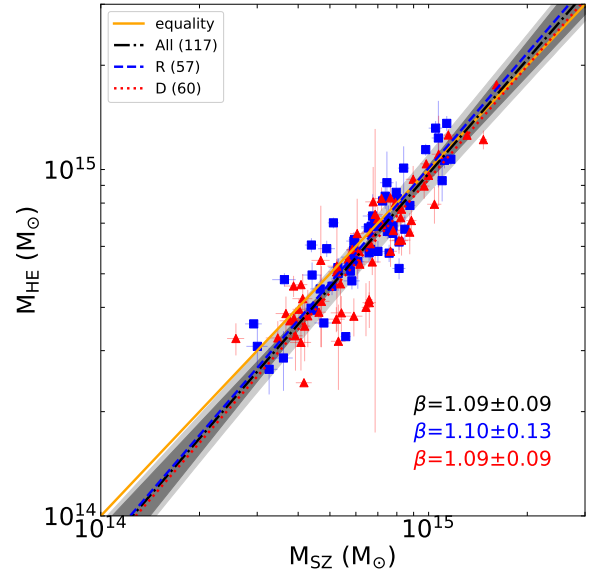


Fig. 3. $M_{\text{HE}} - M_{\text{SZ}}$ relation for the ESZ clusters investigated in L20. Relaxed (i.e., $M_{\text{par}} > 0$) and disturbed (i.e., $M_{\text{par}} < 0$) clusters are shown in blue squares and red triangles, respectively. The dark shaded area represent the 1σ statistical error for the full sample.

There are 12 clusters in common with Schellenberger & Reiprich (2017a) who analyzed the HIFLUGCS (HIGHEST X-ray FLUX Galaxy Cluster Sample, Reiprich & Böhringer 2002) sample with *Chandra* data. The mass profiles were derived with a forward approach using a double β -model for the density profile and several parametric forms which approximate a powerlaw in the outer regions, for the temperature profiles. Our total masses are, on average, $9 \pm 11\%$ lower.

There are 6 clusters in common with the CLASH (Cluster Lensing And Supernova survey with *Hubble*, Postman et al. 2012) sample. The masses were obtained with a backward approach assuming an NFW form for the mass profile and a triple β -model for the density profiles. Our total masses are on average $\sim 5\%$ lower than the masses estimated by Donahue et al. (2014) with *Chandra* data and the difference reduces to $\sim 2\%$ when compared to the masses obtained with *XMM-Newton* data.

Finally, there are 5 clusters in common with XCOP (XMM Cluster Outskirts Project, Eckert et al. 2017). The total masses derived in L20 are on average 11% lower than the masses estimated with the backward method and an NFW model as reported by Ettori et al. (2019). For two clusters (i.e., A3266 and ZwCl1215.1+0400) our total masses are $\sim 30\%$ lower than the XCOP estimate, but are in agreement with the *Planck* masses. Of the XCOP clusters, these two are the ones that deviate more significantly with respect to the *Planck* estimates.

2.4. $M_{\text{HE}} - M_{\text{SZ}}$

The thermal SZ effect directly measures the line-of-sight pressure, and thus, in the adiabatic scenario, the thermal energy content of the cluster gas, which is strongly correlated with the total mass. However, the SZ-derived masses largely rely on prior information on the calibration of the integrated SZ Comptonization (i.e., Y_{SZ}) with cluster masses from X-ray and/or WL analyses. The *Planck* masses have been determined using the X-ray hydrostatic masses to calibrate the $Y_X - Y_{\text{SZ}}$ relation. Therefore, the comparison M_{HE} and M_{SZ} indirectly provides the comparison between our ESZ masses (i.e., M_{HE}) and the hydrostatic masses used by *Planck*.

Table 1. Summary of the $1-b = M_{\text{HE}}/M_{\text{SZ}}$, estimated at R_{500} .

Sample	N	All				
		\tilde{M}_{SZ}	\bar{z}	$1-b$	σ_{SZ}	σ_{HE}
L20 All	117	6.59	0.19	0.93 ± 0.02	5 ± 3	16 ± 2
L20 R	57	6.68	0.20	0.96 ± 0.03	7 ± 4	16 ± 3
L20 D	60	6.51	0.19	0.90 ± 0.03	6 ± 4	15 ± 3
LC ² -single	60	7.56	0.22	0.93 ± 0.02	5 ± 3	17 ± 3
CCCP/MENeCS	25	7.65	0.18	0.94 ± 0.04	11 ± 6	14 ± 6
APEX-SZ	19	8.76	0.28	0.95 ± 0.04	4 ± 3	15 ± 3
LoCuSS	18	7.78	0.22	0.88 ± 0.04	6 ± 4	13 ± 4
WtG	17	8.27	0.29	0.92 ± 0.04	5 ± 3	13 ± 4

Notes. For each subsample we provide the number of clusters, the median SZ masses (i.e., \tilde{M}_{SZ}), and the median redshift (i.e., \bar{z}). The intrinsic scatters are given in percent.

In Fig. 3 we show the comparison between the X-ray (i.e., M_{HE}) and PSZ2 masses (i.e., M_{SZ}). The SZ masses are, on average, 7% higher than the X-ray masses (see Table 1). When we compute the X-ray masses at R_{500} determined by *Planck*, the difference decreases to $\sim 4\%$. When left free to vary, the slope of the $M_{\text{HE}} - M_{\text{SZ}}$ relation is found to be only slightly steeper than the equality line (i.e., $\beta = 1.09 \pm 0.09$) indicating a not significant mass dependence. There is no variation in slope when the sample is split based on the dynamical state of the clusters (see blue and red lines in Fig. 3). Even when selecting the most relaxed (top 30%) and most disturbed clusters in our sample, we do not find any significant dependence. However, there is a mild effect on the normalization with relaxed clusters more in agreement with the *Planck* masses than the disturbed clusters.

When splitting the sample in 3 redshift bins (i.e., $z < 0.15$, $0.15 < z < 0.25$, $z > 0.25$), we found the $M_{\text{HE}}/M_{\text{SZ}}$ ratio to be 0.96 ± 0.04 , 0.90 ± 0.03 , and 0.96 ± 0.04 , respectively. Thus, there are no indication of a dependence on redshift, as also confirmed by the Spearman rank test (i.e., $r = 0.06$ and $p = 0.53$). We also subdivided each redshift bin based on cluster mass but we did not find any trend.

We also investigated if any of the WL subsamples show a significantly different $M_{\text{HE}}/M_{\text{SZ}}$ ratio. The LoCuSS subsample shows the lowest value (i.e., $1-b = 0.88 \pm 0.04$) but it is still consistent within 1σ with the results from the other subsamples.

3. Results

Cluster masses can be directly estimated through X-ray observations, gravitational weak-lensing, and dynamical analysis of cluster galaxies, or alternatively using a mass proxy, such as the SZ flux which relies on prior information on the SZ-mass calibration obtained from X-ray or WL masses. Although each method requires some assumptions and is limited by different systematics, WL and caustic technique do not require the assumption of hydrostatic equilibrium, and numerical studies suggest that they can provide cluster mass estimates with little bias. In the following we compare our ESZ masses (i.e., M_{HE}) with the ones from other mass estimators.

3.1. X-ray versus weak-lensing masses

In this section, we compare X-ray and WL masses estimated within R_{500} which is one of the most common overdensities used for cosmological studies. In Fig. 4 we show the $M_{\text{HE}} - M_{\text{WL}}$ relation, while in Table 2, we provide the values $1-b$ computed for all the investigated subsamples.

