

# An accurate measure of the size of dark matter haloes using the size of galaxies

Claudio Dalla Vecchia<sup>1,2,\*</sup> and Ignacio Trujillo<sup>1,2</sup>

<sup>1</sup> Instituto de Astrofísica de Canarias, C/ Vía Láctea s/n, 38205 La Laguna, Spain

<sup>2</sup> Departamento de Astrofísica, Universidad de La Laguna, Av. del Astrofísico Francisco Sánchez s/n, 38206 La Laguna, Spain

Received 28 July 2025 / Accepted 17 October 2025

## ABSTRACT

The physically motivated definition of galaxy size proposed recently, linked to the farther location of the in situ star formation, considerably reduces the scatter of the galaxy mass–size relation and provides a viable method to infer the galaxy stellar mass from its size. We provide a similar relation correlating the size of galaxies with the size of their dark matter haloes by leveraging the small scatter of the aforementioned relation. We analysed the simulated galaxies of the two main cosmological volumes of the EAGLE simulations and computed the size of the galaxies and their mass when mimicking the observational analysis. For central galaxies, we computed the relation between galaxy size and halo size. We show that the simulated galaxies reproduce the observed stellar mass–size relation’s normalisation and slope. The scatter of this relation, 0.06 dex, matches the intrinsic scatter measured in observation. We then computed the correlation between galaxy size and halo size and found that the relation is steeper than when using the half-mass radius as a measure of size, with the scatter (0.1 dex) being a factor of two smaller than the observed relation. As well, the galaxy-to-halo mass relation derived from the simulations provides a factor of two better scatter than the observed scatter. This opens the possibility of measuring the size of dark matter haloes with greater accuracy (less than 50%, i.e. around six times better than using the effective radius) by using only deep imaging data.

**Key words.** galaxies: evolution – galaxies: fundamental parameters – galaxies: halos – galaxies: structure

## 1. Introduction

A physically motivated definition of galaxy size, given by the farthest radial location where gas could efficiently collapse and turn into stars, was proposed by Trujillo et al. (2020). As a proxy for measuring this position, these authors suggest using the radial location of a stellar mass density contour at  $1 M_{\odot} \text{pc}^{-2}$ , denoted  $R_1$ . This measure is consistent with the gas density threshold for star formation derived theoretically by Schaye (2004). The use of  $1 M_{\odot} \text{pc}^{-2}$  was calibrated using deep observations of Milky Way-sized galaxies at  $z \sim 0$ , and it provides a more representative boundary for galaxies than traditional metrics such as the effective radius,  $R_e$  (Sérsic 1968). Using this definition over a wide range of stellar masses (from  $10^7$  to  $10^{12} M_{\odot}$ ) significantly reduces the scatter in the stellar mass–size relation to about 0.06 dex, thus outperforming conventional measures such as  $R_e$  and the Holmberg radius,  $R_H$  (Holmberg 1958).

Even though  $R_1$  is only a proxy to characterise the physical radius where in situ star formation is or has been taking place, the enormous reduction in the scatter of the stellar mass–size relation underscores the robustness of  $R_1$  as a galaxy size metric. The results of using  $R_1$  also reveal clear trends within the stellar mass–size relation. Galaxies with stellar masses between  $10^7$  and  $10^{11} M_{\odot}$  follow a consistent power-law relationship with a slope of around 1/3. Across the range of galaxy stellar masses studied, the scatter is consistently small ( $\sim 0.06$  dex), providing an accurate measure of the galaxy’s stellar mass by knowing only its physical size.

Observed disc galaxies follow the relation  $R_1 \propto M_{\star}^{1/3}$  over several orders of magnitude in stellar mass. A similar relation has been proposed for neutral hydrogen by Broeils & Rhee (1997). The mass–size relation for neutral hydrogen is  $M_{\text{HI}} \propto D_{\text{HI}}^2$ , where  $D_{\text{HI}}$  is the diameter of the neutral hydrogen disc of a spiral galaxy. Wang et al. (2016), using observations of 500 galaxies, showed that this correlation holds irrespective of galaxy luminosity, fraction of neutral gas, and morphology, with a scatter of 0.06 dex. This points to the non-trivial interplay of gas accretion, star formation efficiency, and stellar feedback in maintaining the two mass–size relations over several orders of magnitude in stellar mass.

Chamba et al. (2022) tested the validity of the hypothesis of using  $R_1$  as an indicator of the radius up to which in situ star formation has occurred. To do this, they identified the position of the stellar edge of galaxies,  $R_{\text{edge}}$ , as the outermost cut-off in their stellar mass radial profile using deep multi-band optical imaging; verified that it corresponds to a change in the radial colour profile and surface brightness profile; and measured the stellar mass surface density it corresponds to. Chamba et al. (2022) find that  $R_{\text{edge}} \sim R_1$  for galaxies with masses similar to the Milky Way (see also Golini et al. 2025). For lower-mass galaxies (dwarfs) or more massive galaxies (ellipticals), the surface mass density of the galaxy edge is slightly different from  $1 M_{\odot} \text{pc}^{-2}$ . It is smaller by a factor of about two for low-mass galaxies and larger by a factor of about three for massive elliptical galaxies. Be that as it may, within a range of five orders of magnitude in stellar mass, the approximation of using  $R_1$  as an indicator of galaxy size is quite good, and its practical implementation is straightforward. One of the main results of Trujillo et al. (2020) is that as long as the definition of galaxy size includes most of the

\* Corresponding author: [claudio.dalla.vecchia@iac.es](mailto:claudio.dalla.vecchia@iac.es)

stellar mass, the mass–size relation should have a small scatter (e.g. Miller et al. 2019). Sánchez Almeida (2020) has shown that for a fixed stellar mass, stellar profiles with different Sérsic indexes all cross at a similar stellar mass surface density, and they concluded that this explains the small scatter when using  $R_1$  as the size.

In a subsequent paper, Buitrago & Trujillo (2024) studied the evolution of the stellar mass–size relation with redshift for a sample of Milky Way-like galaxies, directly measuring  $R_{\text{edge}}$  (instead of  $R_1$ ) of galaxies up to  $z = 1$ . They showed that the edge is located at higher stellar surface densities at a higher redshift and that disc galaxies of the same stellar mass are smaller with increasing redshift, with their size inversely proportional to  $(1 + z)$ . This shows that using  $R_1$  as a proxy for galaxy size is a valid option only at  $z \sim 0$ . The dependence of galaxy size on redshift reveals the connection between the growth of the dark matter halo and the galaxy disc, as the former grows in size with the same redshift scaling (Navarro et al. 1997).

In a companion paper, Arjona-Gálvez et al. (2025) show that when using  $R_1$  as the definition of size, galaxies extracted from a variety of zoom-in simulations performed with different models for galaxy formation all share the same mass–size relation. When studying its evolution in simulations, the relation does not change with redshift; rather, the galaxies move on the relations as they grow in mass and size. The difference between this and the observations of Buitrago & Trujillo (2024) is that the stellar surface density threshold is kept fixed with redshift in the analysis, whilst observations show that the stellar surface density at which disc truncation is observed is larger at a higher redshift. What the simulations have in common is that their star formation models all try to reproduce the observed (local) Kennicutt-Schmidt law and its surface density threshold for star formation (Kennicutt 1998; Martin & Kennicutt 2001; Bigiel et al. 2008), which is assumed to hold at any redshift. Combining this with a fixed stellar surface density may result in galaxies of the same mass having the same  $R_1$  at any redshift.

