

# Halo concentrations from extended Press–Schechter merger histories

Andrew J. Benson<sup>1</sup>,<sup>1★</sup> Aaron Ludlow<sup>3</sup> and Shaun Cole<sup>4</sup>

<sup>1</sup>*Carnegie Observatories, 813 Santa Barbara Street, Pasadena, CA 91101, USA*

<sup>2</sup>*accidentally clicked "+Affiliation" - please remove*

<sup>3</sup>*International Centre for Radio Astronomy Research, 7 Fairway, Crawley, 6009 Perth, WA, Australia*

<sup>4</sup>*Department of Physics, Science Laboratories, Institute for Computational Cosmology, Durham University, South Road, Durham DH1 3LE, UK*

Accepted 2019 February 21. Received 2019 February 1; in original form 2018 December 14

## ABSTRACT

We apply the model relating halo concentration to formation history proposed by Ludlow et al. to merger trees generated using an algorithm based on excursion set theory. We find that while the model correctly predicts the median relation between halo concentration and mass, it underpredicts the scatter in concentration at fixed mass. Since the same model applied to  $N$ -body merger trees predicts the correct scatter, we postulate that the missing scatter is due to the lack of any environmental dependence in merger trees derived from excursion set theory. We show that a simple modification to the merger tree construction algorithm, which makes merger rates dependent on environment, can increase the scatter by the required amount, and simultaneously provide a qualitatively correct correlation between environment and formation epoch in the excursion set merger trees.

**Key words:** galaxies: haloes – dark matter.

## 1 INTRODUCTION

It is now very well established that the density profiles of cold dark matter (CDM) haloes can be well described by simple, universal forms such as the NFW profile Navarro, Frenk & White (1996, 1997) or, somewhat more accurately, by the Einasto profile (Gao et al. 2008; Ludlow & Angulo 2017). The density profile is thought to play a significant role in shaping the properties of galaxies that form in dark matter haloes, as it contributes significantly to the rotation curve, and gravitational potential.

In semi-analytic models of galaxy formation, the density profile of the halo directly affects determinations of galaxy sizes (Cole et al. 2000; Jiang et al. 2018), and in many models affects the mass loading of outflows from galaxies (Cole et al. 2000; Benson 2012). While semi-analytic models of galaxy formation are most often applied to dark matter halo merger trees extracted from  $N$ -body simulations, from which dark matter profiles can be measured directly, they can also be applied to merger trees generated using other techniques, such as those based on excursion set theory (Bond et al. 1991; Lacey & Cole 1993; Cole et al. 2000; Parkinson, Cole & Helly 2008). Such approaches have some advantages over the use of  $N$ -body merger trees, for example, allowing much finer time resolution to be achieved (which can affect the results of semi-analytic models; Benson et al. 2012), rapid exploration of different dark matter physics (Benson et al. 2013), and avoidance of numerical noise issues that occur in  $N$ -body haloes at low particle number (Benson 2017a).

However, merger trees built from excursion set theory have, so far, only provided halo masses – they say nothing about other key physical properties of haloes such as their density profiles, spins, or environment. Typically, these quantities are assigned to haloes in such trees by appealing to an empirical correlation with halo mass and redshift, or by drawing at random from some measured distribution. In the case of halo concentrations for example, the usual approach is to assign a concentration based on a concentration–mass–redshift relation measured from an  $N$ -body simulation. This approach has two significant disadvantages. First, concentration is not uniquely determined by halo mass and redshift – there is significant scatter in concentration at fixed mass and redshift. As such, this scatter will be missing from calculations that depend on halo structure (e.g. galaxy sizes; Jiang et al. 2018). This scatter cannot be incorporated by simply adding a random perturbation around the median concentration relation as the offset from the median is expected to be correlated across time in any given halo. The second disadvantage of this approach is that it is known that halo structure correlates with the formation history of the halo (Navarro et al. 1997; Bullock et al. 2001; Ludlow et al. 2013).

