

# Biases in the estimation of velocity dispersions and dynamical masses for galaxy clusters

A. Ferragamo<sup>1,2,\*</sup>, J. A. Rubiño-Martín<sup>1,2</sup>, J. Betancort-Rijo<sup>1,2</sup>, E. Munari<sup>3,4</sup>, B. Sartoris<sup>3,4</sup>, and R. Barrena<sup>1,2</sup>

<sup>1</sup>Instituto de Astrofísica de Canarias (IAC), C/ Vía Láctea s/n, E-38205 La Laguna, Tenerife, Spain

<sup>2</sup>Universidad de La Laguna, Departamento de Astrofísica, C/ Astrofísico Francisco Sánchez s/n, E-38206 La Laguna, Tenerife, Spain

<sup>3</sup>Dipartimento di Fisica, Sezione di Astronomia, Università di Trieste, Via Tiepolo 11, I-34143 Trieste, Italy

<sup>4</sup>INAF/Osservatorio Astronomico di Trieste, Via Tiepolo 11, I-34143 Trieste, Italy

**Abstract.** Using a set of 73 numerically simulated galaxy clusters, we have characterised the statistical and physical biases for three velocity dispersion and mass estimators, namely biweight, gapper and standard deviation, in the small number of galaxies regime ( $N_{\text{gal}} \leq 75$ ), both for the determination of the velocity dispersion and the dynamical mass of the clusters via the  $\sigma$ - $M$  relation. These results are used to define a new set of unbiased estimators, that are able to correct for those statistical biases. By applying these new estimators to a subset of simulated observations, we show that they can retrieve bias-corrected values for both the mean velocity dispersion and the mean mass.

## 1 Introduction

Several authors have used the velocity dispersion mass proxy to study and characterise scaling relations between SZ and dynamical mass [1–3]. In order to have sample with enough statistical power, it is necessary to estimate the velocity dispersion for hundreds of galaxy clusters (GCs). Although this goal can be achieved through spectroscopic follow-up [e.g. 4, 5], these studies are extremely expensive in terms of observational time and data reduction. For these reasons, it is extremely difficult to estimate radial velocities for more than few members (typically  $\sim 20$ ) for each cluster target.

In this article we present our study of statistical and physical biases introduced in the estimation of velocity dispersion and dynamical mass.

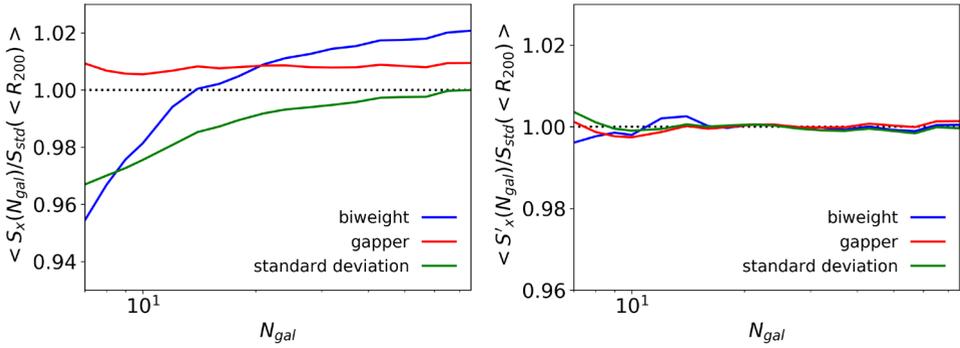
## 2 Statistical biases in Velocity Dispersion estimation

For our analysis we use a sample of 73 simulated massive clusters selected from the simulations described in [6]. Our selected sample contains clusters with masses  $M_{200} > 2 \times 10^{14} M_{\odot}$  and located at five redshifts between  $0.12 \leq z \leq 0.82$ .

There are several method to estimate mean and scale of a distribution. We focus our attention on the estimators that, in the last decades, became standard tools in GC analyses,

---

\*e-mail: ferragamo\_ext@iac.es



**Figure 1.** *Left panel:* mean velocity dispersion  $S_X/S_{std}(< R_{200})$  as a function of the number of galaxies  $N_{gal}$ . *Right panel:* corrected velocity dispersion estimators  $S'_X/S_{std}(< R_{200})$  as a function of  $N_{gal}$ . The dispersion  $S_X(N_{gal})$  is calculated for the standard deviation (green line), biweight (blue line), and gapper (red line) estimators.