Given the close connection between the new definition of galaxy size and the physical process of in situ star formation, as well as the observation that the size thus defined scales with the expected evolution of dark matter halo size, in this paper we explore the correlation between galaxy size,  $R_1$ , and dark matter halo size,  $R_{200}$ , defined as the radius of the sphere with an average mass density 200 times the critical density of the Universe. We find that the simulations reproduce the observed relation between stellar mass and size remarkably well. Moreover, they predict that the relation between  $R_1$  and  $R_{200}$  is extremely tight so that when given  $R_1$ , one can estimate  $R_{200}$  with an uncertainty of less than 50%. This is a factor of more than two better than the predicted relation between the effective radius and  $R_{200}$  (Kravtsov 2013) and the predicted stellar-to-halo mass relation for observed galaxies when the galaxy size is given by the radius containing 80% of the light (Mowla et al. 2019).

In this work, we analyse the two main cosmological hydrodynamical simulations of the EAGLE project (Schaye et al. 2015; Crain et al. 2015) for the stellar mass range  $\approx 10^{7.5} - 10^{12} M_{\odot}$ . We describe the data used in the simulations and the analysis in the following section. In Section 3, we compare the results from the simulations with observations. In Section 4, we link the size of the galaxies to the size of their halo, showing that this correlation can predict the halo mass and size with a higher accuracy when compared to using the half-mass radius or  $R_e$ . A discussion and our conclusions are presented in Section 6.

## 2. Analysis of the simulations

### 2.1. Simulations and simulation data

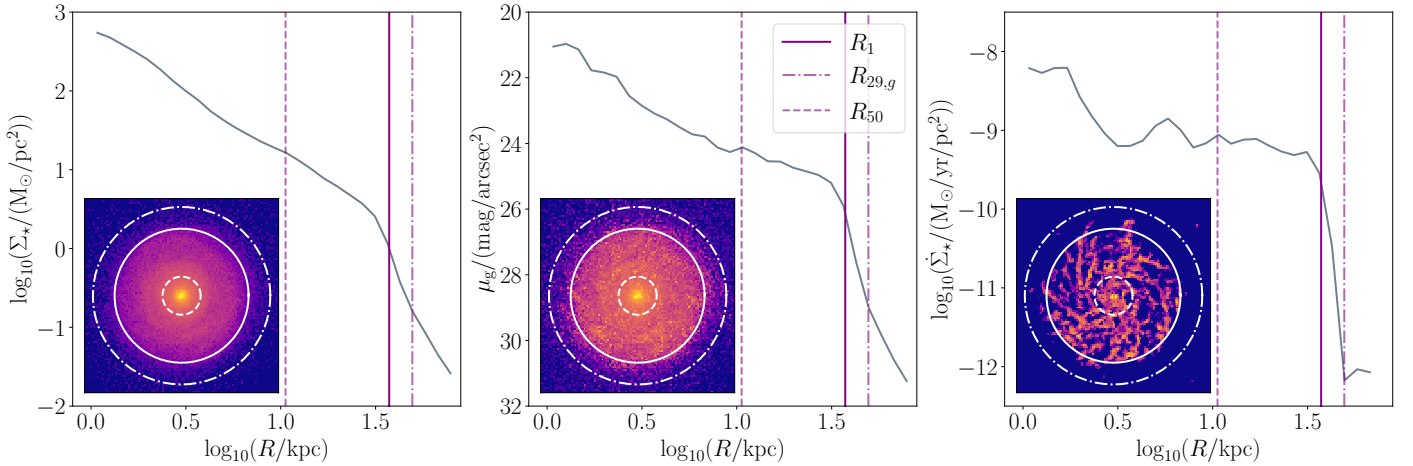
The simulations employed in this work were performed with the EAGLE model for galaxy formation and evolution. For more information on the EAGLE model and its calibration against global relations of the observed galaxy population (galaxy stellar mass function, stellar mass–size relation, and stellar mass–black hole mass relation), we refer to Schaye et al. (2015) and Crain et al. (2015). For more details on the numerical algorithms describing the photo-ionisation equilibrium cooling, star formation, stellar evolution, stellar feedback, black hole growth, and feedback we refer to Schaye & Dalla Vecchia (2008), Wiersma et al. (2009a,b), Dalla Vecchia & Schaye (2012), Rosas-Guevara et al. (2015), respectively. A description of the benefits of the hydrodynamic scheme used in the EAGLE model can be found in Schaller et al. (2015a) (see also Durier & Dalla Vecchia 2012). For the analysis presented here, the raw particle data at  $z = 0$  and the public database<sup>1</sup> were used (McAlpine et al. 2016; The EAGLE Team 2017).

The EAGLE model implements the sub-grid prescription for star formation of Schaye & Dalla Vecchia (2008). Briefly, the model is based on the conversion of the Kennicutt-Schmidt empirical law (Kennicutt 1998) that correlates the surface rate of star formation to the surface density of gas, to a volumetric law that expresses the rate of star formation as a function of gas pressure. The density threshold above which the gas is able to form stars is assumed to vary with the metallicity of the gas (Schaye 2004). Therefore, star formation in EAGLE occurs from the threshold density set by the metallicity of the gas and beyond. This threshold decreases with increasing metallicity. Details on the numerical implementation of the dependence on metallicity of the density threshold for star formation can be found in Schaye et al. (2015). We emphasise that the theoretical predictions of Schaye (2004) are used as a starting hypothesis for the definition of galaxy size by Trujillo et al. (2020).

The post-processing of the simulations was performed with the friends-of-friends algorithm (e.g. Einasto et al. 1984; Davis et al. 1985) and the SubFind algorithms (Springel et al. 2001; Dolag et al. 2009), which define the distribution of overdense regions in the simulated cosmological volumes and their substructure. We further post-processed the  $z = 0$  outputs of the EAGLE simulations to compute dust-attenuated stellar absorption spectra that we convolved with filter response curves and integrated in the wavelength bands of the Sloan Digital Sky Survey (Abazajian et al. 2003). We employed the EMILES stellar spectra library (Vazdekis et al. 2016) and the same methodology described in Negri et al. (2022). More details are given in that work.