Based on these correlations, Ludlow et al. (2014, see also Correa et al. 2015) and Ludlow et al. (2016) developed models that relate the density profiles of dark matter haloes to their assembly histories. The model of Ludlow et al. (2016) is based upon the time evolution of the total mass of progenitor haloes collapsed (the ‘collapsed mass history’, or CMH), and is able to correctly predict concentrations for both CDM and warm dark matter (WDM) power spectra, while the Ludlow et al. (2014) model, which is based on the ‘mass accretion history’ (MAH; the mass of the main progenitor halo) fails to reproduce the concentration–mass relation in WDM. As shown by

\* E-mail: [abenson@carnegiescience.edu](mailto:abenson@carnegiescience.edu)

Ludlow et al. (2014, 2016), these models work well for haloes that are relaxed and in equilibrium, for which concentration is mainly set by formation time and is independent of other factors (such as substructure, recent collapse times, etc.). As such, the models are relevant to a biased set of simulated merger trees, since selecting relaxed haloes naturally biases trees to those that form early (on average). While these models were developed and tested primarily on assembly histories extracted from  $N$ -body simulations, they were also shown to be applicable to simple, spherical collapse models. As such, they can be similarly applied to assembly histories derived from excursion set-based merger trees.

It has also been shown (e.g. Zehavi et al. 2018) that the formation times of haloes correlate with environment. Any model that relates concentration to halo assembly history (and therefore formation time) should also consider the relation with environment. In excursion set theory, the large-scale environment of a halo has no effect on its assembly history (Bond et al. 1991), at least in the standard case where the power spectrum is filtered using a window function that is sharp in  $k$ -space.

In this work, we examine the outcome of applying the Ludlow et al. (2016) model for halo concentration to merger trees built using the Parkinson et al. (2008) algorithm. We further explore how a simple model for the effects of environment on halo assembly can be introduced into this algorithm, and show that this improves agreement with the distribution of concentrations derived from  $N$ -body simulations.

The remainder of this paper is organized as follows. In Section 2, we describe how we measure the concentrations of  $N$ -body haloes, and estimate the uncertainties on these measurements. We then describe our algorithm for introducing an environmental dependence into the Parkinson et al. (2008) tree-building algorithm, and how the parameters of this model are constrained. In Section 3, we show the results of this model and compare it to those from  $N$ -body simulations. Finally, in Section 4 we discuss the implications of our results.

## 2 METHODS

Our goal in this work is two-fold: to introduce a model for halo concentrations into the Parkinson et al. (2008) tree-building algorithm that is dependent upon the formation history of each halo, and to compare the results of that model to those from cosmological  $N$ -body simulations. This will require determining the full distribution of concentration parameter as a function of halo mass. To that end, we first describe how we measure concentrations of a sample of  $N$ -body haloes and, importantly, how we determine the uncertainties in these measurements arising from the finite number of particles with which each halo is represented (Benson 2017a,b).

### 2.1 Concentrations of $N$ -body haloes

#### 2.1.1 Fitting procedure

We make use of the *Copernicus Complexio* (COCO) simulations of Hellwing et al. (2016) to determine the distribution of  $N$ -body halo concentrations. As described by Hellwing et al. (2016), haloes were identified in the COCO simulations using the friends-of-friends algorithm (Davis et al. 1985), with a linking length parameter of  $b = 0.2$ . For each halo a mass  $M_{200c}$  is determined as the mass within a sphere centred on the halo particle with the minimum gravitational potential, and enclosing a mean density equal to 200 times the

critical density. This is a widely used definition of halo mass, and we will use it in this work to calibrate our model. Once calibrated in this way the model can easily be applied to other halo mass definitions (see Section 2.2.1). As shown by Ludlow et al. (2012), haloes of a given mass that collapse very recently (i.e. double their mass in less than a crossing time) have concentrations that are biased relative to the normal concentration-formation time relation. Such haloes, which are dynamically young and not in equilibrium, do not have well-defined concentrations. Concentrations of a single such halo will vary widely over short time-scales if the halo is undergoing a merger or rapid/transient but substantial accretion for example. As such, concentration models based on mean MAH such as the one used in this work cannot be applied to such haloes and we follow Ludlow et al. (2016) and exclude any haloes that more than doubled their mass in the last 1.25 crossing times, or approximately the last 3.7 Gyr.