**Table 1.** Best-fit parameters to be used in the parametric function given in equation 1, describing the bias of the three estimators. See text for details.

	<i>BWT</i>	<i>GAP</i>	<i>STD</i>
<i>D</i>	$0.72 \pm 0.03$	0	0.25
<i>B</i>	$-0.0225 \pm 0.0002$	$-0.0080 \pm 0.0002$	$-0.0037 \pm 0.0003$
$\beta$	$1.28 \pm 0.03$	1	1

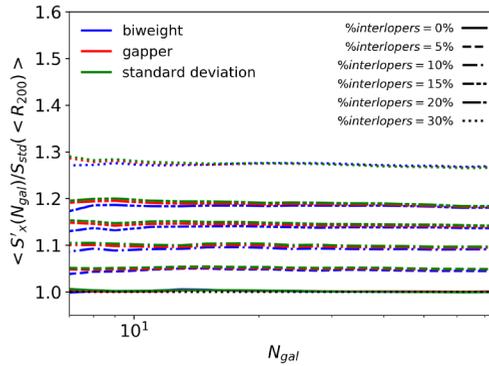
namely biweight and gapper, compared with the standard deviation. A detailed description of these estimators can be found in [7]. In order to investigate bias and variance of the three scale estimators  $S_X(N_{gal})$ , being  $X=$  “std”, “bwt” or “gap”, for each one of the three cases, we have explored the low number of galaxies regime between  $N_{gal} = 8$  and  $N_{gal} = 75$ . We have generated 2500 configurations of randomly selecting galaxies within the projected circle of radius  $R_{200}$  and then we averaged the  $73 \times 2250$  velocity dispersions (750 for each mayor axis).

Left panel of Fig. 1 shows the average  $S_X(N_{gal})$  normalised with respect to  $S_{std}(< R_{200})$ , which represents the velocity dispersion of all the galaxies in the simulation within a circle of projected radius  $R_{200}$ , and calculated using the standard deviation estimator. In the low- $N_{gal}$  regime, all estimators are biased showing very different behaviours. The standard deviation estimator (green line) shows a dependence with the number of elements used for the estimation. However, this dependence can be theoretically predicted to be  $1 - 1/(4(N_{gal} - 1))$ . The biweight (blue line) shows a stronger drop for  $N_{gal} \leq 30$  underestimating the reference dispersion by up to 4% at  $N_{gal} = 7$ . Finally, the gapper (red line) shows an estimate of the velocity dispersion almost constantly biased at any  $N_{gal}$ .

In order to construct an unbiased velocity dispersion estimators,  $S'_X(N_{gal})$ , we use a parametrisation of the curves in left panel of Fig. 1:

$$S'_X(N_{gal}) \equiv S_X(N_{gal}) \left( 1 + \left( \left( \frac{D}{(N_{gal} - 1)} \right)^\beta + B \right) \right). \quad (1)$$

Table 1 shows the best-fit values for the parameters  $D$ ,  $\beta$  and  $B$ , for each one of the three estimators (biweight, gapper and standard deviation). Right panel of Fig. 1 shows that the corrected estimators  $S'_X(N_{gal})$  are actually unbiased by construction.



**Figure 2.** Effect of interlopers on  $S'_X/S_{std}(< R_{200})$ , as a function of the number of galaxies. The velocity dispersion,  $S'_X(N_{gal})$ , is first computed using a pure galaxie sample (solid lines). We also evaluated the response of biweight (blue lines), standard deviation (green lines), and gapper (red lines) using samples contaminated by 0% (solid line), 5% (dashed lines), 10% (dot dashed lines), 15% (three-dot-dashed lines), 20% (two-dot-long-dashed lines) and 30% (dotted lines) of interlopers at any  $N_{gal}$ .