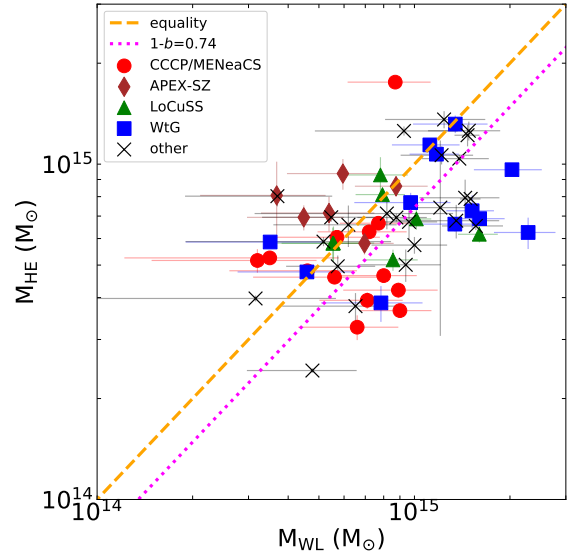


Fig. 4. Comparison between X-ray and WL masses from LC²-single. Red circles, brown diamonds, green triangles, and blue squares represent the masses obtained by Herbonnet et al. (2020), Klein et al. (2019), Okabe & Smith (2016), and Applegate et al. (2014), respectively. The magenta line represents the average $M_{\text{HE}}/M_{\text{WL}}$ ratio.

When the LC²-single catalog is used, we find that the average ratio between HE and WL masses is $1-b = 0.74 \pm 0.06$. This number is almost independent of the dynamical state: relaxed and disturbed clusters have mass ratios of $1-b = 0.75 \pm 0.08$ and $1-b = 0.73 \pm 0.12$, respectively. By splitting the sample in three redshift bins, we find no dependence of the mass ratio on redshift (i.e., $1-b = 0.72 \pm 0.09$ for $z < 0.17$, $1-b = 0.75 \pm 0.12$ for $0.17 < z < 0.29$, and $1-b = 0.76 \pm 0.12$ for $z > 0.29$). Instead, we find indication for a dependence on the total mass with smaller $1-b$ values for high mass clusters and a better agreement for low mass systems. In detail, for $M_{\text{WL}} \leq 6.5 \times 10^{14} M_{\odot}$ we found $1-b = 1.15 \pm 0.11$, for $6.5 \times 10^{14} M_{\odot} < M_{\text{WL}} \leq 10^{15} M_{\odot}$ we found $1-b = 0.73 \pm 0.11$, and for $M_{\text{WL}} > 10^{15} M_{\odot}$ we found $1-b = 0.62 \pm 0.08$. This is also confirmed by the Spearman test which gives a significant anti-correlation (i.e., $r = -0.67$, $p < 0.01$).

Also for WL there are significant differences between masses reported by distinct groups (e.g., see also Umetsu et al. 2014; Sereno & Ettori 2015). Here we provide a comparison between our masses, M_{HE} , and those estimated by different WL projects.

There are 27 clusters in common with the sample analyzed by Herbonnet et al. (2020, hereafter H20) who determined the masses for 48 clusters in the Multi Epoch Nearby Cluster Survey (MENeCS), and 52 clusters from the Canadian Cluster Comparison Project (CCCP). We find $M_{\text{HE}}/M_{\text{WL}} = 0.77 \pm 0.10$. Among the common clusters, there are 17 relaxed clusters (i.e., $M_{\text{par}} > 0$) for which we obtain $M_{\text{HE}}/M_{\text{WL}} = 0.83 \pm 0.12$ and 10 disturbed clusters (i.e., $M_{\text{par}} < 0$) for which we obtain $M_{\text{HE}}/M_{\text{WL}} = 0.73 \pm 0.21$. Given the large uncertainties, the tendency of a better agreement for relaxed clusters is only marginal. The correlation between bias and dynamical state measured by M_{par} is also not supported by the Spearman test (i.e., $r = -0.05$, $p = 0.81$).

There are 19 clusters in common with Klein et al. (2019, hereafter K19) who presented the WL analysis for clusters from the APEX-SZ survey (see also Nagarajan et al. 2019). Our HE masses are in agreement (i.e., $M_{\text{HE}}/M_{\text{WL}} = 1.02 \pm 0.12$) with the WL masses by K19.

Table 2. Summary of the $1-b = M_{\text{HE}}/M_{\text{WL}}$ values, estimated at R_{500} .

All						
Sample	N	\tilde{M}_{WL}	\tilde{z}	$1-b$	σ_{WL}	σ_{HE}
LC ² -single	62	8.37	0.21	0.74 ± 0.06	27 ± 8	30 ± 6
CCCP/MENeaCS	27	7.40	0.18	0.77 ± 0.10	10 ± 9	40 ± 8
APEX-SZ	19	8.51	0.28	1.02 ± 0.12	37 ± 10	28 ± 10
LoCuSS	18	8.24	0.22	0.76 ± 0.09	23 ± 9	24 ± 8
WtG	17	11.73	0.29	0.61 ± 0.12	31 ± 15	26 ± 11
Relaxed						
Sample	N	\tilde{M}_{WL}	\tilde{z}	$1-b$	σ_{WL}	σ_{HE}
LC ² -single	34	8.65	0.21	0.75 ± 0.08	27 ± 9	18 ± 8
CCCP/MENeaCS	17	7.10	0.17	0.83 ± 0.12	21 ± 13	25 ± 10
APEX-SZ	11	8.75	0.27	0.90 ± 0.17	42 ± 14	27 ± 13
LoCuSS	14	9.42	0.22	0.74 ± 0.10	20 ± 10	26 ± 8
WtG	10	11.46	0.23	0.71 ± 0.17	22 ± 18	32 ± 14
Disturbed						
Sample	N	\tilde{M}_{WL}	\tilde{z}	$1-b$	σ_{WL}	σ_{HE}
LC ² -single	28	8.09	0.29	0.73 ± 0.12	31 ± 14	45 ± 10
CCCP/MENeaCS	10	8.35	0.19	0.73 ± 0.21	10 ± 13	61 ± 17
APEX-SZ	8	7.57	0.30	1.19 ± 0.15	14 ± 15	15 ± 14
LoCuSS	4	5.89	0.24	0.93 ± 0.29	9 ± 20	22 ± 29
WtG	7	15.20	0.41	0.48 ± 0.18	16 ± 17	18 ± 16

Notes. The data for the CCCP/MENeaCS, APEX-SZ, LoCuSS, and WtG samples have been taken from H20, K19, O16, and A14, respectively. For each subsample we provide the number of clusters, the median WL mass (i.e., \tilde{M}_{WL}), and the median redshift (i.e., \tilde{z}). Masses are in units of $10^{14} M_{\odot}$. The intrinsic scatters are given in percent.

There are 18 clusters in common with Okabe & Smith (2016, hereafter O16), who analyzed the LoCuSS clusters. The average mass ratio is $1-b = 0.76 \pm 0.09$. It is worth mentioning that LoCuSS shows the larger deviation in $M_{\text{HE}}/M_{\text{SZ}}$ (i.e., $12\% \pm 4$ offset, see Sect. 2.4).

Finally, there are 17 clusters in common with Weighing the Giants (WtG), the sample analyzed by Applegate et al. (2014, hereafter A14) who reported WL masses for 51 massive clusters. We find $M_{\text{HE}}/M_{\text{WL}} = 0.61 \pm 0.12$, a much larger discrepancy than the one observed with respect to the other samples. There is a clear correlation between the dynamical state (i.e., M_{par}) and $M_{\text{HE}}/M_{\text{WL}}$ (i.e., Spearman $r = 0.52$, $p = 0.03$). In fact, when splitting the sample in relaxed and disturbed clusters, we find $M_{\text{HE}}/M_{\text{WL}} = 0.71 \pm 0.17$ and $M_{\text{HE}}/M_{\text{WL}} = 0.48 \pm 0.18$, respectively.

The intrinsic scatter σ_{WL} of the different samples varies from $\sim 10\%$ for CCCP/MENeaCS to $\sim 35\%$ for APEX-SZ clusters. Also for WtG the scatter is $>30\%$ but is probably boosted by the clear dependence of the mass ratio on the dynamical state. In fact, both subsamples of relaxed and disturbed clusters show a smaller scatter (i.e., 15–20%), although with large uncertainties. The scatters σ_{WL} and σ_{HE} are often at the same level. An exception is CCCP/MENeaCS which show a significantly larger σ_{HE} for the disturbed systems.

