Submitted for publication in ApJL
Preprint typeset using LaTeX style emulateapj v. 01/23/15

QUANTIFYING AND CONTROLLING BIASES IN DARK MATTER HALO CONCENTRATION ESTIMATES

C. N. Poveda-Ruiz [1] J. E. Forero-Romero [1] J. C. Muñoz-Cuartas [2]
[1]Departamento de Física, Universidad de los Andes, Cra. 1 No. 18A-10, Edificio Ip, Bogotá, Colombia
[2]Instituto de Física - FCEN, Universidad de Antioquia, Calle 67 No. 53-108, Medellín, Colombia
Submitted for publication in ApJL

ABSTRACT

We use bootstrapping to estimate the bias of concentration estimates on N-body dark matter halos as a function of particle number. We find that algorithms based on the maximum radial velocity and radial particle binning tend to overestimate the concentration by $15\% - 20\%$ for halos sampled with 200 particles and by 7%-10% for halos sampled with 500 particles. To control this bias at low particle numbers we propose a new algorithm to estimate halo concentrations based on the integrated mass profile. The method uses the full particle information without any binning, making it reliable in cases when low numerical resolution becomes a limitation for other methods. This method reduces the bias to $1\% - 3\%$ for halos sampled with 200-500 particles. The velocity and density methods have to use halos with least ~ 4000 particles in order to keep the biases down to the same low level. We also show that the mass-concentration relationship could be shallower than expected from numerical experiments once the biases of the different concentration measurements are taken into account. These results show that bootstrapping and the concentration estimates based on the integrated mass profile are valuable tools to probe the internal structure of dark matter halos in numerical simulations.

Subject headings: Galaxies: halos — Galaxies: high-redshift — Galaxies: statistics — Dark Matter — Methods: numerical

1. INTRODUCTION

In the current structure formation paradigm the properties of galaxies are coupled to the evolution of their dark matter (DM) hosting halo. In this paradigm the sizes and dynamics of galaxies are driven by the DM distribution inside a halo. This has motivated the detailed study a DM halo internal structure in the last three decades.

The internal DM distribution in a halo is usually parameterized through the density profile. In a first approximation this profile is spherically symmetric; the density only depends on the radial coordinate. One of the most popular radial parameterizations is the Navarro-Frenk-White (NFW) profile (Navarro et al. 1997). This profile can be considered as universal (Navarro et al. 2010), assuming that one is not interested in the very central region of the dark matter halo where galaxy formation takes place, and where the effects of baryon physics on dark matter distribution are still unknown. This profile is a double power law in radius, where the transition break happens at the so-called scale radius, r_s. The ratio between the scale radius and the halo virial radius R_v is known as the concentration $c = R_v/r_s$.

The concentration of the NFW profile provides a conceptual framework to study simulated dark matter halos as a function of redshift and cosmological parameters. In many numerical studies (Neto et al. 2007; Macciò et al. 2008; Duffy et al. 2008; Muñoz-Cuartas et al. 2011; Prada et al. 2012; Ludlow et al. 2014) the results are summarized through the mass-concentration relationship; that is the distribution of concentration values at a fixed halo mass and redshift. The success of such numerical experiments rests on a reliable algorithm to estimate the concentration. Such an algorithm should provide unbiased results and must be robust when applied to simulations of different numerical resolution.

There are two established algorithms to estimate the concentration parameter of a N-body dark matter halo. The first method takes the halo particles and bins them into logarithmic radii to estimate the density in each bin, then it proceeds to fit the density as a function of the radius to the NFW profile. A second method uses an analytic property of the NFW profile that relates the maximum of the ratio of the circular velocity to the virial velocity, $V_{\rm circ}/V_{\rm vir}$. The concentration can be then found as the root of an algebraic equation dependent on this maximum value.

The first method is straightforward to apply but presents two disadvantages. First, it requires a large number of particles in order to have a proper density estimate in each bin. This makes the method robust only for halos with at least 10^2 particles. The second problem is that there is not a way to estimate the optimal radial bin size, different choices may produce different results for the concentration.

The second method solves the two problems mentioned above. It works with low particle numbers and does not involve data binning. However, it effectively takes into account only a single data point and discards the behavior of the ratio $V_{\rm circ}/V_{\rm vir}$ below and above its maxima. Small fluctuations on the value of this maximum can yield large perturbations on the estimated concentration parameter.