### 3 Biases by interlopers contamination

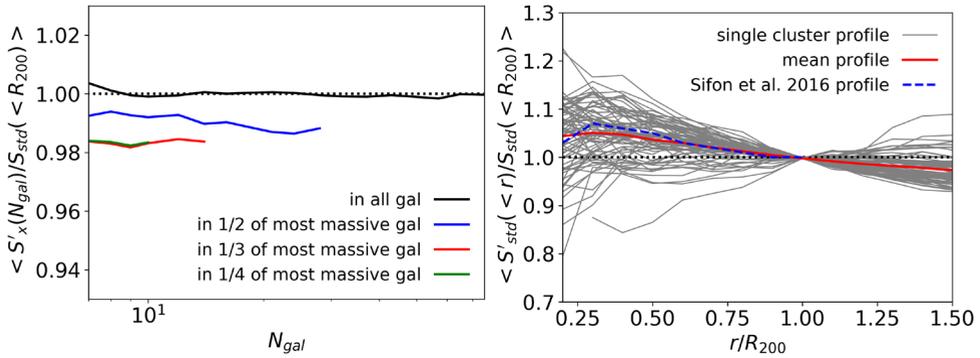
Any spectroscopic sample of cluster members is contaminated by galaxies belonging to the large scale structure that surrounds the cluster itself. This population of ‘pseudo cluster members’, called interlopers, modifies the velocity distribution and therefore affects the estimation of velocity dispersion. According to the definition of interlopers given in [8], one must distinguish between two very different types of contaminants: i. galaxies gravitationally bound to the clusters that are outside the virial sphere (according to the definition given in [9]), but that, due to projection effects appear within a smaller radius; ii. background/foreground galaxies with similar redshifts to the cluster, but belonging to the large scale structure that surrounds the cluster itself. Detailed study of these interlopers is beyond the scope of this work. Here, we illustrate the robustness of the three estimators in exam, by fixing the fraction of contaminants at any  $N_{gal}$ . Fig. 2 shows the results obtained for five different fractions of interlopers: 5% (dashed lines), 10% (dot-dashed lines), 15% (three-dot-dashed lines), 20% (two-dot-long-dashed lines) and 30% (dotted lines). We note that the three estimators are similarly affected by interloper contamination. The effect of the interlopers contamination consists of a velocity dispersion overestimation which is as high as their relative fraction (up to 30%) representing the most damaging source of biases in the velocity estimation estimation.

### 4 Physical biases in Velocity Dispersion estimation

Observational strategies and technical limitations generally force us to observe only massive clusters members sampled in a fraction of  $R_{200}$ . We studied how these limitations might produce biased velocity dispersion estimates.

#### 4.1 Effects due to the selected fraction of massive galaxies

In the ideal case we can observe any cluster member regardless of its brightness. However, the telescope aperture limits the detection magnitude and prevents us from detecting faint



**Figure 3.** *Left panel:* Mean (bias) of  $S'_{std}(N_{gal})/S_{std}(< R_{200})$  as a function of the number of galaxies  $N_{gal}$ , calculated by choosing galaxies within 100% (black solid line), 1/2 (blue solid line), 1/3 (red solid line), and 1/4 (green solid line) of the complete cluster member samples. *Right panel:* Average velocity dispersion profile within a given enclosed radius  $r$ ,  $\langle S'_{std}(< r)/S_{std}(< R_{200}) \rangle$ , normalised to  $R_{200}$ . The red line represents the mean at each radius of the individual 73 simulated GC profiles (grey lines). The dashed blue line represents the [2] profile, which is almost coincident with our derived profile.

objects, for a fixed integration time. Therefore, cluster samples contain only a fraction of the brightest galaxy members, which are also the most massive.

In order to simulate this effect, we mimicked observational conditions by selecting three percentages of all visible galaxies in the simulation, i.e. 50 %, 33 %, and 25 %, by sorting the cluster members by mass and dividing the sample in 2, 3, and 4 mass bins, starting from the most massive object. For each case, we reproduce the procedure explained in Sect. 2.

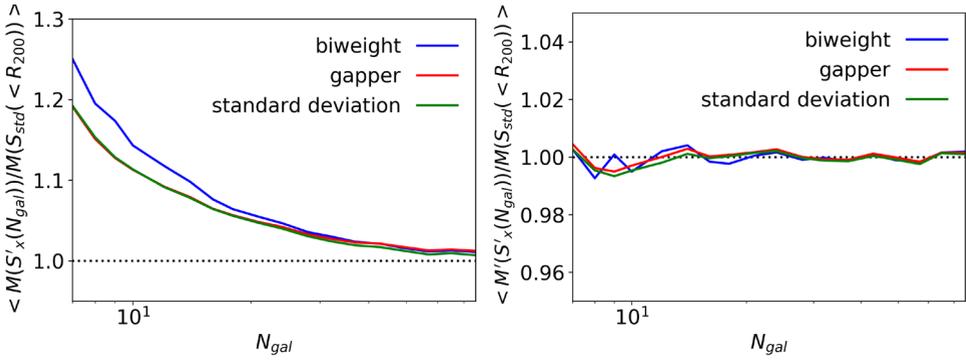
Fig. 3 (left panel) shows  $S'_{std}(N_{gal})/S_{std}(< R_{200})$  as function of  $N_{gal}$  calculated with the corrected standard deviation estimator, and using galaxies picked up from 100 % (black line), 1/2 (blue line), 1/3 (red line) and 1/4 (green line) of the complete cluster member samples. We see that the velocity dispersion is sensitive to the fraction of massive galaxies used to estimate it, reaching a bias of almost 2% using only the most massive fraction. However, it is almost insensitive to  $N_{gal}$ . For this reason, we can use the curves in the left panel of Fig. 3 to correct this physical effect.