3.2. X-ray versus caustic masses

An alternative technique to estimate cluster masses is through the caustic method, that makes use of galaxy dynamics. The caustic, a trumpet-shaped region in the diagram of line-of-sight velocity versus projected distance, defines which galaxies lie inside and outside a cluster assuming that a system cannot have a velocity larger than the escape velocity. Caustic-derived masses

Table 3. Summary of the mass ratio between X-ray and dynamical cluster masses.

All						
Sample	N	\tilde{M}_{dyn}	\tilde{z}	$1-b$	σ_{dyn}	σ_{HE}
HeCS-omnibus	25	7.10	0.16	1.30 ± 0.14	76 ± 11	8 ± 8
HeCS-omnibus ($N > 135$)	12	9.31	0.18	1.03 ± 0.10	26 ± 11	13 ± 10
HeCS-omnibus ($N < 135$)	13	5.80	0.14	1.72 ± 0.27	103 ± 20	18 ± 12
HeCS-omnibus ($N > 100$)	18	7.36	0.18	1.04 ± 0.11	26 ± 12	14 ± 11
HeCS-omnibus ($N < 100$)	7	2.74	0.14	2.34 ± 1.48	121 ± 41	19 ± 15
HIFLUGCS	12	4.89	0.08	0.95 ± 0.17	49 ± 18	14 ± 14
Relaxed						
Sample	N	\tilde{M}_{dyn}	\tilde{z}	$1-b$	σ_{dyn}	σ_{HE}
HeCS-omnibus	14	7.46	0.19	1.37 ± 0.21	99 ± 19	8 ± 9
HIFLUGCS	7	3.77	0.09	1.28 ± 0.23	37 ± 23	23 ± 15
Disturbed						
Sample	N	\tilde{M}_{dyn}	\tilde{z}	$1-b$	σ_{dyn}	σ_{HE}
HeCS-omnibus	11	7.10	0.11	1.17 ± 0.16	41 ± 15	25 ± 14
HIFLUGCS	5	9.15	0.07	0.69 ± 0.23	12 ± 19	22 ± 25

Notes. The comparison is done at R_{200} for the HeCS-omnibus caustic masses by Sohn et al. (2020) and at R_{500} for the HIFLUGCS masses estimate from the velocity dispersion by Zhang et al. (2017). Masses are in units of $10^{14} M_{\odot}$. The intrinsic scatters are given in percent.

(i.e., M_{c}) are insensitive to the physical processes that might cause hydrostatic biases and are subject to different systematic uncertainties than lensing masses.

One of the advantages of the caustic approach is that it does not rely on the hypothesis of dynamical equilibrium. The drawback is that a reliable measurement of the underlying dark matter distribution requires a large number of member galaxies (e.g., Serra & Diaferio 2013).

There are 25 clusters in common with Sohn et al. (2020) who obtained the caustic masses at R_{200} for the HeCS-omnibus cluster sample (see also Maughan et al. 2016). For a comparison with their masses, we computed the M_{200} masses for the L20 sample. We found that the caustic masses are, on average, smaller than the X-ray-derived masses with no indication for a relation with the dynamical state of the clusters (see Table 3 and Fig. 5). However, we note that the clusters showing the largest difference between X-ray and caustic masses tend to have a small N_{200} , which is the number of spectroscopic members within R_{200} . For instance, when we split the sample based on N_{200} , we find $M_{\text{HE}}/M_{\text{c}} = 1.04 \pm 0.11$ for clusters with $N_{200} > 135$ and $M_{\text{HE}}/M_{\text{c}} = 1.72 \pm 0.27$ for clusters with $N_{200} < 135$. This is also supported by the good anti-correlation between $M_{\text{HE}}/M_{\text{c}}$ and N_{200} (i.e., Spearman $r = -0.54$, $p < 0.01$). Also, the scatter significantly decreases when clusters with few member galaxies are removed, and becomes comparable with the one observed for WL samples. The low bias and relatively small scatter σ_{dyn} is found also with different cuts on N_{200} and start to show significant deviations when clusters with $N_{200} < 100$ are included in the fit.

3.3. X-ray versus virial masses

As an alternative to the caustic method, one can derive dynamical masses using the virial theorem (i.e., M_{v} hereafter). This method is based on the hypothesis of dynamical equilibrium, spherical symmetry, and on the additional assumptions that all the galaxies have the same mass and their spatial and velocity

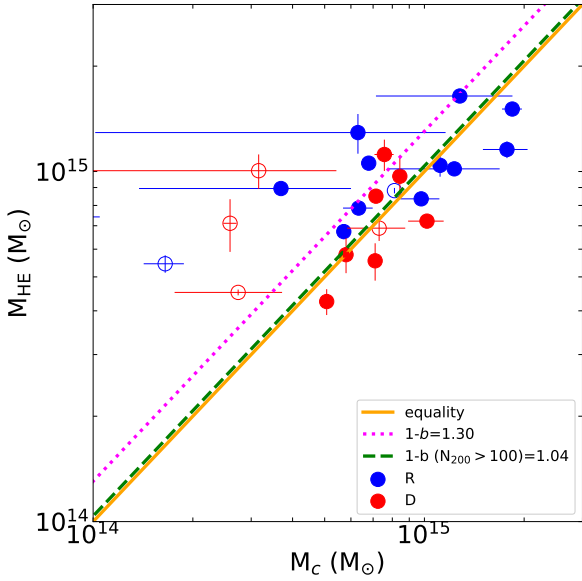


Fig. 5. Comparison between X-ray and caustic masses for the 25 clusters in common with [Sohn et al. \(2020\)](#). Relaxed (*R*) and disturbed (*D*) clusters are plotted with blue and red circles, respectively. Filled and empty circles represent the clusters with $N_{200} > 100$ and $N_{200} < 100$, respectively. The dotted magenta line represents the average M_{HE}/M_c ratio. The dashed green line represents the average M_{HE}/M_c ratio obtained after removing the clusters with less than 100 galaxies within R_{200} .

distribution follow those of dark matter particles. Despite these approximations, the virial theorem has proven to be a good estimator even for a small number of cluster galaxies (e.g., see [Biviano et al. 2006](#) and [Munari et al. 2013](#)).

[Zhang et al. \(2017\)](#) obtained the dynamical masses for the HIFLUGCS sample based on the velocity dispersion of the member galaxies. For the 12 clusters in common with the sample in L20, we find reasonably good agreement with a mass ratio consistent with zero (i.e., $1-b = 0.95 \pm 0.18$). However, relaxed and disturbed objects clearly populate different regions of the $M_{\text{HE}} - M_v$ plane (see Fig. 6) although larger samples are clearly needed to obtain firmer conclusions. When the HE masses are computed at R_{500} provided by [Zhang et al. \(2017\)](#), we obtain a mass ratio of 0.97 ± 0.10 and a significant reduction of the scatter (i.e., $\sigma_{\text{dyn}} = 18 \pm 12\%$ and $\sigma_{\text{HE}} = 8 \pm 7\%$).

4. Discussion

4.1. X-ray hydrostatic masses

[Roza et al. \(2014\)](#) and [Sereno & Ettori \(2015\)](#) reported that the X-ray properties estimated by competing groups can show large discrepancies. They found that total masses can differ by a factor of two even for relaxed clusters. The major sources of the discrepancies have been ascribed to differences in the data analysis, to cross-calibration problems between instruments, and to different techniques to recover the mass.

As seen in Sect. 2.3 the total masses obtained in L20 are systematically lower than the masses published by other authors (at least for the ones discussed in this paper). However, while the discrepancy can be up to 30% for individual clusters, the average difference in sample pairs is only a few percent with respect to the values by [Donahue et al. \(2014\)](#) and [Planck Collaboration XI \(2011\)](#) and $\sim 10\%$ with respect to [Ettori et al. \(2019\)](#), [Martino et al. \(2014\)](#), and [Schellenberger & Reiprich \(2017a\)](#).

