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

AVOIDING BIASED ESTIMATES OF DARK MATTER HALO CONCENTRATIONS

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 ApJ

ABSTRACT

We present a new algorithm to estimate the concentration of N-body dark matter halos using 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. We compare the new algorithm against two other methods: maximum radial velocity measurements and radial particle binning. Tests on mock and N-body halos show that the accuracy of the new method to measure concentrations varies with halo resolution, outperforming the other two methods. We also esteimate the mass-concentration relationship on N-body data. We find that in the probed mass range ($10^{10}h^{-1}\mathrm{M}_{\odot} < M_h < 10^{14}h^{-1}\mathrm{M}_{\odot}$) the three methods give consistent results within the statistical uncertainties. There is a small deviation at low masses, $M < 10^{12}h^{-1}\mathrm{M}_{\odot}$, where the new method yields lower median concentration values by 25% compared to the velocity method; the difference is explained by the bias that the velocity method introduces. From these results we believe that this new method is a promising tool to probe the internal structure of dark matter halos.

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). If 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) one can also assume that this profile is universal. (Navarro et al. 2010). 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 10^3 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 paper we 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. Here we provide a detailed description of the method and a comparison against the

two traditional techniques by measuring its performance both on mock and N-body halos.

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 paper 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*

The new method 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. In this process we discard the particle at the center.

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. We then use an Affine Invariant Markov Chain Monte Carlo implemented in the python module emcee (Foreman-Mackey et al. 2013) to sample the likelihood function distribution defined by $\mathcal{L}(c) \propto \exp(-\chi^2(c)/2)$ where the $\chi^2(c)$ is written as