To measure concentrations of haloes via their  $N$ -body representations we construct histograms of the number of particles in the halo in 31 spherical shells with logarithmically spaced radii between  $3.46 \times 10^{-2} r_{200c}$  and  $0.764 r_{200c}$ , where  $r_{200c}$  is the radius enclosing a mean density equal to 200 times the critical density, and then fit these using an Einasto profile (Einasto 1965), which has been found to be a good fit to the density profile of cosmological  $N$ -body haloes (Gao et al. 2008). In performing the fit we include only those bins that satisfy the inequalities:

$$r \geq 2\epsilon, \quad (1)$$

$$\kappa(r) \geq \kappa_{\text{conv}}, \quad (2)$$

where  $\epsilon$  is the softening length (230pc for the COCO simulations), and  $\kappa(r) = t_{\text{relax}}(r)/t_{\text{circ}}(r_{200c})$  is a convergence criterion defined by Power et al. (2003; their equation 20). We choose a value of  $\kappa_{\text{conv}} = 1$  to ensure the profile is minimally affected by numerical relaxation. (Power et al. 2003 advocate  $\kappa = 0.6$ , so we are more conservative in defining convergence as larger  $\kappa$  implies better convergence, Navarro et al. 2010.) We find that this relaxation criterion excludes only 1.2 per cent of haloes. We furthermore retain only those haloes for which at least 16 bins satisfy the above inequalities to ensure that the fit is well constrained.

To find the best-fitting profile for each halo we minimize a goodness-of-fit measure:

$$\phi^2 = \sum_{i=1} \log^2 \left( \frac{N_i^{(n)}}{N_i^{(e)}} \right), \quad (3)$$

where  $N_i^{(n)}$  is the number of particles in the  $i$ th bin of the  $N$ -body profile, and  $N_i^{(e)}$  if the mean number of particles expected in that bin assuming an Einasto profile.<sup>1</sup>

<sup>1</sup>We also considered an alternative approach in which we maximized the Poisson likelihood  $\log \mathcal{L} = \sum_i -N_i^{(e)} + N_i^{(n)} \log N_i^{(e)} - \log \Gamma(N_i^{(n)} + 1)$ , which may be expected to be valid if the number of particles in each bin obeys Poisson statistics (Benson 2017b). We found that this leads to small but significant differences in the resulting distribution of concentration parameters. Specifically, the scatter in  $\log_{10} c_{200c}$  at fixed mass is decreased by around 0.02 dex. However, we find that the Poisson model is not a good description of the distribution of particle number in each bin, presumably because of fluctuations introduced by subhaloes and other internal structure. This does, however, point to the need for a more careful understanding of this issue for precision measurements of halo concentrations.

Our Einasto profiles are described by three parameters:  $m_0$  (the total mass of the halo in units of the nominal  $M_{200c}$  mass reported by SUBFIND),  $r_{-2}$  (the radius at which the logarithmic slope of the density profile equals  $-2$ , measured in units of the nominal  $r_{200c}$  radius reported by SUBFIND), and  $\alpha$  (the shape parameter of the Einasto profile). We choose to fix  $\alpha = 0.18$  (consistent with the typical shape of haloes reported by Gao et al. 2008) as it is generally not well constrained by the data, and explore broad ranges of the remaining two parameters:  $m_0 = (0.5, 1.5)$  and  $r_s = (0.01, 0.50)$ . The range for  $m_0$  is chosen to be sufficiently broad to encapsulate all plausible profiles, and that on  $r_s$  is chosen to encompass the plausible range of concentrations (Gao et al. 2008). Once the best-fitting parameters for the Einasto profile,  $(\hat{m}_0, \hat{r}_{-2})$ , have been determined we use them to compute the value of  $\hat{r}_{200c}$  (i.e. the radius enclosing a mean density equal to 200 times the critical density in the best-fitting profile in units of the nominal  $r_{200c}$  radius reported by SUBFIND, and which may differ slightly from unity) for this profile, and then compute a concentration  $c_{200c} = \hat{r}_{200c}/\hat{r}_{-2}$ .