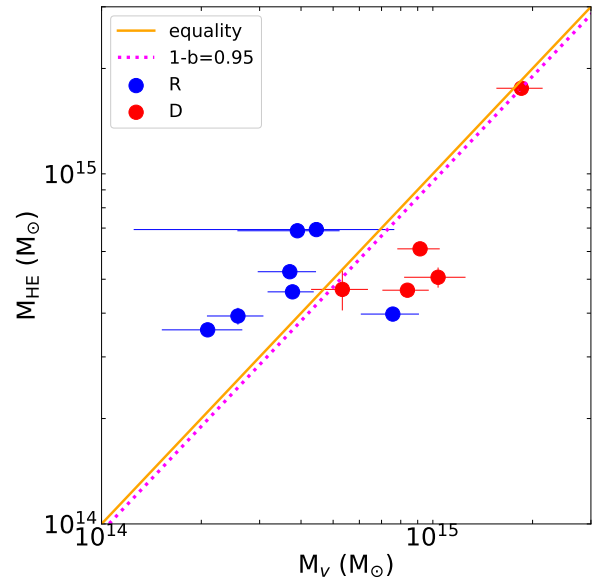


Fig. 6. Comparison between X-ray and virial masses for the 12 HIFLUGCS clusters in common with [Zhang et al. \(2017\)](#). Relaxed (*R*) and disturbed (*D*) clusters are plotted with blue and red circles, respectively. The dotted magenta line represents the average M_{HE}/M_v ratio.

In Sect. 2.3 we showed that part of these differences is indeed associated with the mass reconstruction methodology. Traditionally, as discussed in [Ettori et al. \(2013\)](#) and [Pratt et al. \(2019\)](#), two main approaches have been used to determine the mass profile. In one case, a parametric mass model (e.g., NFW) is assumed, while in the second, two functional forms are fitted to the density and temperature profiles and propagated through the HE equation to derive the mass profile. In this paper, we refer to the masses estimated with the first and second methods as M_{HE} and M_{for} , respectively. Our results show that the M_{for} masses are, on average, 9% smaller than the M_{HE} masses. One of the reasons is a poor radial sampling (i.e., few temperature measurements) of some clusters for which the model extrapolation can diverge (see, e.g., [Bartalucci et al. 2018](#) for a more extended discussion). In fact, when we compare the 25 clusters with at least ten temperature bins, the average difference between the M_{HE} and M_{for} masses is less than 1%. When the X-ray data allow only a few temperature measurements, the assumption of the NFW profile, which is a physically motivated model and is found to provide a good fit for massive clusters, can be used to obtain mass estimates without introducing strong systematics if the temperature profile is measured at least out to $0.5R_{500}$, as shown by [Schellenberger & Reiprich \(2017b\)](#). Our results are also in agreement with [Ettori et al. \(2019\)](#) and [Bartalucci et al. \(2018\)](#), who analyzed samples of 13 and 5 clusters with high X-ray quality data and found that different mass estimation methods yield reasonably consistent results.

Another cause for the lower masses in L20 is the use of the total (i.e., neutral and molecular) column density, which has a significant impact on the measured temperatures in the regions with high molecular contributions. The inclusion of the molecular component allows a better fit to the X-ray data (see [Lovisari & Reiprich 2019](#) for more details) and, for very high total column densities (i.e., $N_{\text{H}} > 10^{21} \text{ cm}^{-2}$), can lead to estimated temperatures (and therefore total masses) that are 10–15% lower than when only the neutral component is used. However, for most clusters the effect is limited to a few percent.

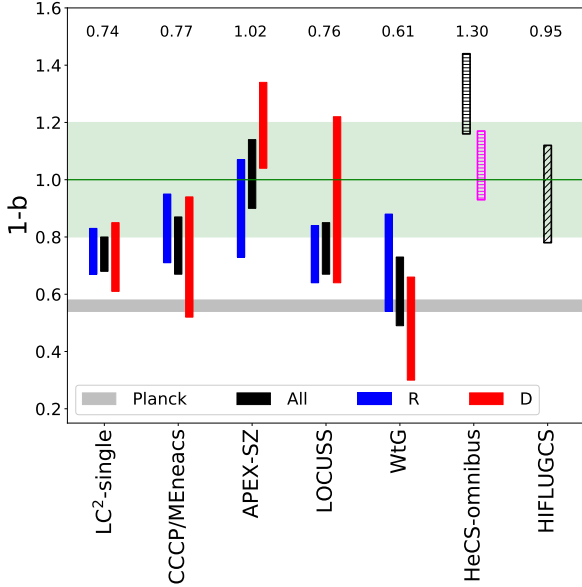


Fig. 7. Hydrostatic mass bias for each of the subsamples investigated in this paper. Filled color bars represent the mass bias estimated using WL masses, while horizontal and diagonal hatches show the estimates using masses from caustics and velocity dispersion, respectively. In magenta we show the bias obtained with caustics after removing the clusters with few member galaxies. The bar lengths enclose the 1σ estimates and the central values are also reported on the top of the panel. The gray line represents the bias necessary to reconcile *Planck* number counts and CMB. The green shaded region indicates the area defined as $\pm 20\%$ centered on $1-b = 0$.

Table 4. Summary of the $M_{\text{HE}}/M_{\text{WL}}$ ratio estimated at $R = 1$ Mpc.

Sample	N	All		$1-b$	σ_{WL}	σ_{HE}
		\bar{M}_{WL}	\bar{z}			
LC ² -single All	62	8.37	0.21	0.83 ± 0.04	21 ± 4	22 ± 5
LC ² -single R	34	8.65	0.21	0.84 ± 0.05	20 ± 6	14 ± 6
LC ² -single D	28	8.09	0.29	0.83 ± 0.08	23 ± 10	30 ± 7

Notes. Masses are in units of $10^{14} M_{\odot}$. The intrinsic scatters are given in percent.

4.2. $M_{\text{HE}} - M_{\text{SZ}}$ and $M_{\text{SZ}} - M_{\text{WL}}$

The PSZ2 masses were estimated from the $M - Y_{\text{SZ}}$ relation calibrated using the $M - Y_{\text{X}}$ relation by Arnaud et al. (2010), which links the hydrostatic mass to the Y_{X} parameter, the X-ray equivalent of the integrated Compton parameter Y_{SZ} . Considering uncertainties and scatter, the $M - Y_{\text{X}}$ relation for the ESZ sample obtained in L20 is in good agreement with the one by Arnaud et al. (2010), in particular in the low mass regime. However, in the high mass regime, the difference is about 5–6%. Therefore, the difference of 4% that we observe at the same radius (i.e., R_{500} from *Planck*) can be easily explained by the small offset in the $M - Y_{\text{X}}$ relation.

We found a mild mass-dependent trend between X-ray and *Planck* SZ masses as previously obtained by other studies. Schellenberger & Reiprich (2017a) suggest that this effect is caused by the *Planck* selection function and the intrinsic scatter of the hydrostatic mass estimates. However, since the *Planck* masses are obtained using the $M - Y_{\text{X}}$ relation, it is also possible that there is a mass dependent effect introduced by the calibration of the scaling relations.

The agreement between the X-ray and SZ masses is better for relaxed clusters ($4 \pm 3\%$) than disturbed clusters ($11 \pm 3\%$). Since the relation used to estimate the *Planck* masses is the same, and there is no significant difference in the $M - Y_{\text{X}}$ for relaxed and disturbed systems (e.g., see L20), the larger $M_{\text{HE}}/M_{\text{SZ}}$ ratio for disturbed systems may indicate that the SZ signal is somehow impacted (i.e., overestimated). For instance, due to the large *Planck* beam, the integrated SZ signal may include unidentified substructures along the line of sight, which instead could be masked in the X-ray analysis. Projection effects are also a source of intrinsic scatter (e.g., Krause et al. 2012), and indeed we note that the scatter in the low mass regime is higher than the one for massive clusters (10% versus 4%).