We analysed the  $z = 0$  output of two EAGLE simulations. The reference EAGLE volume of  $(100 \text{ Mpc})^3$  – the ‘reference’ simulation – initially contained  $2 \times 1504^3$  dark matter and gas particles, had a physical spatial resolution of  $\epsilon = 0.7 \text{ kpc}$ , had particle masses of  $m_{\text{gas}} = 1.81 \times 10^6 M_{\odot}$  and  $m_{\text{dm}} = 9.70 \times 10^6 M_{\odot}$ , for gas and dark matter, respectively. The recalibrated EAGLE volume of  $(25 \text{ Mpc})^3$  – the ‘recalibrated’ simulation – initially contained  $2 \times 752^3$  dark matter and gas particles, had a physical spatial resolution of  $\epsilon = 0.35 \text{ kpc}$ , and had particle masses of  $m_{\text{gas}} = 2.26 \times 10^5 M_{\odot}$  and  $m_{\text{dm}} = 1.21 \times 10^6 M_{\odot}$ . For the recalibrated simulation, the EAGLE galaxy formation model was calibrated in order to match the same global relations

<sup>1</sup> <http://icc.dur.ac.uk/Eagle/>



**Fig. 1.** From left to right, projected density maps and radial profiles of stellar mass, stellar luminosity, and gas star formation rate for a simulated spiral galaxy of stellar mass  $3.8 \times 10^{10} M_{\odot}$  selected from the recalibrated simulation. The vertical dashed, solid, and dot-dashed lines mark the positions of  $R_{50}$ ,  $R_1$ , and  $R_{29,g}$ , respectively. For this star forming galaxy,  $R_1$  is a better measure of the size of the galaxy for its proximity to the edge (truncation radius) of the disc. The edge of the disc coincides with a net drop in the star formation rate surface density, the starting hypothesis for the new definition of galaxy size.

at  $z = 0$  reproduced within the largest volume (Crain et al. 2015). The recalibrated simulation was used to extend the analysis down to galaxy stellar masses  $\approx 2 \times 10^7 M_{\odot}$  (about 1 dex smaller stellar masses than in the reference simulation), after restricting the minimum stellar mass to about 100 stellar particles. Finally, the initial conditions for the aforementioned simulations were generated with the cosmological parameters inferred by the Planck Collaboration XVI (2014):  $\Omega_0 = 0.307$ ,  $\Omega_{\Lambda} = 0.693$ ,  $\Omega_b = 0.04825$ ,  $h = 0.6777$ ,  $\sigma_8 = 0.8288$ , and  $n_s = 0.9611$ .

## 2.2. Determination of galaxy size and stellar mass

We describe here how we measured the size of the simulated galaxies. The  $R_1$  radius has been defined as the projected distance from the centre of the galaxy to the location where the stellar surface density reaches  $1 M_{\odot} \text{pc}^{-2}$ . We determined  $R_1$  by first rotating the galaxy face-on using the angular momentum vector of the stellar component, thus avoiding the uncertainties introduced by de-projecting the stellar distribution. Because of the relatively low spatial and mass resolution of the simulations, we did not calculate maps of the stellar mass surface density as usually done in observational work. Indeed, given the stellar particle mass of about  $10^6 M_{\odot}$  for the reference simulation, at the threshold surface density of  $1 M_{\odot} \text{pc}^{-2}$ , the particle number density would be about  $1 \text{kpc}^{-2}$ , making any projected surface density very noisy. We therefore opted to use the stellar surface density profile derived by averaging the projected stellar mass in circular annuli. We set the centre of the galaxy at the position of the minimum of the gravitational potential of the halo, as provided by the sub-halo finder. This is a good estimate of the centre of the galaxy, as the offset of the stellar component with respect to the dark matter halo is less than the spatial resolution of the simulation for the majority of galaxies (Schaller et al. 2015b).  $R_1$  was then derived by interpolating the surface density profile at the position of the threshold surface density.

We determined the galaxy stellar mass as the projected stellar mass within a radius of constant surface brightness. This was done in order to consistently include the same bias introduced by observers in determining the relation. Indeed, there is no direct correlation between the measurement of  $R_1$  and  $M_{\star}$  in Trujillo et al. (2020).  $R_1$  was derived by converting the sur-

face brightness into surface stellar mass density, while  $M_{\star}$  was derived by integrating the stellar mass above within the surface brightness threshold of  $29 \text{ mag arcsec}^{-2}$  in the  $g$  band. This corresponds to a stellar surface density threshold lower than  $1 M_{\odot} \text{pc}^{-2}$ . In order to calculate the position of the threshold in surface brightness, we converted the stellar mass surface density by inverting Equation (1) of Bakos et al. (2008):

$$\log_{10} \Sigma_{\star} = \log_{10}(M/L)_{\lambda} - 0.4(\mu_{\lambda} - m_{\text{abs},\odot,\lambda}) + 8.629, \quad (1)$$

where  $\Sigma_{\star}$  is the stellar surface density in  $M_{\odot} \text{pc}^{-2}$ ,  $M/L$  is the mass-to-light ratio at the wavelength  $\lambda$ , and  $m_{\text{abs},\odot,\lambda}$  is the absolute magnitude of the Sun at the same wavelength. We calculated  $M/L$  with the Sloan Digital Sky Survey  $g$  band magnitude that we computed in the post-processing of the simulations for each stellar particle. From here on, we denote the projected stellar mass within  $R_{29,g}$  as  $M_{\star}$ .

For each galaxy, we also calculated the stellar half-mass radius,  $R_{50}$ , as the radius containing half the projected stellar mass. This radius is a good proxy for  $R_e$ . The computation of  $R_1$ ,  $R_{50}$ , and  $R_{29,g}$  was performed for all simulated galaxies with a mass larger than 100 initial gas particle masses ( $2.26 \times 10^7$  and  $1.81 \times 10^8 M_{\odot}$  for the recalibrated and reference simulations, respectively). However, the algorithm failed to converge for a fraction of galaxies that are barely resolved or have a disturbed morphology. We marked them and did not use them in the analysis. The final selection of the samples of galaxies used in the analysis is described in Appendix A.

We show in Figure 1 one example of a simulated spiral galaxy with a stellar mass of  $3.8 \times 10^{10} M_{\odot}$ . From left to right, we show the radial profiles for the projected stellar surface density, the surface brightness, and the star formation surface density. The insets are the corresponding maps and have an arbitrary logarithmic colour scale. The circles in the maps and the vertical lines in the profiles plots indicate the values of the half-mass radius,  $R_{50}$  (dashed line);  $R_{29,g}$  (dash-dotted line); and the galaxy size,  $R_1$  (solid line). For the galaxy shown in Figure 1,  $R_1$  is an excellent proxy for measuring the edge of the star forming disc and thus a good approximation of the physically motivated definition of galaxy size. The stellar disc truncation is close to  $R_1$ , and the gas surface density steeply declines beyond  $R_1$  and so does the star formation surface density. If the galaxy ceased to

form stars for the rest of its evolution and without contribution to its stellar mass from mergers, the edge of the in situ star formation would still be marked by  $R_1$ .

### 3. Results

The results of the study of the galaxy mass–size relation in simulations is presented here. We show in Figure 2 the correlation between  $R_1$  and galaxy stellar mass for all galaxies in the sample, central and satellite<sup>2</sup>. The sample was selected as described in Appendix A. We emphasise that even when using the same definition of  $R_1$ , the analysis of the simulated data differs substantially from that of observational data. The large dots in the figure are the median values of  $R_1$  measured as described in Section 2.2, while the small dots are individual galaxies where the stellar mass bin contains fewer than ten galaxies. The shaded area represents the 16% and 84% quantiles in each bin. The simulated relation follows the trend of observations, with large slopes for stellar masses below  $10^{8.5} M_\odot$  and above  $10^{10.5} M_\odot$  with respect to the mass range in between. As mentioned in the introduction, Trujillo et al. (2020) explains that this is due to the fixed surface density threshold employed in their analysis. The same trend is shown in the figure of Appendix B, where we show the calculated relation for the samples of central<sup>3</sup> and satellite galaxies.