In this letter we use bootstrapping to estimate the bias and standard deviation on the concentration estimates as a function of particle number. We show that the two standard methods to estimate concentrations have increasing biases for decreasing particle numbers.

This motivates us to present a third alternative based on fitting the integrated mass profile. This approach has two advantages with respect to the above mentioned

methods. It does not involve any data binning and does not throw away data points. This translates into a robust estimate even at low resolution/particle numbers. Furthermore, since the method does not require any binning, there is no need to tune numerical parameters. This method provides a new independent method to estimate the concentration parameter.

2. BASIC PROPERTIES OF THE NFW DENSITY PROFILE

Let us review first the basic properties of the NFW density profile. This shall help us to define our notation.

2.1. *Density profile*

The NFW density profile can be written as

$$\rho(r) = \frac{\rho_c \delta_c}{r/r_s(1+r/r_s)^2}, \tag{1}$$

where $\rho_c \equiv 3H^2/8\pi G$ is the Universe critical density, H is the Hubble constant, G and the universal gravitational constant, δ_c is the halo dimensionless characteristic density and r_s is the scale radius. This radius marks the point where the logarithmic slope of the density profile is equal to -2, the transition between the power law scaling $\rho \propto r^{-1}$ for $r < r_s$ and $\rho \propto r^{-3}$ for $r > r_s$.

We define the virial radius of a halo, r_v, as the boundary of the spherical volume that encloses a density of Δ_h times the mean density of the Universe. The corresponding mass M_v, the virial mass, can be written as $M_v = \frac{4\pi}{3}\bar{\rho}\Delta_h r_v^3$. From these virial quantities we define new dimensionless variables for the radius and mass $x \equiv r/r_v$ and $m \equiv M(<r)/M_v$.

In this letter we use $\Delta_h = 740$, a number roughly corresponding to 200 times the critical density at redshift z=0.

2.2. *Integrated mass profile*

From these definitions we can compute the total mass enclosed inside a radius r:

$$M(<r) = 4\pi\rho_c\delta_c r_s^3 \left[\ln\left(\frac{r_s+r}{r_s}\right) - \frac{r}{r_s+r}\right], \tag{2}$$

or in terms of the dimensionless mass and radius variables

$$m(<x) = \frac{1}{A}\left[\ln(1+xc) - \left(\frac{xc}{xc+1}\right)\right], \tag{3}$$

where

$$A = \ln(1+c) - \left(\frac{c}{c+1}\right), \tag{4}$$

and the parameter c corresponds to the concentration $c \equiv r_v/r_s$.

From this normalization and for later convenience we define the following function

$$f(x) = \ln(1+x) - \left(\frac{x}{x+1}\right). \tag{5}$$

The most interesting feature of Eq. (3) is that the concentration is the only free parameter to describe the integrated mass profile.

2.3. *Circular velocity profile*

It is also customary to express the mass of the halo in terms of the circular velocity $V_c = \sqrt{GM(<r)/r}$. From this we can define a new dimensionless circular velocity $v(<x) \equiv V_c(<r)/V_c(<r_v)$, using the result in Eq. 3 we have:

$$v(<x) = \sqrt{\frac{1}{A}\left[\frac{\ln(1+xc)}{x} - \frac{c}{xc+1}\right]}, \tag{6}$$

This normalized profile always shows a maximum provided that the concentration is larger than $c > 2$. It is possible to show that for the NFW profile the maximum is provided by

$$\max(v(<x)) = \sqrt{\frac{c}{x_{\max}}\frac{f(x_{\max})}{f(c)}}, \tag{7}$$

where $x_{\max} = 2.163$ (Klypin et al. 2016) and the function $f(x)$ corresponds to the definition in Eq. (5).

3. METHODS TO ESTIMATE THE CONCENTRATION FROM N-BODY SIMULATIONS

3.1. *Estimates from the density and velocity profiles*

To date, there are two standard methods to estimate concentrations in dark matter halos extracted from N-body simulations. The first method tries to directly estimate the density profile. It takes all the particles in the halo and bins them in the logarithm of the radial coordinate from the halo center. Then, it estimates the density in each logarithmic bin counting the particles and dividing by the corresponding shell volume. At this point is possible to make a direct fit to the density as a function of the radial coordinate. This method has been broadly used for more than two decades to study the mass-concentration-redshift relation of dark matter halos.