Sereno et al. (2017), for a subsample of 35 PSZ2 clusters, found that the *Planck* masses are biased low with respect to WL masses by $27 \pm 11(\text{stat}) \pm 8(\text{sys})\%$. Instead, the bias obtained with the heterogeneous LC² catalog is $\sim 13\%$ (see Sereno et al. 2015). For the 62 ESZ clusters in common with LC², we estimate a mass ratio of $1-b = 0.80 \pm 0.06$ (corresponding to a bias of $b = 20 \pm 6\%$), essentially in agreement with previous findings. Due to the large uncertainties, it is hard to say if there is any trend with the dynamical state, although, based on the average values, disturbed clusters show a slightly better agreement between the SZ and WL masses (i.e., $1-b = 0.78$ and $1-b = 0.83$ for disturbed and relaxed clusters, respectively). This is in agreement with the above-mentioned suggestion of an overestimated SZ signal for disturbed clusters. However, another possibility is that the WL masses are on average slightly underestimated.

4.3. M_{HE} and M_{WL} comparison

We compared our estimates of the hydrostatic mass with constraints obtained with WL analysis from different groups (see Sect. 3.1). The mass ratio obtained with the LC² compilation is $1-b = 0.74 \pm 0.06$ (see Table 2 and Fig. 7). This is consistent with predictions from hydrodynamical simulations (e.g., Rasia et al. 2012; Biffi et al. 2016). Moreover, as discussed by Sereno & Ettori (2015), given the relation between mass and overdensity-radius, mass differences are inflated when computed at R_{500} . Therefore, we also compare the masses obtained within a physical aperture of 1 Mpc (see Table 4), keeping in mind that this aperture refers to a physically different region in clusters with different mass. The average difference decreases by $\sim 9\%$ with also a substantial reduction of the intrinsic scatter.

The large differences between the total masses in L20 and the LC² sample is partially driven by the larger mass ratio (see Table 2 and Fig. 7) observed with WtG (masses from A14) and CLASH (masses from Umetsu et al. 2016) programs, whose results agree. In fact, of the 62 matched clusters, 18 of the matched clusters are included either in WtG or CLASH.

The mass ratio observed, when considering the WL masses from A14, is $1-b = 0.61 \pm 0.12$, which corresponds to a bias of $39 \pm 12\%$. This is consistent with the value required to reconcile *Planck* cluster number counts and *Planck* primary CMB measurements. Our result is in agreement with the finding by von der Linden et al. (2014) who found an overall mass ratio of $M_{\text{Planck}}/M_{\text{WtG}} = 0.69 \pm 0.07$. Also, although the uncertainties are large, this is the only sample for which the hydrostatic bias for disturbed systems is significantly higher than the bias obtained for relaxed systems, which is what would be expected if the WL masses are indeed unbiased and the hydrostatic bias is larger for dynamical active clusters.

The other samples show a higher mass ratio (corresponding to a lower hydrostatic bias, i.e., 20–30%) and a less clear

dependence on the dynamical state. The results from K19 show an anti-correlation even though not statistically significant. Since, relaxed clusters are expected to have a smaller mass bias than disturbed objects (e.g., Biffi et al. 2016), the lack of correlation of the mass ratios with the dynamical state may suggest that some systematics in the WL analyses are still not accounted for. For instance, it is worth noting that WL masses can be significantly underestimated because of massive sub-clumps (Meneghetti et al. 2010) or uncorrelated large-scale matter projections along the line of sight (Becker & Kravtsov 2011). For instance, the lensing masses are typically overestimated when the cluster is elongated along the line of sight, and underestimated when it is elongated on the plane of the sky. If we focus on the more relaxed clusters, we find that the mass ratios, derived from the different subsamples, tend to agree better, and correspond to biases on the order of 5–20%. Instead, when disturbed clusters are considered, the mass ratios vary a lot from sample to sample. One of the reasons is that each of the lensing subsamples contains only a small fraction of disturbed clusters (e.g., only four in O16). This makes possible for the lensing bias of individual systems to not be fully averaged out. We also stress that, although the subdivision between relaxed and disturbed clusters using M_{par} , is only relative, SZ surveys are thought to be very close to being mass selected, and therefore the M_{par} distribution is probably indicative of the cluster population in the Universe.

We investigated any dependence of the mass ratio on cluster mass and redshift (see Sect. 3.1). We do not find any redshift dependence of the mass ratio but our clusters span a relatively small range of redshifts and we cannot place conclusive constraints. Instead, we find a moderate mass dependence, with the most massive clusters showing a larger bias in agreement with the finding by von der Linden et al. (2014) and Eckert et al. (2019), while Sereno et al. (2015) and Sereno & Ettori (2017) found the bias to be redshift rather than mass dependent. Naively, one would expect the more massive objects to be on average more disturbed than the least massive clusters, and therefore with a large bias. However, in L17, we did not find any correlation between the different morphological parameters (i.e., a measure of the dynamical state) and the total mass (see also Böhringer et al. 2010). This is confirmed by the Spearman rank test for the full L20 sample ($r = 0.02$, $p = 0.80$) and for the subsample of 62 clusters with WL masses ($r = -0.09$, $p = 0.46$).

4.4. M_{HE} and M_c comparison

In addition to the widely used WL technique to constrain the hydrostatic mass bias, we also compared the X-ray masses with caustic masses. The comparison with caustic masses determined at R_{200} by Sohn et al. (2020) required a significant extrapolation of the X-ray data, so our finding should be treated with caution. We find that the caustic masses are significantly lower than the X-ray masses in agreement with the finding by Ettori et al. (2019), Maughan et al. (2016), and Sereno et al. (2015). However, Serra & Diaferio (2013) found that only with a large number of spectroscopic members can the velocity field be properly sampled and return a correct mass. In particular, they showed that for $N = 50$ the mass estimate is biased low by $\sim 20\%$, while for $N = 200$ the bias is consistent with zero. After removing the clusters with fewer member galaxies, the agreement between X-ray and caustic measurements is good. Moreover, there is no indication of any dependence of the mass ratio on the dynamical state of the clusters, supporting the finding by Maughan et al. (2016). This, together with the overall trend observed with the comparison to WL masses, suggests that there are still unknown

systematics that introduce significant scatter in the relations that prevent a complete and proper estimate of the hydrostatic mass.

Previous comparisons of caustic masses with X-ray masses by Maughan et al. (2016) and with *Planck* masses (calibrated with hydrostatic X-ray masses) by Rines et al. (2016) also show little or no bias. Instead, Ettori et al. (2019) found that the caustic masses of 6 X-COP clusters, that are nearby massive objects for which the determinations of caustics might be more problematic, are significantly underestimated with respect to the hydrostatic masses. However, given the large scatter observed in the $M_{\text{HE}} - M_c$ relation, a comparison with a much larger sample is required to obtain more conclusive results.

4.5. M_{HE} and M_v comparison

The velocity dispersion of member galaxies provides another way to estimate cluster masses under the hydrostatic equilibrium assumption. We find agreement when the L20 X-ray masses are compared to the masses estimated based on the velocity dispersion. In fact, for the 12 clusters in common with Zhang et al. (2017) we find $1-b = 0.95 \pm 0.17$ in agreement with the finding by Ettori et al. (2019). On the contrary, Amodeo et al. (2017) reported a bias of $1-b = 0.64 \pm 0.11$ for a sample of 17 *Planck* clusters. However, it is still debated whether the velocity dispersion of the galaxies is a good tracer for the velocity dispersion of the dark matter particles. In fact, differences like the so-called velocity bias, have been reported to be at the $\pm 10\%$ level (e.g., Munari et al. 2013; Armitage et al. 2018), and could partially explain the different hydrostatic bias derived in different studies. Although, there is no general agreement on the magnitude or even the sign of the velocity bias, Rines et al. (2016), using a sample of 123 clusters, found that the *Planck* masses are not significantly biased compared to dynamical mass estimates unless, there is a significant velocity bias. We also note that our results show a significant difference between the mass ratio determined for relaxed and disturbed clusters. Although the number of objects in each subsample is small, this systematic difference suggests that the hydrostatic equilibrium assumption does affect quite differently the M_{HE} and M_v masses, and therefore that some systematics are not accounted for in at least one of the two methods.

The dynamical masses have large scatter (see Table 3) which prevents robust determination with small samples. However, the scatter for the caustic is significantly reduced when clusters with a limited number of member galaxies are excluded from the sample.