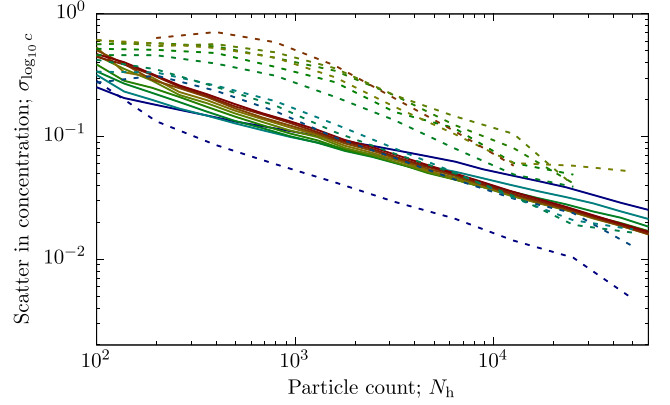
### 2.1.2 Uncertainties in $N$ -body concentration estimates

As has been explored by Trenti et al. (2010) and Benson (2017a,b), measurements of halo properties from  $N$ -body simulations are subject to noise arising from the discrete sampling of the underlying distribution function by a finite number of particles. To explore the effects of the discreteness noise on our concentration measures we perform two Monte Carlo experiments.

In the first, we generate spherically symmetric Einasto density profiles with a range of different concentrations, and sample from each profile using different numbers of particles via a Poisson process (Benson 2017b). These  $N$ -body representations are then fit using the procedure described above [without any application of the Power et al. (2003) convergence criterion, which does not apply in this simple experiment]. This is repeated for many different random realizations of the particle distribution, and for many different total number of particles in the halo to quantify the uncertainty and bias in the recovered concentration.

In our second experiment, we utilize  $N$ -body haloes from the Millennium Simulation (Springel et al. 2005), fitting them using the same method as described above. However, we sample particles uniformly at random from each halo (with replacement) to generate a large number of random realizations of the halo, and repeat the fitting procedure on each realization. This sampling is initially done at a rate of  $\mu = 1$  (where we define  $\mu$  as the mean number of times that any given particle will appear in a random realization), and then repeated with lower sampling rates to generate realizations of the haloes with fewer particles. This allows us to assess the uncertainty and bias in the measured concentrations for realistic, cosmological haloes.

The results of these experiments are shown in Fig. 1. The scatter,  $\sigma_{\log_{10} c_{200c}}$ , is shown as a function of the number of particles in a halo,  $N_h$ , on the  $x$ -axis, and as a function of concentration (shown by colour – with  $c_{200c} = 4$  shown in blue, increasing to  $c_{200c} = 20$  shown in red). Solid lines show the results of fitting Monte Carlo realizations of idealized, spherical, Einasto profiles. Dashed lines indicate results from fitting cosmological  $N$ -body haloes, with each dashed line corresponding to the mean scatter measured over all cosmological haloes of that concentration. Cosmological haloes that deviate significantly from the expected scaling are excluded – in these cases the haloes are not well described by Einasto



**Figure 1.** Scatter in the concentration parameter,  $\log_{10} c_{200c}$ , derived from Monte Carlo experiments. The scatter,  $\sigma_{\log_{10} c_{200c}}$ , is shown as a function of the number of particles in a halo,  $N_h$ , on the  $x$ -axis, and as a function of concentration (shown by colour – with  $c_{200c} = 4$  shown in blue, increasing to  $c_{200c} = 20$  shown in red). Solid lines show the results of fitting Monte Carlo realizations of idealized, spherical, Einasto profiles. Dashed lines indicate results from fitting cosmological  $N$ -body haloes, the particles of which were resampled (with replacement) at different rates and fit to estimate concentration. Each dashed line corresponds to the mean scatter measured over all cosmological haloes of that concentration. Cosmological haloes that deviate significantly from the expected scaling are excluded.

profiles and so we do not expect our fits to give meaningful results.