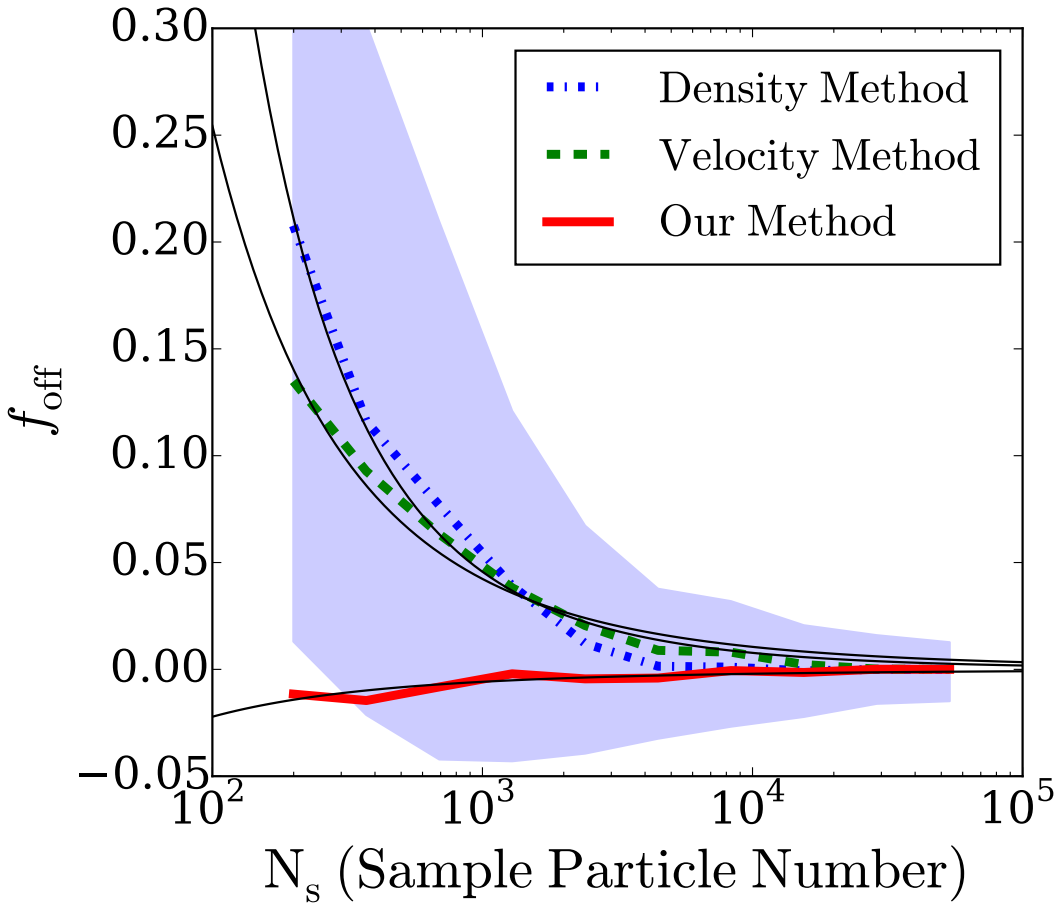

FIG. 1.— Average value of the fractional offset ($f_{\rm off}$Eq. 12) between the concentration in a halo at high resolution versus its downsampled versions. Each point is the result of 100 downsampling iterations. The velocity method noticeably overestimates the concentration up to a factor of 0.25, while the new method only underestimates the concentrations by up to a factor of 0.05.

$$\chi^2(c) = \sum_{i=1}^{N} [\log m_i - \log m(< x_i; c)]^2, \qquad (8)$$

where $m(< x_i; c)$ 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. From the χ^2 distribution we find the optimal value of the concentration and its associated uncertainty.

4. RESULTS

We apply the three methods mentioned above on two different halo samples. The first sample is composed by mock halos generated to have known concentration values in perfect spherical symmetry following an NFW profile. Using this sample we quantify the ability to recover the expected concentration as a function of particle number. The second sample comes from a publicly available N-body cosmological simulation to test the new method on more realistic and complex data to show the impact on the mass-concentration relationship.

4.1. *Retrieving known concentrations from a mock halo sample*

The method we use to generate the mock halos is based on the integrated mass profile. We start by fixing the desired concentration c and total number of particles N in the mock halo. With these values we define the mass element as $\delta m = 1/M$, corresponding to the mass of each particle such that the total mass is one. Then for each particle, $i = 1, \ldots, N$, we find the value of r_i such that

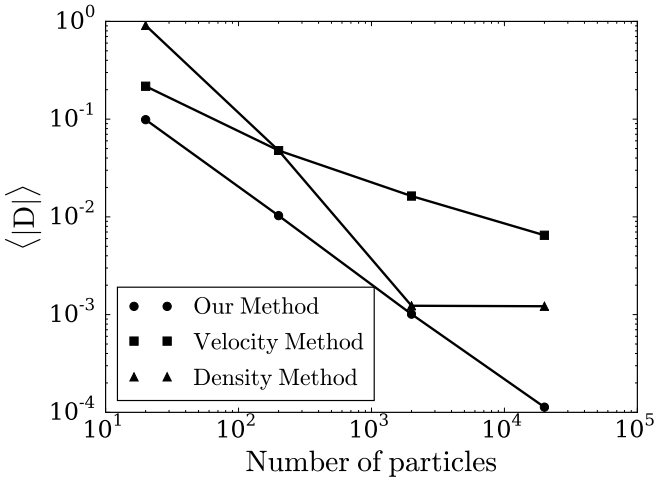

FIG. 2.— Average value of the relative error in the concentration estimate, $\langle |D| \rangle$, as a function of the particle number in the set of mock halos. Different symbols represent different methods. The new method provides the most accurate estimate at fixed particle number.

the difference

$$m(< r_i; c) - i \cdot \delta m \tag{9}$$

is zero using Ridders' method.

The value of r_i is the radius of the i-th particle of the mock halo. Then we generate random polar and azimuthal angles θ and ϕ for each particle to ensure spherical symmetry. Finally these three spherical coordinates are transformed into Cartesian coordinates $(r, \theta, \phi) \rightarrow (x, y, z)$.

We generate in total 400 mock halos split into four different groups of 100 halos each. The four groups differ in the total number of particles for their halos: 20, 200, 2000 and 20000. Inside each group the halos have random concentration values in the range $1 < c < 20$ with a uniform distribution. We find the concentration values for all these halos using the density, velocity and integrated mass methods described in the previous section.