For comparison, in Figure 2 we also plot the median distribution of projected stellar half-mass radii,  $R_{50}$ , with its 16% and 84% quantiles (purple solid line and shaded area). The scatter around the median is more than double of that for  $R_1$ . Moreover, the correlation between  $R_{50}$  and stellar mass is very weak in the mass range  $10^{8.5} < M_\star/M_\odot < 10^{10.5}$ , where the mass of galaxies with similar  $R_{50} \approx 2\text{--}3$  kpc spans over two orders of magnitude. The weak correlation and its large scatter provide a relation with little information on the properties of galaxies. However, for a given stellar mass, the scatter around  $R_{50}$  can provide insight into how gas has been redistributed by feedback processes within a star forming galaxy (e.g. Crain et al. 2015; Rohr et al. 2022).

In order to provide an analytic fit to the relation, we chose the extended double power-law function defined as follows:

$$y = A \left( \frac{x}{x_s} \right)^\alpha \left( 1 + \left( \frac{x}{x_s} \right)^s \right)^{(\beta-\alpha)/s}, \quad (2)$$

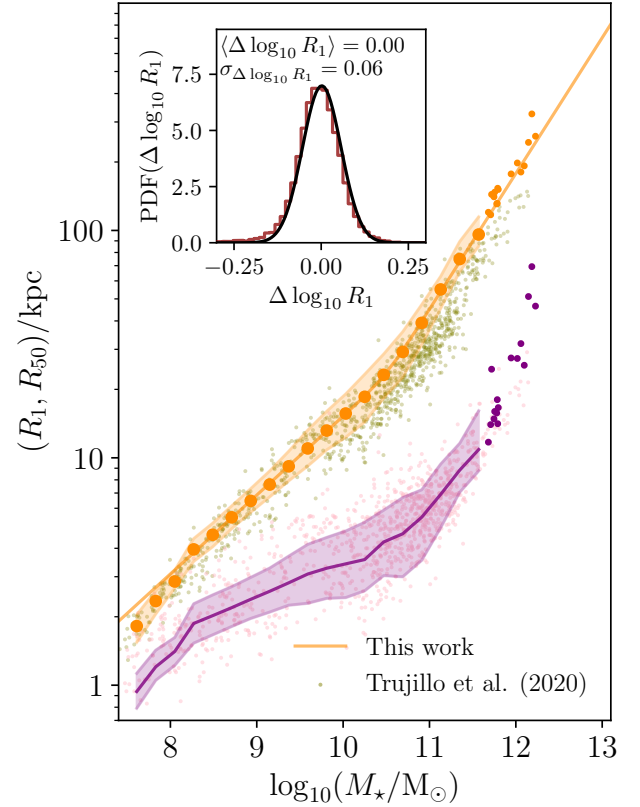
with free parameters ( $A, x_s, \alpha, \beta, s$ ). Parameter  $A$  sets the normalisation,  $x_s$  is the scale at which the slope transitions between the asymptotic values  $\alpha$  and  $\beta$ , and  $s$  is the strength of this transition. Here, we assumed  $(x, y, x_s) \equiv (M_\star, R_1, M_s)$ . We then fit the median values for  $M_\star > 10^{8.5} M_\odot$ , excluding the low-mass end drop of  $R_1$  with decreasing mass<sup>4</sup>. The  $\chi$ -square minimisation gives the following set of parameters for the best-fit function:  $(A, M_s, \alpha, \beta, s) \approx (22.8 \text{ kpc}, 3.21 \times 10^{10} M_\odot, 0.35, 0.60, 4.25)$ . The asymptotic slopes,  $\alpha = 0.35$  and  $\beta = 0.60$ , are remarkably close to the observed values given by Trujillo et al. (2020) (and the high-mass slope given by Mowla et al. 2019).

In Appendix B, we provide similar fits for central and satellite galaxies and a table with the double power-law fit parameters and their errors for the three sub-samples of galaxies. As a check

<sup>2</sup> Trujillo et al. (2020) did not discriminate between central and satellite galaxies.

<sup>3</sup> A central galaxy is defined by SUBFIND as the galaxy sitting at the potential minimum of its friends-of-friends halo.

<sup>4</sup> We refer the reader interested in the low-mass regime to Arjona-Gálvez et al. (2025) for a similar study employing high-resolution simulations.



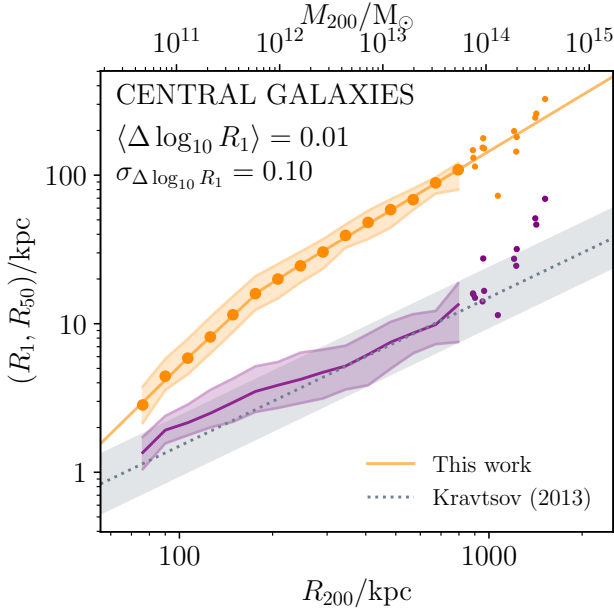
**Fig. 2.** Correlation between  $R_1$  and stellar mass for all galaxies in the sample (orange symbols). The median values in each bin are plotted together with the 16% and 84% quantiles (shaded area). For bins with fewer than ten galaxies, single galaxies are plotted as small dots. The orange solid line depicts the double power-law fit for  $M_\star > 10^{8.6} M_\odot$  (extrapolated below that mass). We show the relation between  $R_{50}$  and stellar mass for all galaxies in the sample in purple. For visual comparison, the galaxies in the sample of Trujillo et al. (2020) are represented with small dots, both  $R_1$  and  $R_{50}$ . The distribution of the residuals around the best-fit linear relation for all galaxies is shown in the inset. The distribution has been fitted with a Gaussian (solid line) with a dispersion of  $\sigma_{\Delta \log_{10} R_1} = 0.06$  dex.

of the quality of the fit to the binned values, we also calculated the best-fit linear function to individual galaxies within the mass range  $10^{8.5} \leq M_\star/M_\odot \leq 10^{10.5}$  (dashed line), which yielded the correlation

$$\log_{10} \left( \frac{R_1}{\text{kpc}} \right) = (0.343 \pm 0.001) \log_{10} \left( \frac{M_\star}{M_\odot} \right) - (2.258 \pm 0.009), \quad (3)$$

which successfully matches the asymptotic slope of the fit to the median.