We fit the dependence of  $\sigma_{\log_{10} c_{200c}}$  on particle number and concentration using simple polynomial fits. For idealized haloes we find that the scatter in concentrations can be described by the model

$$\begin{aligned} \log_{10} \sigma_{\log_{10} c_{200c}} &= \alpha(c_{200c}) + b \log_{10} N_h, \\ \alpha(c_{200c}) &= -1.04 + 2.03 \log_{10} c_{200c} - 0.53 \log_{10}^2 c_{200c}, \\ b &= -0.47 \pm 0.06, \end{aligned} \quad (4)$$

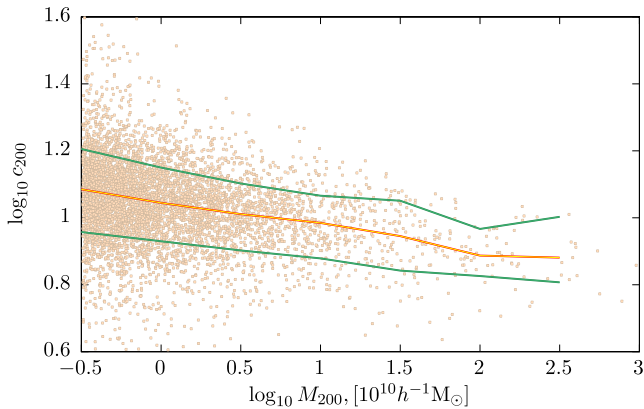
while for cosmological haloes the scatter is described by

$$\begin{aligned} \log_{10} \sigma_{\log_{10} c_{200c}} &= \alpha(c_{200c}) + b \log_{10} N_h, \\ \alpha(c_{200c}) &= -0.20 + 1.46 \log_{10} c_{200c} - 0.25 \log_{10}^2 c_{200c}, \\ b &= -0.54 \pm 0.06. \end{aligned} \quad (5)$$

In both cases, the dependence of the relation on particle number is consistent with the expected  $N_h^{-1/2}$  scaling due to Poisson sampling. The normalization is significantly higher for the cosmological haloes than for the idealized haloes.

Achieving an uncertainty in concentration parameter less than 0.1 dex requires  $N_h \gtrsim 2 \times 10^3$  for idealized haloes, and  $N_h \gtrsim 6 \times 10^3$  for cosmological haloes. When fitting our model to match  $N$ -body data, we will use the model of equation (5) to forward model the scatter in concentration parameter.

We also considered the bias in concentration – defined as the difference between the measured and true concentrations. (In the case of cosmological haloes, the mean concentration found when fitting to the halo particles sampled at a rate of  $\mu = 1$  is taken to be our estimate of the true concentration.) Idealized haloes show a trend of positive bias in fits, while cosmological haloes show no significant trend in bias – the bias estimates for individual haloes



**Figure 2.** Concentrations of haloes in the COCO simulation as a function of halo mass. Points show a random subsample of all haloes fitted. The solid yellow line shows the median concentration as a function of mass, while the solid green lines show the 16th and 84th percentiles of the distribution.

scatter around zero. As such, we assume no bias when modelling concentrations.

### 2.1.3 Concentrations of COCO haloes

We fit profiles from the COCO simulations (Hellwing et al. 2016) using the procedure described in Section 2.1.1. Results are shown in Fig. 2. Points show a random subsample of all haloes fitted. The solid yellow line shows the median concentration as a function of mass, while the solid green lines show the 16th and 84th percentiles of the distribution. The expected scatter in concentration arising from finite particle number effects (as computed using the model of equation 5) is negligibly small, less than 0.035 dex, in this figure as all haloes shown have  $N_p > 25\,000$  particles.