## 4.2 Effect of aperture sub-sampling

There are evidence in the literature that the velocity dispersion has a radial dependence [e.g., 2, 9]. This implies that sampling galaxies in different fractions of the cluster's virial radius should lead to a biased velocity dispersion. In order to investigate and quantify this effect, we averaged the 73 velocity dispersions calculated using all galaxies inside a cylinder of variable radius  $0.2 \leq r/R_{200} \leq 1.2$ . Fig. 3 (right panel) shows that the velocity dispersion is (on average) overestimated in region smaller than  $R_{200}$  and slightly underestimated for  $r > R_{200}$ , in analogy with that recover by [2].

## 5 Statistical bias in the estimation of $M_{200}$

When we use velocity dispersion as a mass proxy we apply a scaling relation  $\sigma_{1D} - M$ , it is a power law that has been previously calibrated either with simulations [6, 10, 11]. The left panel of Fig. 4 shows how applying the scaling relation  $\sigma_{1D} - M$  to the unbiased velocity dispersion estimator  $S'_x$  the resulting masses  $\langle M(S'_x)/M(S_{std}(< R_{200})) \rangle$  are biased.



**Figure 4.** *Left panel:* mean of  $M(S'_X(N_{gal}))/M(S'_{std}(< R_{200}))$ , which represent the standard mass estimator, described in [6], applied to normal and unbiased velocity dispersion estimators, standard deviation (green), gapper (red), and biweight (blue). *Right panel:* mean of the corrected mass estimator  $M'(S'_X(N_{gal}))/M(S'_{std}(< R_{200}))$ .

**Table 2.** Best-fit parameters for the function describing the bias in  $\langle M(S'_X)/M(S_{std}(< R_{200})) \rangle$  for simulated clusters, as described in equation 2.

	<i>BWT</i>	<i>GAP</i>	<i>STD</i>
$E'$	$1.31 \pm 0.03$	$1.50 \pm 0.03$	$1.53 \pm 0.03$
$F'$	1	1	1
$\gamma'$	$1.24 \pm 0.03$	$1.17 \pm 0.04$	$1.11 \pm 0.04$

We can calculate analytically the functional form of the mass estimator as a function of the number of galaxies  $\frac{1-2\alpha}{4\alpha^2(N_{gal}-1)}$ , with  $A = 1177.0 \text{ km s}^{-1}$  and  $\alpha = 0.364$  parameters of the [6] scaling relation.

As in Sect. 2, we fitted a parametric form description of the bias as a function of  $N_{gal}$  by using three parameters ( $E', F'$  and  $\gamma'$ , listed in Table 2):

$$M'(S'_X(N_{gal})) = M(S'_X(N_{gal})) \left[ \frac{1 - E'\alpha}{(E'\alpha)^2(N_{gal} - 1)^{\gamma'}} + F' \right]^{-1}. \quad (2)$$

The right panel of Fig. 4 shows the bias of  $M'(S'_X)$ . The new mass estimator is actually unbiased by construction.

Finally, in order to test the corrections described in previous sections, we have applied them to a set of mock cluster catalogues. To do this we have simulated a realistic observational strategy based on the observations described in [4, 5]. We generated 100 mock samples out of the 73 GCs object simulated in this study. For each of these samples we calculated the mean ratio between the estimated and the reference velocity dispersion of each cluster, as well as the velocity dispersion calculated using  $S'_X$  defined in eq. 1 with the parameters in Table 1. Averaging over all the mock samples we obtained a biased mean velocity dispersion

$$\begin{aligned} \langle S_{std}(N_{gal}, r)/S_{bwt}(< R_{200}) \rangle &= 0.96 \pm 0.02, \\ \langle S'_{std}(N_{gal}, r)/S_{bwt}(< R_{200}) \rangle &= 1.00 \pm 0.02. \end{aligned} \quad (3)$$

The normal  $S_{std}$  estimator shows to be biased, whereas the  $S'_X$  lead to an unbiased estimation of the velocity dispersion.