We quantify the difference between the expected c_{in} and measured c_{out} values by

$$D = c_{out}/c_{in} - 1, \tag{10}$$

and

$$\langle |D| \rangle = \frac{1}{|\mathcal{H}|} \sum_{\mathcal{H}} |D|, \tag{11}$$

where $\mathcal{H}$ corresponds to a set of mock halos, and $|\mathcal{H}|$ is the number of haloes in $\mathcal{H}$. A large discrepancy between the estimated and input concentrations are quantified by a large $\langle |D| \rangle$.

Figure 2 shows the behaviour of $\langle |D| \rangle$ as a function of halo particle number for the three different methods.

At fixed particle numbers the new method always shows the lowest $\langle |D| \rangle$ values compared to the other two. Its accuracy is on the order of 10% for 20 particles in the halo, decreasing to 0.01% for halos with 20000 particles. The dependence of $\langle |D| \rangle$ with the particle number N goes approximately as $\langle |D| \rangle \propto N^{-1}$.

The method based on the maximum of the circular velocity shows a similar behaviour with $\langle |D| \rangle \propto N^{-1/3}$. Its accuracy is $2-5$ times lower than in the new method, on the order of 20% for 20 particle halos and 0.5% for 20000 particle halos. The method based on the direct density fit shows the lowest accuracy for a low particle number. As it is discussed in (Muñoz-Cuartas et al. 2011), density binning for halos with particle numbers below ~ 200 leads to a biased estimation of the mass density profile, and therefore to a biased estimation of the concentration parameter. This behaviour is evident in the large values of $\langle |D| \rangle$ for halos with number of particles below ~ 200 and an intermediate accuracy between the other two methods for a high particle number.

We also find that the bias from different concentrations at fixed particle number is negligible with respect to the effects induced by different particle numbers.

4.2. *Measuring the concentration in halos from a cosmological simulation*

The results presented in the previous section show that in an idealized setup (pure NFW density profiles, perfect spherical halos in total isolation) the new method performs better than the other two methods commonly used in the literature.

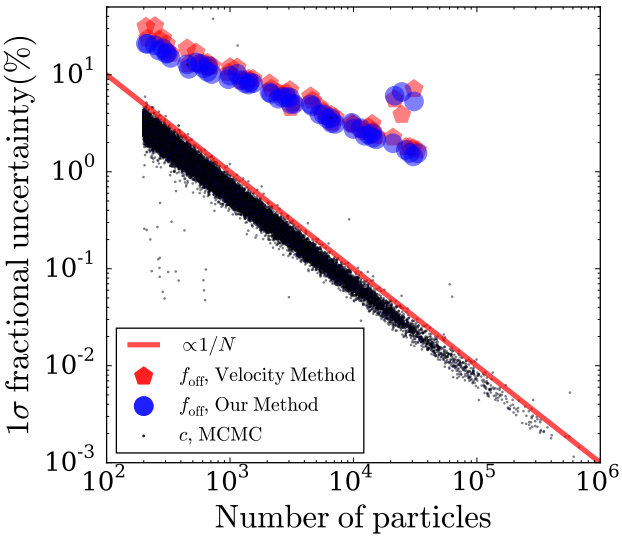

FIG. 3.— Uncertainties on different concentration estimates. The large symbols show the standard deviation of the fractional offset ($f_{\rm off}$ as defined in Eq. 12) from the downsampling experiments. The small dots correspond to the 1σ uncertainty on the concentration for each halo as estimated by the MCMC algorithm used for the new method. The solid lines shows the $1/N$ tendency, which also corresponds to the particle number dependence of the $\langle |D| \rangle$ statistic (i.e. the accuracy to retrieve a known concentration in mock halos) of the new method. Uncertainties due to particle number fluctuations are always larger than the statistical uncertainties from the fitting process.

However, simulated dark matter halos in a cosmological context are complex. They are not isolated, many of them experience mergers that take them out of the equilibrium. Halos may have also plenty of substructure. All these effects disturb the mass density profile, taking it away from the ideal NFW profile. In order to explore the performance of the new method in this setting, we use it to estimate the concentration parameter in dark matter halos obtained from a cosmological N-body simulation and compare it with the other two classical methods.