5. Summary and conclusion

We have compared the masses obtained with X-ray, WL, and dynamical analyses of a *Planck*-selected sample of clusters. Our main findings are the following:

- X-ray hydrostatic mass estimates with a forward approach (M_{for}) and a backward approach (M_{HE}) yield consistent results, as long as only the clusters with sufficiently large number of temperature measurements in the radial range of interest are considered. Otherwise, M_{for} tend to be underestimated with respect to M_{HE} .
- The X-ray masses reported by competing groups and investigated in this paper show discrepancies lower than 10% on average, although tensions up to 30% can be still present in single objects, as consequence of systematic differences (mainly in the spectroscopic analysis (see also Sereno & Ettori 2015)).
- The X-ray masses are smaller than the *Planck* SZ masses, which are derived iteratively from scaling relations

calibrated using masses not corrected for the hydrostatic bias. However, the difference most probably arises from a small offset between the $M - Y_X$ relation determined for the ESZ sample and the one used to calibrate the *Planck* $M - Y_{SZ}$ relation. Interestingly, we find that the difference is higher for the disturbed systems which could be an indication that the SZ signal is overestimated for these systems.

- Based on the compilation of WL masses from different works, the estimated M_{HE}/M_{WL} mass ratio is $1-b \sim 0.7-0.8$. This corresponds to an hydrostatic bias of 20–30%, which is smaller than that required to resolve the discrepancy between *Planck* cluster number counts and primary CMB. However, the bias derived with different subsamples gives significantly different results, for example, the comparison with 17 WtG clusters yields a mass bias of $\sim 40\%$ (which would substantially reduce the tension) while, for example, the comparison with 19 APEX-SZ clusters points to a zero bias. Therefore, because of the intrinsic uncertainties in the WL calibrations we cannot provide a final conclusion.
- Unlike the WL masses, the dynamical masses, either from caustic or velocity dispersion, favor a scenario where X-ray hydrostatic masses have little or no bias requiring a different explanation to solve the tension between SZ cluster counts and primary CMB. However, the dynamical masses show a large scatter and are affected by some systematics (e.g., the velocity bias) which are source of uncertainty in our final result.
- We do not observe a significant dependence of the mass ratio $1-b$ with the dynamical state of the clusters. The only subsamples showing a clear trend are WtG (WL masses) and HIFLUGCS (dynamical masses), while for the others the effect is small or shows an opposite behavior. This suggests that there are systematics not accounted for in the different analyses. However, since most of the subsamples of dynamically disturbed objects have fewer than 10 clusters, the dependence on the dynamical state needs to be confirmed with larger samples.

Acknowledgements. The authors thank the anonymous referee for useful comments and suggestions that helped improve and clarify the presentation of this work. L.L., S.E., and M.S. acknowledge financial contribution from the contracts ASI-INAF Athena 2015-046-R.0, ASI-INAF Athena 2019-27-HH.0, “Attività di Studio per la comunità scientifica di Astrofisica delle Alte Energie e Fisica Astroparticellare” (Accordo Attuativo ASI-INAF n. 2017-14-H.0), and from INAF “Call per interventi aggiuntivi a sostegno della ricerca di main stream di INAF”. G.S. acknowledges support from *Chandra* grant GO5-16126X. W.R.F. and C.J. acknowledge support from the Smithsonian Institution and the *Chandra* High Resolution Camera Project through NASA contract NAS8-03060. F.A.-S. acknowledges support from *Chandra* grant GO3-14131X.