The slope of the size-mass relation increases to a larger value above the mass  $M_s \approx 10^{10.5} M_\odot$ . In numerical modelling, this is a characteristic mass that marks the transition between supernova and black hole feedback dominance for several numerical models (e.g. Bower et al. 2017; McAlpine et al. 2017; Pillepich et al. 2018). Observations show a similar characteristic mass. Trujillo et al. (2020) give a value close to  $10^{10.8} M_\odot$ , whilst Mowla et al. (2019) quote a mass of  $10^{10.2} M_\odot$ . In both cases, the characteristic stellar mass marks the morphological transition from disc to elliptical galaxies, which is, from theoretical studies, closely connected to the transition between stellar and



**Fig. 3.** Correlation between  $R_1$  and  $R_{200}$  for central galaxies. Median (dots and solid lines) and 16% and 84% quantiles (shaded areas) are shown. We quote in the legend the average of the scatter for the entire sample of central galaxies and with respect to the analytic fit as well as the dispersion around the mean. The purple line and shaded area are the relation between  $R_{50}$  and  $R_{200}$  and its 16% and 84% quantiles. The dotted line is the linear correlation proposed by Kravtsov (2013) for early-type galaxies,  $R_{50} = 0.015 R_{200}$ , with a scatter of 0.2 dex (shaded area). The top  $x$ -axis gives the corresponding halo mass,  $M_{200}$ .

black hole feedback and to the transition in the process of stellar mass growth from gas accretion to dry mergers with increasing halo mass.

Finally, we show in the inset in Figure 2 the distribution of residuals around the best-fit stellar mass–size relation and in the same mass range used for the calculation of the best-fit analytic function for the full sample of galaxies. The distribution is well represented by a Gaussian with a central value of  $\langle \Delta \log_{10} R_1 \rangle = 0.00$  dex and a dispersion of  $\sigma_{\Delta \log_{10} R_1} = 0.06$  dex, in agreement with the intrinsic scatter measured by Trujillo et al. (2020). We show in the next section that this very small scatter yields a tighter correlation between galaxy and halo size than what was inferred by Kravtsov (2013) and a smaller scatter in the stellar-to-halo mass relation than that observed by Mowla et al. (2019) with their definition of galaxy size.

#### 4. Correlation between galaxy and halo sizes

In this section, we present the main result of this work, which is the study of the correlation between galaxy and halo sizes. We show in Figure 3 the relation between  $R_1$  and  $R_{200}$ . The top horizontal axis gives the mass of the halo,  $M_{200}$ , proportional to  $R_{200}^3$ . By definition,  $M_{200}$  and  $R_{200}$  are the mass and radius of a sphere centred on the potential minimum and with an average density 200 times the critical density of the Universe:  $M_\delta = (4/3)\pi\delta\rho_{\text{crit}}R_\delta^3$ , with  $\delta = 200$ .

We restricted the sample to only central galaxies because the calculation of the mass and radius of their haloes is more reliable. The data distribution is shown as the median in equally spaced logarithmic bins (orange dots), with the orange shaded area delimiting the 16% and 84% quantiles. The correlation

between  $R_1$  and stellar mass can be described by a double power-law function. We employed the extended double power-law function of Equation (2), with  $(x, y, x_s) \equiv (R_{200}, R_1, R_s)$ , whose  $\chi$ -square minimisation fit provides the parameter values  $(A, R_s, \alpha, \beta, s) \approx (16.4 \text{ kpc}, 175.8 \text{ kpc}, 2.04, 1.25, 15.1)$ . The scale radius,  $R_s$ , corresponds to the halo mass of about  $10^{11.8} M_\odot$ . All fitted parameters with errors are given in Table B.2. Given the tight correlation between  $R_1$  and  $M_*$ , the above correlation can be interpreted as the stellar-to-halo mass relation (see Arjona-Gálvez et al. 2025, for more details, and later in this section for the comparison with observations).

For comparison, in Figure 3 we also plot the relation between  $R_{50}$  and  $R_{200}$  (purple line and shaded area) and the corresponding linear relation of Kravtsov (2013) (dotted line). The  $R_{50} = 0.015 R_{200}$  relation was derived using a large sample of observed galaxies with stellar masses between  $10^5$  and  $10^{12} M_\odot$ , and it qualitatively agrees with the EAGLE simulation data. The scatter measured by Kravtsov (2013),  $\approx 0.2$  dex, is twice that of the simulated galaxies,  $\sigma_{\Delta \log_{10} R_1} = 0.1$  dex.

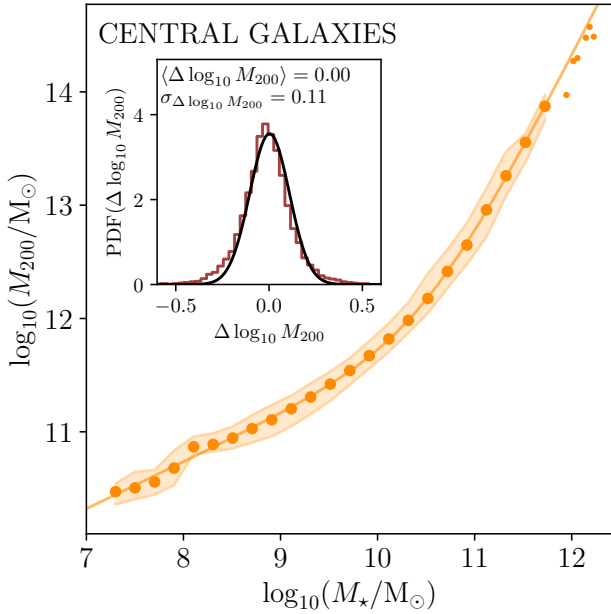
Kravtsov (2013) assumed that the relation between galaxy and halo sizes is linear, with a well-defined proportionality for all galaxies and their haloes (see also Somerville et al. 2018, for a similar hypothesis). The reasoning is that simulations show a very tight distribution of the spin parameter,  $\lambda$ , of dark matter haloes at any scale (e.g. Peebles 1969; Bullock et al. 2001), and if the halo angular momentum dictates the size of the galaxy’s disc, there should be a well-defined proportionality between the distribution of stars in the disc and the size of its dark matter halo. At a fixed stellar mass, variations of the stellar half-mass radius should reflect the dispersion around the average value of  $\lambda$ . Although we found some indication that the scatter in the mass–size relation may correlate with  $\lambda$  at a fixed stellar mass (not shown in this work), this correlation is weak.

The  $R_1$ – $R_{200}$  relation asymptotically approaches the slope  $\beta \approx 1.25$  for halo masses larger than  $\approx 10^{11.5} M_\odot$ , which is steeper than what is inferred for observed galaxies. We conclude that given the small scatter and prominent slope in the  $R_1$ – $R_{200}$  relation, the physical size of observed galaxies can be used to infer the size (and mass) of their host haloes with noticeable accuracy. For a galaxy such as the Milky Way, using  $R_1$  to infer  $R_{200}$  would be six times more precise than using  $R_{50}$ .

As a final exercise, we analysed the stellar-to-halo mass relation, shown in Figure 4. We computed the best-fit parameters for the double power-law function of Equation (2), with  $(x, y, x_s) \equiv (M_*, M_{200}, M_s)$ , which gives the values  $(A, M_s, \alpha, \beta, s) \approx (5.8 \times 10^{11} M_\odot, 3.13 \times 10^{10} M_\odot, 0.41, 1.69, 0.98)$ , confirming the change of slope at  $M_s \approx 10^{10.5} M_\odot$ . The residual distribution is shown in the inset of Figure 4, with the calculated dispersion around the mean of 0.11 dex, which is half the value reported by Mowla et al. (2019) in their work.