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 data is publicly available at http://www.cosmosim.org/. More 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 $4 \times 10^{10} \leq M_{\rm FoF}/h^{-1}{\rm M}_\odot \leq 10^{14}$. 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

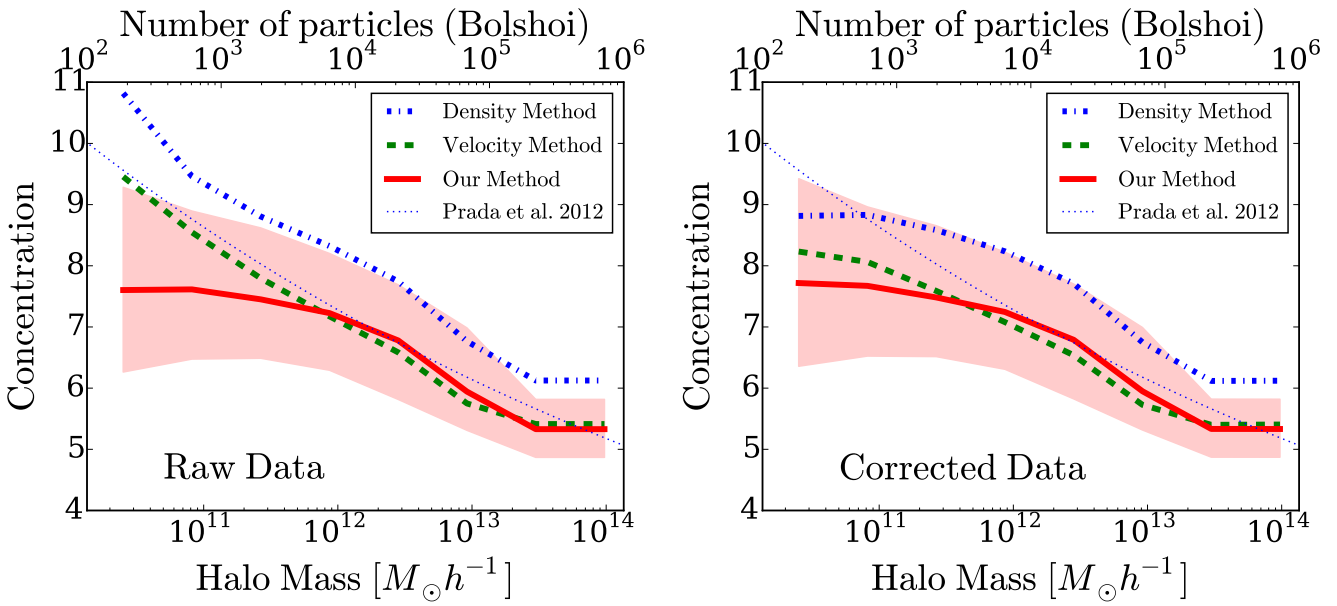

FIG. 4.— Mass-concentration relationship for the three different methods used on the same cosmological N-body data. All halos are used regardless of their relaxation state. The lines correspond to the median concentration values in each mass bin. The shaded region presents 10 and 90 per cent spread. For clarity we only show the spread for the concentrations estimated using the new method. The other two methods have a similar spread. 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. The corrections depends on the number of particles following the results of the downsampling experiments presented in Figure 1.

$(2.6 \times 10^{10} h^{-1} \mathrm{M}_\odot)$.

4.2.1. *Convergence Test using Downsampling*

We perform first a convergence test on the N-body data to find possible biases in the concentration estimates as a function of particle number. We select 8 halos from the simulation that have at least 3×10^5 particles. We then randomly downsample the total number of particles in each halo by factors of $10, 20, 45, 100, 215, 460, 1000$ to measure its concentration with both the integrated mass and the velocity method. For each downsampling factor we repeat the operation 100 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

$$f_{\rm off} = c_N / c_{N_{max}} - 1, \tag{12}$$

to account for the offset between the concentration at a given downsampled particle number c_N and the baseline $c_{N_{max}}$.