A second method uses the circular velocity profile. It finds the value of x for which the normalized circular velocity $v(<x)$ shows a maximum. Using this value it solves numerically for the corresponding value of the concentration using Eq. (7). This method has been most recently used by Klypin et al. (2016) to study the mass-concentration-redshift relation using the Multidark Simulation Suite.

3.2. *Our proposal: estimate from the integrated mass profile*

Here we propose a new method to estimate the concentration. It uses the integrated mass profile defined in Eq. (3). We build it from N-body data following the next steps. First, we define the center of the halo to be at the position of the particle with the lowest gravitational potential. Then we rank the particles by their increasing radial distance from the center. From this ranked list of $i = 1, N$ particles, the total mass at a radius r_i is $M_i = i \times m_p$, where r_i is the position of the i-th particle and m_p is the mass of a single computational particle.

We divide the enclosed mass mass M_i and the radii r_i by their virial values to obtain the dimensionless variables m_i and x_i. Once the mass profile is expressed in dimensionless variables the concentration is the only free parameter.

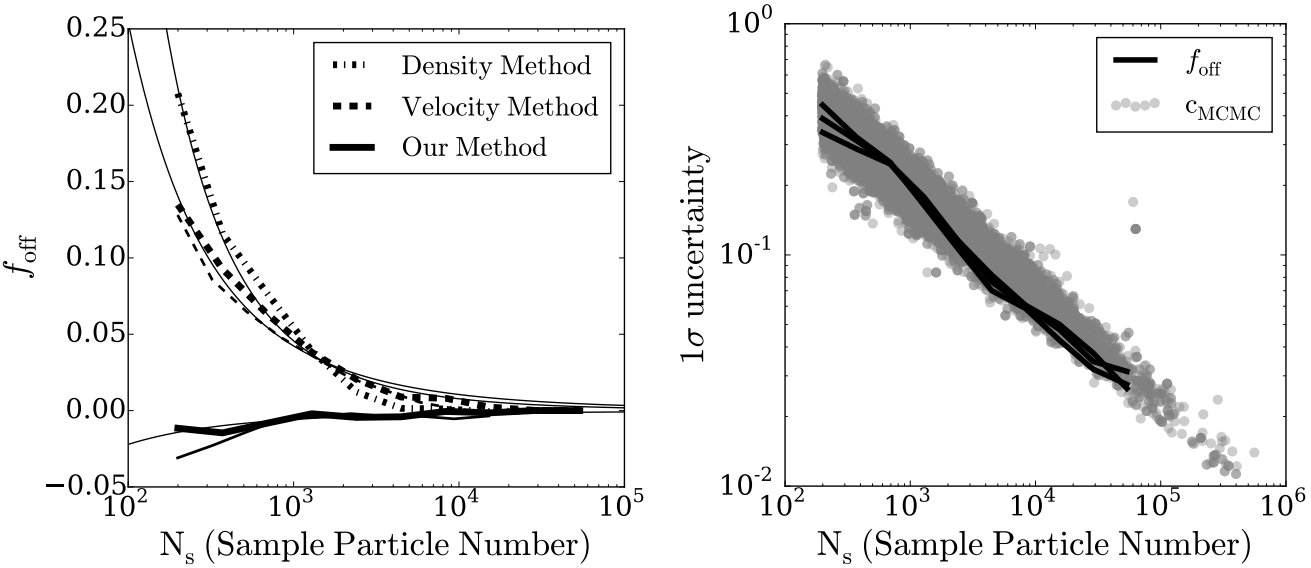

FIG. 1.— Left panel. Average value of the fractional offset ($f_{\rm off}$ (Eq. 11) between the concentration in a halo at high resolution and its bootstrapped versions. Thick (thin) lines correspond to massive halos in the Bolshoi (Via Lactea) simulation. For each case close to 10000 bootstrapping samples are used. The velocity method noticeably overestimates the concentration up to a factor of 0.20, while the new method only underestimates the concentrations by up to a factor of 0.04. Right panel. 1σ uncertainties on the offset and the concentration estimates from the integrated mass method. Lines show the width between the 14 and 86 percentiles of the $f_{\rm off}$ distribution at fixed particle number. The lines include the results for the three methods: maximum velocity, density and integrated mass. Grey symbols correspond to the 1σ uncertainty on each of the halos in the Bolshoi Simulation as a result of the MCMC sampling. To allow a fair comparison with $f_{\rm off}$, this uncertainty has been normalized by the preferred concentration value for each halo. In both cases (bootstrapping for individual halos and MCMC fit of a halo population) the 1σ is of the same order with a decreasing behavior as $1/\sqrt{N_s}$.