#### 5. Discussion

The definition of galaxy size of Trujillo et al. (2020) is physically motivated by the theoretical prediction that the onset of star formation requires specific conditions for the gas to collapse into star forming clouds (Schaye 2004). Accretion of gas into the galaxy and its subsequent cooling provide the material for the galaxy to grow both in stellar mass and size through star formation. Therefore, the physical edge of a star forming galaxy can be defined by the extension of its in situ star formation activity. The threshold gas surface density for star formation given by Schaye (2004) can be converted into a stellar surface density



**Fig. 4.** Correlation between  $M_{200}$  and  $M_*$  for central galaxies. Median (dots and solid lines) and 16% and 84% quantiles (shaded areas) are shown. We quote in the legend the average of the scatter for the entire sample of central galaxies and with respect to the analytic fit, together with dispersion around the mean. We plot in the inset the distribution of residuals for the analytic fitting function and all the galaxies in the sample.

that, even if the galaxy stops forming stars, marks the maximum extension of its in situ stellar mass buildup.

We have shown in this work that the observed galaxy mass–size relation as defined by Trujillo et al. (2020) holds in cosmological hydrodynamic simulations of the formation of galaxies. The remarkable agreement between simulations and observations can be used to infer the size and mass of dark matter haloes with an accuracy higher than 50%. The advantage of this definition of size is that the mass–size relation not only has reduced scatter but is also steeper for intermediate mass galaxies when compared to other empirical definitions of size, such as  $R_{50}$ . This is what makes it a powerful (around six times better than  $R_{50}$ ) predictor of galaxy and halo mass.

We stress that  $R_1$  is only a proxy of the physically motivated definition of size, calibrated for late-type galaxies of mass similar to that of the Milky Way. On the other hand, as shown by Trujillo et al. (2020), Arjona-Gálvez et al. (2025), and this work,  $R_1$  provides a consistently tight relation between size and stellar mass over many orders of magnitude in mass. When applied to early-type galaxies, where the extent of in situ star formation has been modified by mergers, the threshold that defines the size should be between 3 and  $10 M_{\odot} \text{pc}^{-2}$  (Chamba et al. 2022), suggesting that in situ star formation has happened at higher gas densities and larger redshift.

In a companion article, Arjona-Gálvez et al. (2025) have shown that simulations performed with different galaxy formation models all reproduce the same mass–size relation when  $R_1$  defines the galaxy size. How the interplay of gas accretion, star formation, and stellar feedback yields to the same result for a variety of numerical models is not trivial to understand. What all models have in common is that they are trying to reproduce the Kennicutt-Schmidt law and its threshold for star formation. Moreover, it seems that independent of the overall efficiency of

star formation, galaxies of the same mass end up having the same size, although living in different haloes (see their Fig. 8, top panel).

The finding of Arjona-Gálvez et al. (2025) highlights a caveat for the prediction of the halo size. If the numerical model was not tuned to abundance matching models (e.g. Moster et al. 2013; Girelli et al. 2020) and galaxies were under- or over-massive with respect to the halo in which they were formed, the prediction of the halo size would suffer systematic errors. It is, however, worth mentioning that empirical abundance matching models do not agree among them for intermediate- and low-mass galaxies, and the systematic difference would also depend on which model is adopted. On the other hand, between the models, the difference in halo mass for a fixed stellar mass is less than a factor of two and slightly larger for different numerical models.

The question that remains open is why (simulated and observed) galaxies with stellar masses spanning many orders of magnitude and different morphologies follow such a tight mass–size relation. Sánchez Almeida (2020) investigated how the scatter in the mass–size relation significantly decreases when the size is measured at a fixed surface density, showing that galaxies with the same stellar mass always share at least one radius with identical surface density. Miller et al. (2019) hinted to this when defining the size of galaxies as the radius containing a given fraction of the stellar light, and they showed that when increasing the fraction of light, star forming and quiescent galaxies tend to move closer in the mass–size plane (see their Fig. 2). None of these explanations are physical, and we will investigate this question further in the future.

## 6. Conclusions

With this work we have explored the relation between galaxy and halo size by applying the physically motivated definition of galaxy size of Trujillo et al. (2020) to simulated galaxies. We leveraged the public data of the EAGLE simulation project, namely, the galaxies produced in two simulated cosmological volumes of different resolutions, but calibrated to give similar global relations at  $z = 0$ . In the following, we provide a summary of our conclusions:

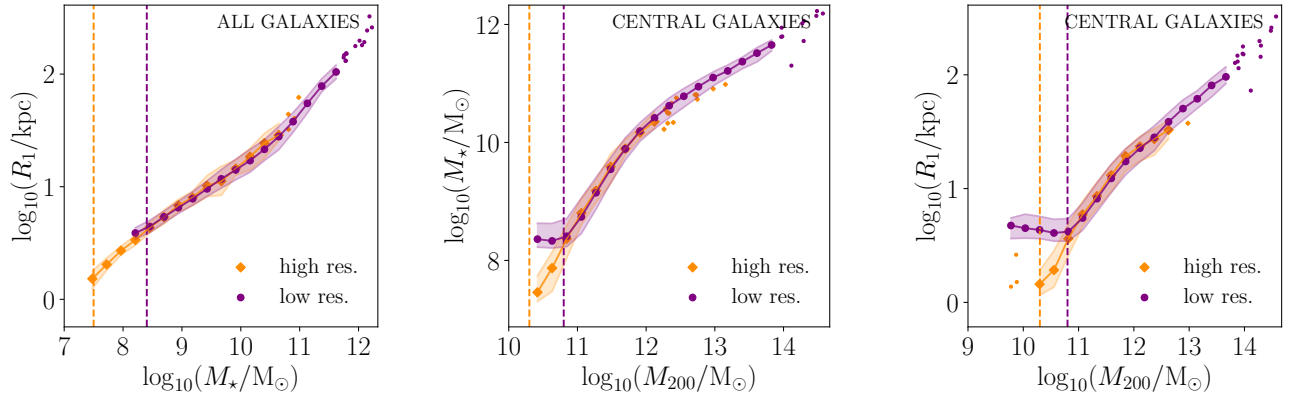
- We prove that the observed relation between galaxy stellar mass and size, as defined by Trujillo et al. (2020), holds in cosmological, hydrodynamical simulations of galaxy formation over five orders of magnitude in stellar mass. The remarkable agreement between the simulations and observations can be used to precisely infer the galaxy stellar mass from its observed physical size (Mowla et al. 2019; Sánchez Almeida 2020). We provide a fit to the distribution for galaxies more massive than  $10^{8.5} M_{\odot}$  (Figure 2 and Table B.1). We recovered the same scatter found by Trujillo et al. (2020).
- We propose the theoretical prediction that the size ( $R_{200}$ ) and mass ( $M_{200}$ ) of the dark matter halo hosting the galaxy can be accurately inferred by knowing the size of its central galaxy. Figure 3 and Table B.2 quantify this prediction. The small scatter in the galaxy mass–size relation is a factor of two smaller than the observed scatter in the galaxy-to-halo size relation, whereas the slope of the relation greatly decrease the range of uncertainty when inferring the stellar mass of intermediate mass galaxies. These confirm the superiority of the galaxy size definition by nearly a factor of six when compared to observations using the half-mass radius (Kravtsov 2013).