Figure 1 summarizes our results. The plot shows the average value of $f_{\rm off}$ as a function of the particle number. As expected for large enough sample particle numbers, $N_s > 5 \times 10^3$, the results of the two algorithms coincide and $\langle f_{\rm off} \rangle \sim 0$. For lower number of particles, the results of the two algorithms deviate from the expected concentration value. The velocity method overestimates by a factor $\langle f_{\rm off} \rangle \sim 0.25$ the concentration values for the lowest sample particle numbers, $N_s \sim 100$. Around the same sampling scale, the new algorithm shows a more stable behaviour with an underestimate by a factor of $\langle f_{\rm off} \rangle \sim 0.02$. This shows that in the more complex setup of a cosmological simulation the new algorithm continues to be stable below the 5% level for low particle numbers.

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

$$f_{\rm off} = \frac{A}{(1 + \log_{10} N_s)^B}. \tag{13}$$

for the dataset shown in the panel we have $A = 3385 \pm 2065$, $B = 7.99 \pm 0.48$ for the velocity method and $A = -110 \pm 253$, $B = 7.11 \pm 1.80$ for the new method, where 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.

Figure 3 summarizes different uncertainty results. Large symbols show the standard deviation on $f_{\rm off}$ from the downsampling experiment. Small dots show the uncertainty on each individual concentration value as estimated by the MCMC algorithm. The line shows the $\propto N^{-1}$ locus. This figure demonstrates that the statistical uncertainties are at least one order of magnitude smaller than the uncertainty associated to downsampling experiment.

4.2.2. *Mass-Concentration Relationship*

Figure 4 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 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. However, to the best of our knowledge, a systematic and mod-

ern study of the mass-concentration relationship spanning a wide range in mass, using a fixed dataset but different fitting methods has not been presented in the literature.

We then notice that the new method closely follows the results of the velocity method at higher masses, $M_h > 10^{12}h^{-1}\mathrm{M}_\odot$ or equivalently for $> 5 \times 10^3$ particles. For masses below that threshold the new method saturates at a concentration value of ~ 8 while the velocity method continues to give higher concentration values.

We hypothesize that the increase in the results for the velocity method below 5×10^4 particles come from the systematic bias described in the previous subsection. To test the general consistency of this hypothesis we correct the concentration values in the velocity and integrated mass methods by a factor of $1/(1 + f_{\mathrm{off}})$, using the definition in Equation (12) and the parameters obtained from the data presented in Figure (1). The correction brings into perfect agreement the results between the velocity and the integrated mass method. This suggests that the concentration at lower masses might very well reach a plateau in the halo mass range $10^{10} < M_h/h^{-1}\mathrm{M}_\odot < 10^{12}$.

5. CONCLUSIONS

In this paper we presented a new method to estimate the concentration of dark matter halos in N-body simulations. We tested the method on mock halo data to study the impact of total number of particles and input concentration on the retrieved values. We compared these results against two other methods commonly used in the literature to estimate concentrations. For all methods, the accuracy in retrieving the input concentration increases with the number of particles as summarized in Figure 2. The new method systematically outperforms the other two.

We also applied the method to halos extracted from a cosmological N-body simulation. Using this data we showed that the new method is robust with respect to a decreasing number of particles in the halos. On the other hand we showed that the velocity method has a systematic bias that overestimates the concentration for halos with a low particle number up to a factor of 25%.

We provide for the first time in the literature a direct way to measure the bias in the estimated halo concentration at low particle numbers and a way to use it to make corrections on the estimates. In Section 4.2 we presented a simple method that can be used on any dataset to estimate the errors and correct them when estimating the concentration at low particle numbers.

We also used the N-body simulation to estimate the impact on the mass concentration relationship. Although the three methods are in broad agreement within their uncertainties there are some noticeable differences. The first difference is that the density method produces systematically higher concentrations by a factor of 15% compared to the velocity method. This systematic offset has been reported in different datasets and techniques to measure concentrations, without any conclusive explanation for its origin (Klypin et al. 2016). The second difference is that the velocity and integrated mass methods start to differ for masses below $10^{12}h^{-1}\mathrm{M}_\odot$. We found that a correction of the results by the bias factor found using the same N-body data can bring these two methods into agreement.

These are encouraging results. They 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

Duffy, A. R., Schaye, J., Kay, S. T., & Dalla Vecchia, C. 2008, MNRAS, 390, L64

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