Using the bootstrapping data we find that at a given normalized radius, x, the logarithm of the normalized integrated mass, m, follows a distribution that can approximated by a gaussian with variance

$$\sigma_x^2 = \frac{1-x}{x}\frac{1}{6N}. \quad (8)$$

If the integrated mass values at different radii were independent from each other, this could be used as the basis to develop a simple bayesian likelihood to be sampled with Markov Chain Monte Carlo techniques where the likelihood could be then defined to be $\mathcal{L}(c) \propto \exp(-\chi^2(c)/2)$ with a $\chi^2(c)$ written as

$$\chi^2(c) = \sum_{i=2}^{N-1} \frac{[\log m_i - \log m(<x_i;c)]^2}{\sigma_i^2}, \quad (9)$$

where $\sigma_i^2 = \sigma_x^2$, $m(<x_i;c)$ corresponds to the values in Eq.(3) at $x = x_i$ for a given value of the concentration parameter c, and the i index sums over all the particles in the numerical profile. In this computation the particles $i = 1$ and $i = N$ are discarded to avoid divergent terms in the sum.

However, tests on the bootstrapping data show that the covariance matrix is not diagonal and using $\sigma_i^2 = \sigma_x^2$ would grossly overestimate $\chi^2(c)$ providing too optimistic error bars.

Given the high expense of computing a full covariance matrix for each desired number of particles N, we have calibrated an effective $\sigma_{\rm eff}^2$ such that it keeps the dependence on x that we have discovered for the diagonal elements and also gives similar curves of $\chi^2(c)$-c around the minimum as the full covariance matrix computation.

We found that this effective $\sigma_{\rm eff}^2$ can be approximated as

$$\sigma_{\rm eff}^2 = \frac{1-x}{x}\frac{N^{1.15}}{4.5\times10^3}. \quad (10)$$

We then use $\sigma_i^2 = \sigma_{\rm eff}^2$ to feed an Affine Invariant Markov Chain Monte Carlo implemented in the python module `emcee` (Foreman-Mackey et al. 2013) to sample the likelihood function distribution. From the χ^2 distribution we find the optimal value of the concentration and its associated uncertainty.

4. NUMERICAL SIMULATIONS AND HALO SAMPLES

We use two very different simulations to be sure that the results (bias, uncertainty on the bias) measured using bootstrapping tests are independent of the kind of parent simulation.

We use data from the Bolshoi cosmological simulation that follows the non-linear evolution of a dark matter density field sampled with 2048^3 particles over a cubic box of 250 h^{-1}Mpc on a side. The cosmological parameters use a Hubble parameter $h = 0.73$, a matter density $\Omega_m = 0.3071$ and a normalization of the power spectrum $\sigma_8 = 0.82$. The data is publicly available at http://www.cosmosim.org/. Details about the structure of the database and the simulation can be found in (Riebe et al. 2013).

We use a halo sample containing all the halos located in a cubic sub-volume of 100 h^{-1}Mpc on a side. From this sample we select all the halos at $z = 0$ detected with a Friends-of-Friends (FoF) algorithm with more than 300 particles, meaning that the masses are in the interval $4\times10^{10} \leq M_{\rm FoF}/h^{-1}{\rm M}_\odot \leq 10^{14}$. The FoF algorithm used a linking length of 0.17 times the mean inter-particle distance. This choice translates into an overdensity $\Delta_h \sim 400-700$ dependent on the halo concentration (More et al. 2011).

From this set of particles we follow the procedure spelled out in Section 3 with $\Delta_h = 740$ (corresponding to 200 times the critical density) to select an spherical region that we redefine to be our halo. This choice makes that the overdensities are fully included inside the original FoF particle group. On the interest of providing a fair comparison against the density method we only report results from overdensities with at least 200 particles ($2.6 \times 10^{10} h^{-1} \mathrm{M}_\odot$).