References

- Allen, S. W., Evrard, A. E., & Mantz, A. B. 2011, *ARA&A*, 49, 409
- Amodeo, S., Mei, S., Stanford, S. A., et al. 2017, *ApJ*, 844, 101
- Ansarifard, S., Rasia, E., Biffi, V., et al. 2020, *A&A*, 634, A113
- Applegate, D. E., von der Linden, A., Kelly, P. L., et al. 2014, *MNRAS*, 439, 48
- Applegate, D. E., Mantz, A., Allen, S. W., et al. 2016, *MNRAS*, 457, 1522
- Armitage, T. J., Barnes, D. J., Kay, S. T., et al. 2018, *MNRAS*, 474, 3746
- Arnaud, M., Pratt, G. W., Piffaretti, R., et al. 2010, *A&A*, 517, A92
- Barnes, D. J., Vogelsberger, M., Pearce, F. A., et al. 2020, *MNRAS*, in press, [arXiv:2001.11508]
- Bartalucci, I., Arnaud, M., Pratt, G. W., & Le Brun, A. M. C. 2018, *A&A*, 617, A64
- Battaglia, N., Bond, J. R., Pfrommer, C., & Sievers, J. L. 2012, *ApJ*, 758, 74
- Becker, M. R., & Kravtsov, A. V. 2011, *ApJ*, 740, 25
- Bhattacharya, S., Habib, S., Heitmann, K., & Vikhlinin, A. 2013, *ApJ*, 766, 32
- Biffi, V., Borgani, S., Murante, G., et al. 2016, *ApJ*, 827, 112
- Biviano, A., Murante, G., Borgani, S., et al. 2006, *A&A*, 456, 23
- Biviano, A., Rosati, P., Balestra, I., et al. 2013, *A&A*, 558, A1
- Bocquet, S., Dietrich, J. P., Schrabback, T., et al. 2019, *ApJ*, 878, 55
- Böhringer, H., Pratt, G. W., Arnaud, M., et al. 2010, *A&A*, 514, A32
- Diaferio, A. 1999, *MNRAS*, 309, 610
- Diaferio, A., & Geller, M. J. 1997, *ApJ*, 481, 633
- Donahue, M., Voit, G. M., Mahdavi, A., et al. 2014, *ApJ*, 794, 136
- Donahue, M., Ettori, S., Rasia, E., et al. 2016, *ApJ*, 819, 36
- Eckert, D., Ettori, S., Pointecouteau, E., et al. 2017, *Astron. Nachr.*, 338, 293
- Eckert, D., Ghirardini, V., Ettori, S., et al. 2019, *A&A*, 621, A40
- Ettori, S., Donnarumma, A., Pointecouteau, E., et al. 2013, *Space Sci. Rev.*, 177, 119
- Ettori, S., Ghirardini, V., Eckert, D., et al. 2019, *A&A*, 621, A39
- Foëx, G., Böhringer, H., & Chon, G. 2017, *A&A*, 606, A122
- Gruen, D., Seitz, S., Brimiouille, F., et al. 2014, *MNRAS*, 442, 1507
- Herbonnet, R., Sifón, C., Hoekstra, H., et al. 2020, *MNRAS*, 497, 4684
- Hilton, M., Hasselfield, M., Sifón, C., et al. 2018, *ApJS*, 235, 20
- Hoekstra, H., Herbonnet, R., Muzzin, A., et al. 2015, *MNRAS*, 449, 685
- Israel, H., Reiprich, T. H., Erben, T., et al. 2014, *A&A*, 564, A129
- Klein, M., Israel, H., Nagarajan, A., et al. 2019, *MNRAS*, 488, 1704
- Krause, E., Pierpaoli, E., Dolag, K., & Borgani, S. 2012, *MNRAS*, 419, 1766
- Kravtsov, A. V., & Borgani, S. 2012, *ARA&A*, 50, 353
- Lau, E. T., Kravtsov, A. V., & Nagai, D. 2009, *ApJ*, 705, 1129
- Lovisari, L., & Reiprich, T. H. 2019, *MNRAS*, 483, 540
- Lovisari, L., Forman, W. R., Jones, C., et al. 2017, *ApJ*, 846, 51
- Lovisari, L., Schellenberger, G., Sereno, M., et al. 2020, *ApJ*, 892, 102
- Mahdavi, A., Hoekstra, H., Babul, A., et al. 2013, *ApJ*, 767, 116
- Martino, R., Mazzotta, P., Bourdin, H., et al. 2014, *MNRAS*, 443, 2342
- Maughan, B. J., Giles, P. A., Rines, K. J., et al. 2016, *MNRAS*, 461, 4182
- Mazzotta, P., Rasia, E., Moscardini, L., & Tormen, G. 2004, *MNRAS*, 354, 10
- Medezinski, E., Battaglia, N., Umetsu, K., et al. 2018, *PASJ*, 70, S28
- Meneghetti, M., Rasia, E., Merten, J., et al. 2010, *A&A*, 514, A93
- Miyatake, H., Battaglia, N., Hilton, M., et al. 2019, *ApJ*, 875, 63
- Motl, P. M., Hallman, E. J., Burns, J. O., & Norman, M. L. 2005, *ApJ*, 623, L63
- Munari, E., Biviano, A., Borgani, S., Murante, G., & Fabjan, D. 2013, *MNRAS*, 430, 2638
- Nagarajan, A., Pacaud, F., Sommer, M., et al. 2019, *MNRAS*, 488, 1728
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, *ApJ*, 462, 563
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, *ApJ*, 490, 493
- Oguri, M., & Hamana, T. 2011, *MNRAS*, 414, 1851
- Okabe, N., & Smith, G. P. 2016, *MNRAS*, 461, 3794
- Penna-Lima, M., Bartlett, J. G., Rozo, E., et al. 2017, *A&A*, 604, A89
- Piffaretti, R., & Valdarnini, R. 2008, *A&A*, 491, 71
- Planck Collaboration VIII. 2011, *A&A*, 536, A8
- Planck Collaboration XI. 2011, *A&A*, 536, A11
- Planck Collaboration XXVII. 2016, *A&A*, 594, A27
- Planck Collaboration XXIV. 2016, *A&A*, 594, A24
- Postman, M., Coe, D., Benítez, N., et al. 2012, *ApJS*, 199, 25
- Pratt, G. W., Arnaud, M., Biviano, A., et al. 2019, *Space Sci. Rev.*, 215, 25
- Rasia, E., Ettori, S., Moscardini, L., et al. 2006, *MNRAS*, 369, 2013
- Rasia, E., Meneghetti, M., Martino, R., et al. 2012, *New J. Phys.*, 14, 055018
- Rasia, E., Meneghetti, M., & Ettori, S. 2013, *Astron. Rev.*, 8, 40
- Rasia, E., Lau, E. T., Borgani, S., et al. 2014, *ApJ*, 791, 96
- Reiprich, T. H., & Böhringer, H. 2002, *ApJ*, 567, 716
- Rines, K. J., Geller, M. J., Diaferio, A., & Hwang, H. S. 2016, *ApJ*, 819, 63
- Roze, E., Rykoff, E. S., Bartlett, J. G., & Evrard, A. 2014, *MNRAS*, 438, 49
- Salvati, L., Douspis, M., Ritz, A., Aghanim, N., & Babul, A. 2019, *A&A*, 626, A27
- Schellenberger, G., & Reiprich, T. H. 2017a, *MNRAS*, 469, 3738
- Schellenberger, G., & Reiprich, T. H. 2017b, *MNRAS*, 471, 1370
- Sereno, M. 2015, *MNRAS*, 450, 3665
- Sereno, M. 2016, *MNRAS*, 455, 2149
- Sereno, M., & Ettori, S. 2015, *MNRAS*, 450, 3633
- Sereno, M., & Ettori, S. 2017, *MNRAS*, 468, 3322
- Sereno, M., Ettori, S., & Moscardini, L. 2015, *MNRAS*, 450, 3649
- Sereno, M., Covone, G., Izzo, L., et al. 2017, *MNRAS*, 472, 1946
- Sereno, M., Umetsu, K., Ettori, S., et al. 2018, *ApJ*, 860, L4
- Serra, A. L., & Diaferio, A. 2013, *ApJ*, 768, 116
- Smith, G. P., Mazzotta, P., Okabe, N., et al. 2016, *MNRAS*, 456, L74
- Sohn, J., Geller, M. J., Diaferio, A., & Rines, K. J. 2020, *ApJ*, 891, 129
- Sunyaev, R. A., & Zeldovich, Y. B. 1972, *Comm. Astrophys. Space Phys.*, 4, 173
- Umetsu, K., Medezinski, E., Nonino, M., et al. 2014, *ApJ*, 795, 163
- Umetsu, K., Zitrin, A., Gruen, D., et al. 2016, *ApJ*, 821, 116
- Umetsu, K., Sereno, M., Lieu, M., et al. 2020, *ApJ*, 890, 148
- Vazza, F., Angelinelli, M., Jones, T. W., et al. 2018, *MNRAS*, 481, 120
- Vikhlinin, A., Kravtsov, A., Forman, W., et al. 2006, *ApJ*, 640, 691
- Vikhlinin, A., Kravtsov, A. V., Burenin, R. A., et al. 2009, *ApJ*, 692, 1060
- von der Linden, A., Mantz, A., Allen, S. W., et al. 2014, *MNRAS*, 443, 1973
- Zhang, Y.-Y., Finoguenov, A., Böhringer, H., et al. 2008, *A&A*, 482, 451
- Zhang, Y.-Y., Okabe, N., Finoguenov, A., et al. 2010, *ApJ*, 711, 1033
- Zhang, Y.-Y., Reiprich, T. H., Schneider, P., et al. 2017, *A&A*, 599, A138

Appendix A: Mass ratio with different estimators

In addition to the estimates from LIRA, we also quantify the mass ratios by computing the geometric mean⁵ (hereafter gmean) which provides, similarly to the median (also quoted in the table for an easier comparison with literature), an estimate of the typical mass ratio in the sample of interest (see e.g., Donahue et al. 2016; Okabe & Smith 2016; Umetsu et al. 2020). Errors in the estimates are computed via bootstrap resampling. The results are shown in Table A.1.

Overall, the three different estimators show the same trend when comparing the results for different samples. However, the bias obtained with LIRA are on average larger than what provided by gmean and median. This suggests that some of the objects with the largest bias have also very small statistical

uncertainties, and therefore have a larger weight in the fit with LIRA. To test that, we performed the fit with LIRA assuming that each cluster have the same average error (equal to the median error of the sample). The results are in very good agreement with the estimate from gmean which provides a good measure of the typical bias in the subsample of interest. Of course, a small statistical error is often a sign of a better data quality and should indeed be accounted for during the fit. Nonetheless, if any of the measurements is systematically biased, then small statistical uncertainties associated with it can have significant impact on the interpretation of the results. This is particularly true when using data from different studies because as argued by Rozo et al. (2014), the errors quoted in different papers account for different sources of statistical and systematic uncertainties and they are unable to account for the variance seen in sample pairs.

Table A.1. Summary of the bias between cluster masses estimated at R_{500} for X-ray, SZ, and WL, estimated with gmean, median and LIRA.