The origin of the relation and its connection to gas accretion, star formation, and stellar feedback is still poorly understood. More theoretical work will be dedicated to it in the near future.

*Acknowledgements.* This work has been supported by the Spanish Ministry of Science, Innovation and Universities (*Ministerio de Ciencia, Innovación y Universidades*, MICIU) through the research grants PID2021-122603NB-C22 and PID2022-140869NB-I00. CDV also acknowledges support from MICIU in the early stages of development of this work through grants RYC-2015-18078 and PGC2018-094975-C22. This research made use of computing time on the high-performance computing systems DELMOS and DIVA of the *Instituto de Astrofísica de Canarias*. IT acknowledges support from: the ACHSI, Consejería de Economía, Conocimiento y Empleo del Gobierno de Canarias and the European Regional Development Fund (ERDF) under grant PROID2021010044; the IAC project P/302302, financed by the Ministry of Science and Innovation, through the State Budget, and by the Canary Islands Department of Economy, Knowledge, and Employment, through the Regional Budget of the Autonomous Community. This research also acknowledges support from the European Union through the following grants: UNDAK and ‘Excellence in Galaxies-Twinning the IAC’ of the EU Horizon Europe Widening Actions programmes (project numbers 101159929 and 101158446). Funding for this work/research was provided by the European Union MSCA EDUCADO (GA 101119830). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them. The following Python packages have been used for this research: H5PY (<https://www.h5py.org>), NUMPY (<https://numpy.org>) (Harris et al. 2020), SCIPY (<https://scipy.org>) (Virtanen et al. 2020), MATPLOTLIB (<https://matplotlib.org/>) (Hunter 2007).

## References

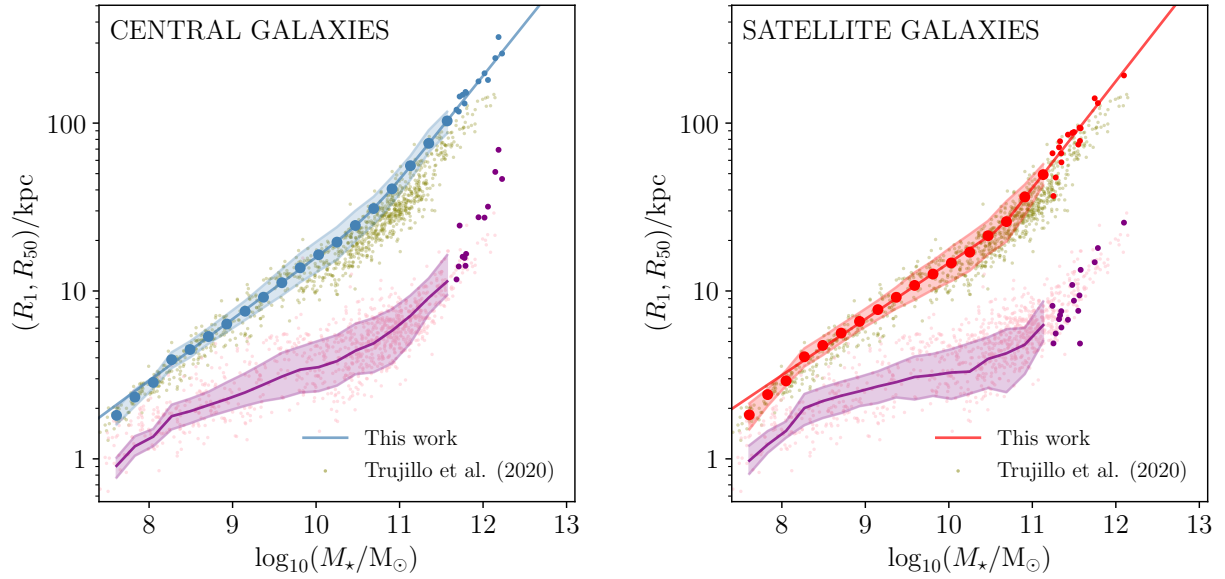
- Abazajian, K., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2003, *AJ*, **126**, 2081
- Arjona-Gálvez, E., Cardona-Barrero, S., Grand, R. J. J., et al. 2025, *A&A*, **699**, A301
- Bakos, J., Trujillo, I., & Pohlen, M. 2008, *ApJ*, **683**, L103
- Bigiel, F., Leroy, A., Walter, F., et al. 2008, *AJ*, **136**, 2846
- Bower, R. G., Schaye, J., Frenk, C. S., et al. 2017, *MNRAS*, **465**, 32
- Broeils, A. H., & Rhee, M. H. 1997, *A&A*, **324**, 877
- Buitrago, F., & Trujillo, I. 2024, *A&A*, **682**, A110
- Bullock, J. S., Dekel, A., Kolatt, T. S., et al. 2001, *ApJ*, **555**, 240
- Chamba, N., Trujillo, I., & Knapen, J. H. 2022, *A&A*, **667**, A87
- Crain, R. A., Schaye, J., Bower, R. G., et al. 2015, *MNRAS*, **450**, 1937
- Dalla Vecchia, C., & Schaye, J. 2012, *MNRAS*, **426**, 140
- Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, *ApJ*, **292**, 371
- Dolag, K., Borgani, S., Murante, G., & Springel, V. 2009, *MNRAS*, **399**, 497
- Durier, F., & Dalla Vecchia, C. 2012, *MNRAS*, **419**, 465
- Einasto, J., Klypin, A. A., Saar, E., & Shandarin, S. F. 1984, *MNRAS*, **206**, 529
- Girelli, G., Pozzetti, L., Bolzonella, M., et al. 2020, *A&A*, **634**, A135
- Golini, G., Trujillo, I., Zaritsky, D., et al. 2025, *A&A*, **700**, A91
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Nature*, **585**, 357
- Holmberg, E. 1958, *Meddelanden fran Lunds Astronomiska Observatorium Serie II*, **136**, 1
- Hunter, J. D. 2007, *Comput. Sci. Eng.*, **9**, 90
- Kennicutt, R. C. 1998, *ApJ*, **498**, 541
- Kravtsov, A. V. 2013, *ApJ*, **764**, L31
- Martin, C. L., & Kennicutt, R. C. 2001, *ApJ*, **555**, 301
- McAlpine, S., Helly, J. C., Schaller, M., et al. 2016, *Astron. Comput.*, **15**, 72
- McAlpine, S., Bower, R. G., Harrison, C. M., et al. 2017, *MNRAS*, **468**, 3395
- Miller, T. B., van Dokkum, P., Mowla, L., & van der Wel, A. 2019, *ApJ*, **872**, L14
- Moster, B. P., Naab, T., & White, S. D. M. 2013, *MNRAS*, **428**, 3121
- Mowla, L., van der Wel, A., van Dokkum, P., & Miller, T. B. 2019, *ApJ*, **872**, L13
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, *ApJ*, **490**, 493
- Negri, A., Dalla Vecchia, C., Aguerri, J. A. L., & Bahé, Y. 2022, *MNRAS*, **515**, 2121
- Peebles, P. J. E. 1969, *ApJ*, **155**, 393
- Pillepich, A., Nelson, D., Hernquist, L., et al. 2018, *MNRAS*, **475**, 648
- Planck Collaboration XVI. 2014, *A&A*, **571**, A16
- Rohr, E., Feldmann, R., Bullock, J. S., et al. 2022, *MNRAS*, **510**, 3967
- Rosas-Guevara, Y. M., Bower, R. G., Schaye, J., et al. 2015, *MNRAS*, **454**, 1038
- Sánchez Almeida, J. 2020, *MNRAS*, **495**, 78
- Schaller, M., Dalla Vecchia, C., Schaye, J., et al. 2015a, *MNRAS*, **454**, 2277
- Schaller, M., Robertson, A., Massey, R., Bower, R. G., & Eke, V. R. 2015b, *MNRAS*, **453**, L58
- Schaye, J. 2004, *ApJ*, **609**, 667
- Schaye, J., & Dalla Vecchia, C. 2008, *MNRAS*, **383**, 1210
- Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, *MNRAS*, **446**, 521
- Sérsic, J. L. 1968, *Atlas de Galaxias Australes* (Cordoba, Argentina: Observatorio Astronomico)
- Somerville, R. S., Behroozi, P., Pandya, V., et al. 2018, *MNRAS*, **473**, 2714
- Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, *MNRAS*, **328**, 726
- The EAGLE Team 2017, arXiv e-prints [arXiv:1706.09899]
- Trujillo, I., Chamba, N., & Knapen, J. H. 2020, *MNRAS*, **493**, 87
- Vazdekis, A., Koleva, M., Ricciardelli, E., Röck, B., & Falcón-Barroso, J. 2016, *MNRAS*, **463**, 3409
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nat. Meth.*, **17**, 261
- Wang, J., Koribalski, B. S., Serra, P., et al. 2016, *MNRAS*, **460**, 2143
- Wiersma, R. P. C., Schaye, J., & Smith, B. D. 2009a, *MNRAS*, **393**, 99
- Wiersma, R. P. C., Schaye, J., Theuns, T., Dalla Vecchia, C., & Tornatore, L. 2009b, *MNRAS*, **399**, 574