Using the bias corrected velocity dispersions, we calculated cluster masses,  $M(S'_X)$  and  $M'(S'_{\text{std}})$ , obtaining

$$\begin{aligned}\langle M(S'_{\text{std}}(N_{\text{gal}}, r)) / M(S_{\text{bwt}}(< R_{200})) \rangle &= 1.13 \pm 0.07, \\ \langle M'(S'_{\text{std}}(N_{\text{gal}}, r)) / M'(S_{\text{bwt}}(< R_{200})) \rangle &= 1.00 \pm 0.06.\end{aligned}\quad (4)$$

As described above, these normal mass estimator is overestimated, while primed mass estimator,  $M'$  is actually unbiased.

## 6 Conclusions

We have used 73 simulated GCs from hydrodynamic simulations including AGN feedback and star formation, in order to characterise the statistical and physical biases in three velocity dispersion (and mass) estimators frequently used in the literature: the biweight, the gapper, and the standard deviation. We have focused our study in the (common) case of a low number of galaxy members ( $N_{\text{gal}} \lesssim 75$ ).

We showed that each of these estimators (dispersion and mass) presents a statistical bias. Therefore, we propose bias corrected velocity dispersion ( $S'_X$ ) and mass ( $M'(S'_X)$ ) estimators.

We have tested the robustness of the new estimators against the contamination by interlopers. We found that the velocity dispersion estimators  $S'_X$  are similarly affected by the contamination for all the three cases in this low- $N_{\text{gal}}$  limit.

We observed that the most likely sources of physical bias are i) the selection effect the fraction of massive galaxies used to estimate the velocity dispersion; and ii) the fraction of the virial radius explored. We showed that in the first case the bias is estimated to be around 2 % when considering only 1/4 of the most massive galaxies. Concerning the effect produced by the sampling aperture, we find a dispersion radial profile in agreement with [2] results.

## References

- [1] J. Ruel, G. Bazin, M. Bayliss, M. Brodwin, R.J. Foley, B. Stalder, K.A. Aird, R. Armstrong, M.L.N. Ashby, M. Bautz et al., *ApJ***792**, 45 (2014), 1311.4953
- [2] C. Sifón, N. Battaglia, M. Hasselfield, F. Menanteau, L.F. Barrientos, J.R. Bond, D. Crichton, M.J. Devlin, R. Dünner, M. Hilton et al., *MNRAS***461**, 248 (2016), 1512.00910
- [3] S. Amodeo, S. Mei, S.A. Stanford, J.G. Bartlett, J.B. Melin, C.R. Lawrence, R.R. Chary, H. Shim, F. Marleau, D. Stern, *ApJ***844**, 101 (2017), 1704.07891
- [4] Planck Collaboration Int. XXXVI, *A&A***586**, A139 (2016), 1504.04583
- [5] R. Barrena, A. Streblyanska, A. Ferragamo, J.A. Rubino-Martin, A. Aguado-Barahona, D. Tramonte, R.T. Genova-Santos, A. Hempel, H. Lietzen, N. Aghanim et al., *ArXiv e-prints* (2018), 1803.05764
- [6] E. Munari, A. Biviano, S. Borgani, G. Murante, D. Fabjan, *MNRAS***430**, 2638 (2013), 1301.1682
- [7] T.C. Beers, K. Flynn, K. Gebhardt, *AJ***100**, 32 (1990)
- [8] G.W. Pratt, M. Arnaud, A. Biviano, D. Eckert, S. Ettori, D. Nagai, N. Okabe, T.H. Reiprich, *Space Sci. Rev.***215**, 25 (2019), 1902.10837
- [9] G.A. Mamon, A. Biviano, G. Murante, *A&A***520**, A30 (2010), 1003.0033
- [10] A.E. Evrard, J. Bialek, M. Busha, M. White, S. Habib, K. Heitmann, M. Warren, E. Rasia, G. Tormen, L. Moscardini et al., *ApJ***672**, 122-137 (2008), astro-ph/0702241
- [11] A. Saro, J.J. Mohr, G. Bazin, K. Dolag, *ApJ***772**, 47 (2013), 1203.5708