We also use public data from the Via Lactea simulation project (Diemand et al. 2008). This simulation was run for a single isolated halo with a virial mass of the order of $10^{12} h^{-1} \mathrm{M}_\odot$using the parallel tree code PKDGRAV (Stadel 2001). The simulation used $\sim 2 \times 10^8$ particles to resolve this region. The cosmological parameters are different from those in the Bolshoi simulation, with a Hubble parameter $h = 0.73$, a matter density $\Omega_m = 0.238$ and a normalization of the power spectrum $\sigma_8 = 0.74$. The data available to the public corresponds to a downsampled set of 10^5 particles which corresponds to a particle mass of $2.24 \times 10^7 h^{-1} \mathrm{M}_\odot$.

5. RESULTS

5.1. *Bootstrapping to estimate biases*

In our bootstrapping experiments we take halos sampled with at least 10^5 particles and subsample each halo by factors of 2 up to 10^3 to measure the concentration every time the halo is resampled. Every time we keep fixed the virial radius and the center found for the high resolution halo. We tested that leaving the virial radius and center free in each bootstrapping iteration has an effect lower than 1% in the concentration. We extract in total 10000 bootstrapping samples for the Bolshoi and the Via Lactea simulations. In the case of the Bolshoi simulation we select 14 massive halos and create 700 subsamples for each halo. For the Via Lactea simulation only one halo is subsampled 10000 times.

The average concentration value for the largest number of particles, $c_{N_{max}}$, provides with a baseline to compare all the other results. We use the following statistic

$$f_{\rm off} = c_N / c_{N_{max}} - 1, \tag{11}$$

to account for the offset between the concentration at a given downsampled particle number c_N and the baseline $c_{N_{max}}$. For each method we keep its corresponding concentration $c_{N_{max}}$ value at high resolution.

Figure 1 summarizes our results. The plot on the left shows the average value of $f_{\rm off}$ as a function of the particle number. As expected, for large enough sample particle numbers, $N_s > 4 \times 10^3$, the results of the three algorithms coincide below the 1% level. For a lower number of particles, the results deviate from the expected concentration value. For 200 particles the velocity method overestimates by a factor 14% while the density method overestimates it by 20%. Around the same sampling scale, the new algorithm shows a more stable behaviour underestimating the concentration only by a factor of 1%-2%

The line on the same panel shows a fit to the data with the following functional form

$$f_{\rm off} = \frac{A}{(1 + \log_{10} N_s)^B}. \tag{12}$$

for the dataset shown in the panel we have $A = 2842 \pm 1900$, $B = 7.96 \pm 0.54$ for the density method; $A = 239 \pm 131$, $B = 6.23 \pm 0.43$ for the velocity method; $A = -3.66 \pm 5.22$, $B = 4.65 \pm 1.12$ for the integrated mass method. The uncertainties are estimated from the covariance matrix resulting from the non-linear optimization procedure. The lines on Figure 1 are plotted using these best values.

The right panel in 1 shows different uncertainty results. The lines show the difference between the 14 and 86 percentiles in the $f_{\rm off}$ distribution at fixed mass. Each line corresponds to the three different methods to estimate the concentration and the two simulations. This shows that the bootstrapping technique can help us to assign a 1σ uncertainty to the concentration values at a fixed N_s. This uncertainty is shown as normalized by the preferred value of the concentration, which in this case corresponds to $c_{N_{max}}$. The circles in the same figure show the 1σ uncertainty as estimated from the MCMC chains using the integrated mass method on all the relaxed halos in the Bolshoi simulation sample. To allow for a fair comparison with the bootstrapping results, this uncertainty was normalized by the preferred concentration. The MCMC uncertainties follow the same trend as the bootstrapping results.

From these results we defined a universal 1σ fractional uncertainty for the concentration as a function of the number of particles N_s as

$$\sigma_c = \frac{0.40}{\sqrt{N_s/200}}. \tag{13}$$

An interesting consequence of this result is that $\sigma_c = 1$ for $N_s = 32$ particles, independent of the method used to estimate the concentration.

5.2. *Impact on the Mass-Concentration Relationship*

We now compare the results for the mass-concentration relationship obtained with the three different algorithms. This helps to compare the concentration results of the new algorithm and, more importantly, showcase a probable consequence of the biases detected through bootstrapping.