**Fig. A.1.** *Left panel.* Relation between size and stellar mass for all galaxies in the two simulations. Dots and diamonds are the median values in bins of halo mass, and the shaded areas indicate the 16% and 84% quantiles. Single dots are for individual galaxies, in bins containing fewer than ten galaxies. The vertical lines mark the minimum masses considered for the two simulations in the computation of the analytic fits. *Middle panel.* Relation between stellar mass,  $M_*$ , and halo mass,  $M_{200}$ , for all central galaxies in the two simulations. Symbols are as in the left panel. The vertical lines mark the lower halo mass limit of the samples extracted from the two simulations and employed in the computation of the  $R_1-R_{200}$  relation. *Right panel.* Same as the middle panel, but for the  $R_1-M_{200}$  relation.

## Appendix A: Selection of the sample

We show in figure A.1 the distribution of processed galaxies in the  $(M_*, R_1)$  (left panel, all galaxies),  $(M_{200}, M_*)$  (mid panel, only central galaxies) and  $(M_{200}, R_1)$  (right panel, only central galaxies) planes. The median values in bins of halo mass are shown as dots and diamonds for the low- and high-resolution simulations, respectively. The shaded areas indicate the 16% and 84% quantiles. Smaller dots and diamonds are for individual galaxies that are in bins with fewer than ten galaxies. The two simulations overlap remarkably well over more than two orders of magnitude in both stellar mass and halo mass. For the computation of the  $R_1-M_*$  relation, we selected galaxies with  $M_* > 10^{7.5} M_\odot$  from the recalibrated simulations, and with  $M_* > 8 \times 10^{7.5} M_\odot$  from the reference simulation, accordingly to the different mass resolution. These mass thresholds are depicted with vertical dashed lines in the left panel. At the low-mass ends, the relations with  $M_{200}$  (mid and right panels) flatten when reaching the resolution limit. For the computation of the  $R_1-R_{200}$  relation, we then consider central galaxies in haloes with mass above  $M_{\text{thr}} = 10^{10.3} M_\odot$  and  $8 \times M_{\text{thr}}$  for the recalibrated and reference simulations, respectively, and accordingly to their particle mass resolution. This choice is adequate for both  $M_*-M_{200}$  and  $R_1-M_{200}$  relations.



**Fig. B.1.** Correlation between  $R_1$  and stellar mass for central (left panel) and satellite (right panel) galaxies in the sample. The median values for each bin are plotted together with the 16% and 84% quantiles (shaded area). For bins with fewer than ten galaxies, single values are plotted. Both fits to the data are for stellar masses  $M_* > 10^{8.5} M_\odot$ . For comparison, we plot the relation for  $R_{50}$  in purple.

**Table B.1.** Best-fit parameters for the stellar mass–size relation for all galaxies and central and satellite galaxies. The errors are given by the  $\chi$ -square minimisation.

GALAXY MASS–SIZE RELATIONS					
	A/kpc	$M_s/M_\odot$	$\alpha$	$\beta$	$s$
all galaxies	$22.8 \pm 1.0$	$(3.21 \pm 0.35) \times 10^{10}$	$0.35 \pm 0.05$	$0.60 \pm 0.01$	$4.25 \pm 2.50$
central galaxies	$27.1 \pm 1.3$	$(4.32 \pm 0.57) \times 10^{10}$	$0.37 \pm 0.04$	$0.62 \pm 0.02$	$2.48 \pm 0.75$
satellite galaxies	$25.4 \pm 5.7$	$(6.35 \pm 0.48) \times 10^{10}$	$0.31 \pm 0.06$	$0.75 \pm 0.24$	$1.70 \pm 0.82$

**Table B.2.** Best-fit parameters for the galaxy size–halo size relation for central galaxies. The errors are given by the  $\chi$ -square minimisation.

GALAXY SIZE–HALO SIZE RELATION					
	A/kpc	$R_s/\text{kpc}$	$\alpha$	$\beta$	$s$
central galaxies	$16.4 \pm 1.4$	$175.8 \pm 10.4$	$2.04 \pm 0.07$	$1.25 \pm 0.03$	$15.1 \pm 17.9$

## Appendix B: Mass–size relation for central and satellite galaxies

We restricted the sample to central and satellite galaxies, and repeated the analysis of section 3. We show in figure B.1 the stellar mass–size relations for central (left panel) and satellite (right panel) galaxies. The figure is similar to figure 2, where median values are fitted with a double power-law (equation 2) for stellar masses  $M_* > 10^{8.5} M_\odot$ . For the same galaxies we also show the relation between stellar mass and half-mass radius,  $R_{50}$ , in purple. The median is given by the purple solid lines, whereas the shaded areas represent the 16% and 84% quantiles. We report in table B.1 the best-fit parameters for the stellar mass–size relation for all galaxies and for the subsamples of central and satellite galaxies, and in table B.2 the best-fit parameters for the galaxy size–halo size relation for central galaxies. The errors on the parameters are given by the  $\chi$ -square minimisation. The dispersion around the best-fitting function is  $\sigma_{\Delta \log_{10} R_1} = 0.05$  and  $\sigma_{\Delta \log_{10} R_1} = 0.07$  for central and satellite galaxies, respectively.