All												
Sample	$M_{\text{HE}}/M_{\text{WL}}$						$M_{\text{HE}}/M_{\text{SZ}}$					
	N	\tilde{M}_{WL}	\tilde{z}	gmean	median	LIRA	N	\tilde{M}_{SZ}	\tilde{z}	gmean	median	LIRA
LC ² -single	62	8.37	0.21	0.82 ^{+0.05} _{-0.05}	0.86 ^{+0.02} _{-0.05}	0.75 ^{+0.06} _{-0.06}	60	7.56	0.22	0.93 ^{+0.02} _{-0.02}	0.94 ^{+0.01} _{-0.03}	0.93 ^{+0.02} _{-0.02}
CCCP/MENeaCS	27	7.40	0.18	0.85 ^{+0.09} _{-0.07}	0.82 ^{+0.22} _{-0.15}	0.77 ^{+0.10} _{-0.10}	25	7.65	0.18	0.93 ^{+0.03} _{-0.03}	0.90 ^{+0.02} _{-0.02}	0.94 ^{+0.04} _{-0.04}
APEX-SZ	19	8.51	0.28	1.08 ^{+0.10} _{-0.10}	1.20 ^{+0.02} _{-0.22}	1.02 ^{+0.12} _{-0.12}	19	8.76	0.28	0.95 ^{+0.03} _{-0.03}	0.95 ^{+0.01} _{-0.01}	0.95 ^{+0.04} _{-0.04}
LoCuSS	18	8.24	0.22	0.80 ^{+0.06} _{-0.06}	0.79 ^{+0.25} _{-0.11}	0.76 ^{+0.09} _{-0.09}	18	7.78	0.22	0.87 ^{+0.03} _{-0.03}	0.88 ^{+0.01} _{-0.04}	0.88 ^{+0.04} _{-0.04}
WtG	17	11.73	0.29	0.66 ^{+0.08} _{-0.07}	0.65 ^{+0.14} _{-0.16}	0.62 ^{+0.12} _{-0.12}	17	8.27	0.29	0.91 ^{+0.03} _{-0.03}	0.89 ^{+0.04} _{-0.01}	0.92 ^{+0.04} _{-0.04}
Relaxed												
Sample	$M_{\text{HE}}/M_{\text{WL}}$						$M_{\text{HE}}/M_{\text{SZ}}$					
	N	\tilde{M}_{WL}	\tilde{z}	gmean	median	LIRA	N	\tilde{M}_{SZ}	\tilde{z}	gmean	median	LIRA
LC ² -single	34	8.65	0.21	0.84 ^{+0.06} _{-0.06}	0.87 ^{+0.11} _{-0.05}	0.75 ^{+0.08} _{-0.08}	32	7.45	0.21	0.96 ^{+0.03} _{-0.03}	0.94 ^{+0.03} _{-0.03}	0.96 ^{+0.03} _{-0.03}
CCCP/MENeaCS	17	7.10	0.17	0.92 ^{+0.09} _{-0.09}	1.05 ^{+0.02} _{-0.22}	0.83 ^{+0.12} _{-0.12}	15	7.78	0.18	0.96 ^{+0.05} _{-0.04}	0.90 ^{+0.02} _{-0.02}	0.96 ^{+0.05} _{-0.05}
APEX-SZ	11	8.75	0.27	0.96 ^{+0.11} _{-0.11}	1.03 ^{+0.17} _{-0.19}	0.90 ^{+0.17} _{-0.17}	11	8.13	0.27	0.93 ^{+0.05} _{-0.05}	0.94 ^{+0.01} _{-0.05}	0.93 ^{+0.07} _{-0.07}
LoCuSS	14	9.42	0.22	0.77 ^{+0.07} _{-0.07}	0.68 ^{+0.22} _{-0.03}	0.74 ^{+0.09} _{-0.09}	14	7.96	0.22	0.89 ^{+0.04} _{-0.04}	0.89 ^{+0.02} _{-0.04}	0.89 ^{+0.05} _{-0.05}
WtG	10	11.46	0.23	0.80 ^{+0.12} _{-0.11}	0.95 ^{+0.07} _{-0.29}	0.71 ^{+0.17} _{-0.17}	10	8.13	0.23	0.94 ^{+0.05} _{-0.04}	0.90 ^{+0.05} _{-0.02}	0.94 ^{+0.06} _{-0.06}
Disturbed												
Sample	$M_{\text{HE}}/M_{\text{WL}}$						$M_{\text{HE}}/M_{\text{SZ}}$					
	N	\tilde{M}_{WL}	\tilde{z}	gmean	median	LIRA	N	\tilde{M}_{SZ}	\tilde{z}	gmean	median	LIRA
LC ² -single	28	8.09	0.29	0.80 ^{+0.09} _{-0.07}	0.81 ^{+0.06} _{-0.20}	0.74 ^{+0.11} _{-0.11}	28	7.86	0.29	0.89 ^{+0.03} _{-0.03}	0.95 ^{+0.02} _{-0.06}	0.89 ^{+0.03} _{-0.03}
CCCP/MENeaCS	10	8.35	0.19	0.74 ^{+0.15} _{-0.11}	0.63 ^{+0.10} _{-0.05}	0.71 ^{+0.23} _{-0.23}	10	7.12	0.19	0.89 ^{+0.05} _{-0.05}	0.91 ^{+0.06} _{-0.06}	0.96 ^{+0.05} _{-0.05}
APEX-SZ	8	7.57	0.30	1.27 ^{+0.17} _{-0.15}	1.27 ^{+0.30} _{-0.22}	1.19 ^{+0.15} _{-0.15}	8	9.40	0.30	0.98 ^{+0.05} _{-0.05}	1.00 ^{+0.05} _{-0.05}	0.98 ^{+0.06} _{-0.06}
LoCuSS	4	5.89	0.24	0.93 ^{+0.14} _{-0.13}	1.05 ^{+0.06} _{-0.23}	0.93 ^{+0.39} _{-0.39}	4	6.89	0.24	0.82 ^{+0.05} _{-0.05}	0.82 ^{+0.10} _{-0.09}	0.82 ^{+0.16} _{-0.16}
WtG	7	15.20	0.41	0.51 ^{+0.07} _{-0.06}	0.50 ^{+0.01} _{-0.01}	0.48 ^{+0.19} _{-0.19}	7	8.27	0.41	0.86 ^{+0.05} _{-0.05}	0.88 ^{+0.05} _{-0.12}	0.89 ^{+0.08} _{-0.08}

Notes. The uncertainties for gmean and median have been computed from bootstrapping. Masses are in units of $10^{14} M_{\odot}$.

⁵ Opposite to the arithmetic means, the geometric mean is symmetric with respect to an exchange of the numerator and denominator (i.e., $\langle X/Y \rangle = \langle Y/X \rangle^{-1}$).

Appendix B: Comparing the X-ray masses from different studies

In Sect. 2.3 we discussed the differences between pairs of total X-ray mass measurements performed by different groups (i.e., L20, [Planck Collaboration XI 2011](#), [Martino et al. 2014](#), [Donahue et al. 2016](#), [Schellenberger & Reiprich 2017a](#), and [Ettori et al. 2019](#)). In Fig. B.1 we show the comparison between these total masses to better visualize the level of systematic differences between different studies. We note that, apart from the HIFLUGCS and a few CLASH clusters, all the plotted masses have been obtained using *XMM-Newton* data. Therefore, the differences cannot be attributed to cross-calibration difference between X-ray detectors, but instead to differences in the data analysis and/or to the methodology used to derive the mass (but see the discussion in Sect. 4 for the latter point).

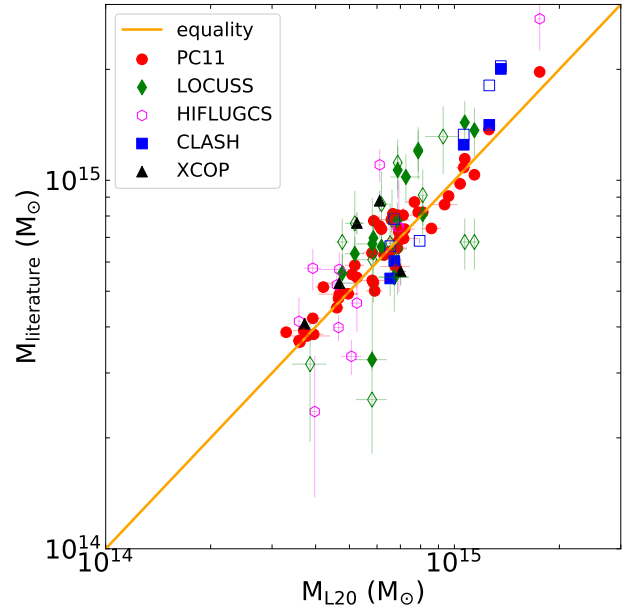


Fig. B.1. X-ray masses obtained in L20 compared with the masses reported in literature. Red circles, green diamonds, magenta hexagons, blue squares, and black triangles represent the comparison with the values by [Planck Collaboration XI \(2011\)](#), [Martino et al. \(2014\)](#), [Schellenberger & Reiprich \(2017a\)](#), [Donahue et al. \(2014\)](#), and [Ettori et al. \(2019\)](#), respectively. Fill and empty symbols represent the literature masses obtained with *XMM-Newton* and *Chandra*, respectively.