Figure 2 shows the mass-concentration relationship for the density, velocity and integrated mass method. The left panel shows the results as they are produced by each of the algorithms, the thin dashed line marks the results by (Prada et al. 2012).

We notice first that the results from the new algorithm follow very closely the velocity algorithm at high masses ($M_h > 10^{12} h^{-1} \mathrm{M}_\odot$ or equivalently for $> 8 \times 10^3$ particles). For lower masses there is a difference between the median of the two methods, but they are still consistent within the statistical uncertainties.

We also notice that the results from the density method have a systematic 15% offset from the velocity methods. This offset was already presented by Prada et al. (2012) for low concentrations ($c < 6$) and high ($M_h > 10^{12} h^{-1} \mathrm{M}_\odot$) halo masses. Recently (Klypin et al. 2016) summarized results for the mass-concentration relationship coming from different methods and datasets to show that similar systematic offsets are present. (Dutton & Macciò 2014) studied the mass-concentration relationship using the maximum velocity and density methods and did not report any significant difference. However,

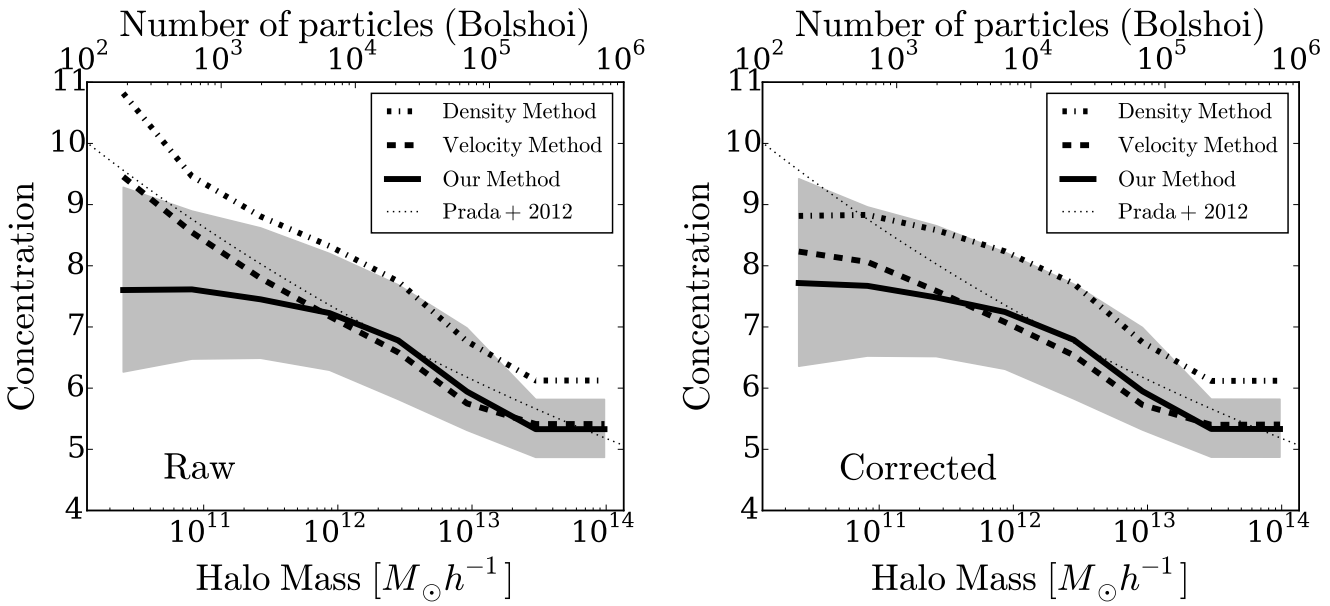

FIG. 2.— Mass-concentration relationship for the three different methods on the Bolshoi data using only relaxed halos. The lines correspond to the median concentration values in each mass bin. The shaded region presents 10 and 90 per cent spread. The three methods have a similar spread but for clarity we only show the spread for the concentrations estimated using the new method. The dotted line corresponds to fits reported by (Prada et al. 2012). . The left panel shows the raw results coming from each algorithm. The right panel introduces a correction on the velocity and integrated mass results following the results of the downsampling experiments presented in Figure 1.

they implemented a modified version of the velocity algorithm that uses the data binning obtained for the density algorithm.

We hypothesize that the increase in the results for the velocity and density methods below 8×10^3 particles come from the systematic bias described in the previous section. To test the general consistency of this hypothesis we correct the concentration values in the velocity and integrated mass methods by a factor of $1/(1+f_{\rm off})$, using the definition in Equation (11) and the parameters obtained from the data presented in Figure (1). The correction brings into perfect agreement the results between the velocity/density methods and the new algorithm.

6. CONCLUSIONS

In this letter we used bootstrapping to quantify the biases on concentration estimates. We found that methods commonly used in the literature present a large bias by overestimating concentrations by factors of 15%-20% for halos with 200 particles, or 7%-10% for halos with 500 particles. This procedure provides a robust technique to quantify the bias in concentration estimates with the advantage that it works without having to run new simulations.

These results motivated us to introduce a method with a robust performance at low particle numbers. This method is based on the integrated mass profile. The results of the algorithm on the same data sample showed a bias of $2\% - 3\%$ for halos with 200 particles and less than 1% for halos with 500 particles or more. To keep the bias of the velocity and density methods below 2% only halos with at least ~ 4000 particles should be taken into account.

Measuring the mass concentration relationship the three methods are in broad agreement within the statistical uncertainties, although there are some noticeable differences. For instance, the density method produces systematically higher concentrations by a factor of 15% compared to the velocity method. This systematic offset has been reported before with the same dataset (Prada et al. 2012) and with different simulations (Klypin et al. 2016) without any conclusive explanation for its origin. A second difference is that the velocity and integrated mass methods start to differ for masses below $10^{12}h^{-1}\mathrm{M}_\odot$ (~ 4000 particles). We found that correcting the mean concentration by the mean bias factor found through bootstrapping brings these two techniques into agreement.

These results show that using the integrated mass profile to find the Dark Matter halo concentration is a tool deserving deeper scrutiny. Further tests with larger simulated volumes, varying numerical resolution and different density profiles are the next natural step to explore the full potential of this new method.

The authors acknowledge the technical support from the new high-performance computing facility at Uniandes. JEF-R acknowledges financial support from Vicerrectoría de Investigaciones at Uniandes through a FAPA project. JCMC acknowledges financial support from "Estrategia de sostenibilidad 2014-2015, Universidad de Antioquia".

REFERENCES

Diemand, J., Kuhlen, M., Madau, P., Zemp, M., Moore, B., Potter, D., & Stadel, J. 2008, Nature, 454, 735
Duffy, A. R., Schaye, J., Kay, S. T., & Dalla Vecchia, C. 2008, MNRAS, 390, L64
Dutton, A. A., & Macciò, A. V. 2014, MNRAS, 441, 3359
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306

Klypin, A., Yepes, G., Gottlöber, S., Prada, F., & Heß, S. 2016, MNRAS, 457, 4340
Ludlow, A. D., Navarro, J. F., Angulo, R. E., Boylan-Kolchin, M., Springel, V., Frenk, C., & White, S. D. M. 2014, MNRAS, 441, 378
Macciò, A. V., Dutton, A. A., & van den Bosch, F. C. 2008, MNRAS, 391, 1940
More, S., Kravtsov, A. V., Dalal, N., & Gottlöber, S. 2011, ApJS, 195, 4
Muñoz-Cuartas, J. C., Macciò, A. V., Gottlöber, S., & Dutton, A. A. 2011, MNRAS, 411, 584
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493
Navarro, J. F., Ludlow, A., Springel, V., Wang, J., Vogelsberger, M., White, S. D. M., Jenkins, A., Frenk, C. S., & Helmi, A. 2010, MNRAS, 402, 21
Neto, A. F., Gao, L., Bett, P., Cole, S., Navarro, J. F., Frenk, C. S., White, S. D. M., Springel, V., & Jenkins, A. 2007, MNRAS, 381, 1450
Prada, F., Klypin, A. A., Cuesta, A. J., Betancort-Rijo, J. E., & Primack, J. 2012, MNRAS, 423, 3018
Riebe, K., Partl, A. M., Enke, H., Forero-Romero, J., Gottlöber, S., Klypin, A., Lemson, G., Prada, F., Primack, J. R., Steinmetz, M., & Turchaninov, V. 2013, Astronomische Nachrichten, 334, 691
Stadel, J. G. 2001, PhD thesis, UNIVERSITY OF WASHINGTON