## 2.2 Merger trees

Our goal is to apply the concentration model of Ludlow et al. (2016) to merger trees generated using an excursion set-based approach. Specifically, we will use the algorithm of Parkinson et al. (2008) – which is itself a modification of the algorithm of Cole et al. (2000) – to construct merger trees. In Appendix A, we describe a small optimization to the tree building algorithm of Cole et al. (2000).

### 2.2.1 Concentrations

To compute concentrations of our merger trees we apply either the Ludlow et al. (2014) or Ludlow et al. (2016) algorithms. In the Ludlow et al. (2016) algorithm, the mean density within the scale radius,  $\langle \rho_{-2} \rangle$ , for a halo is given by

$$\frac{\langle \rho_{-2} \rangle}{\rho_0} = C \frac{\rho_{\text{crit}}(t_c)}{\rho_0}, \quad (6)$$

where  $t_c$  is the time at which the halo had first assembled a mass  $M_{-2} = 4\pi \langle \rho_{-2} \rangle r_{-2}^3 / 3$  into progenitors of mass greater than  $fM$ , where  $M$  is the mass of the halo in question.<sup>2</sup> To find  $r_{-2}$  for any

<sup>2</sup>We note that while the  $N$ -body haloes used in calibrating our model are based on the ‘ $M_{200c}$ ’ mass definition (see Section 2.1.1), we are not restricted to using that mass definition in the generated merger trees used here. Instead, we can adopt any mass definition for the haloes in those trees and simply

given halo in a merger tree, we use an iterative approach. We make an initial guess for  $r_{-2}$  (typically based on the median concentration–mass–redshift relation) and use this, together with the known mass and virial radius of the halo, to determine  $M_{-2}$ . We then search backward in time through the progenitors of the halo until we find the time,  $t_c$ , at which the halo first had a total mass equal to  $M_{-2}$  in progenitor haloes of mass  $fM$  or greater. An estimate of  $\langle \rho_{-2} \rangle$  can then be made from equation (6), which in turn can be used (along with the mass and virial radius of the halo) to compute an updated estimate of  $r_{-2}$ . This new value of  $r_{-2}$  is used in the next iteration of this procedure, which is continued until the value of  $r_{-2}$  has converged to a desired level. In this way, scale radii,  $r_{-2}$ , are found for all haloes in the merger tree.<sup>3</sup>

For the algorithm of Ludlow et al. (2014) we use a similar approach, except that the condition that the mass in all progenitors above mass  $fM$  must equal  $M_{-2}$  at  $t_c$ , we instead require that the mass in the primary (most massive) progenitor equals  $M_{-2}$ .

### 2.2.2 Environmental dependence

It is now well established that the properties and formation histories of dark matter haloes correlate with their environment (Sheth & Tormen 2004; Gao, Springel & White 2005), such that (for high-mass haloes) highly concentrated/old haloes cluster more weakly than low-concentration/young haloes at fixed mass. As discussed by Paranjape, Hahn & Sheth (2018), analytic models of halo formation (e.g. Dalal et al. 2008; Desjacques 2008; Musso & Sheth 2012; Castorina & Sheth 2013) predict such a correlation arising from correlations between the large-scale density environment around the Lagrangian volume of a halo, and the density structure within that proto-halo region. Such correlations can be captured by extensions of the excursion set approach (Castorina & Sheth 2013). However, it is also well established (see Paranjape et al. 2018) that such models cannot explain the reversal of the trend of clustering versus concentration at low masses. This reversal has been hypothesized to be due to the effects of tidal interactions with the non-linear density field (Hahn et al. 2009; Borzyszkowski et al. 2017). In this work, we attempt to incorporate only the first of these effects, developing a simple, empirical model that aims to capture the effects of large-scale density environment on the formation of haloes, leaving consideration of the effects of the non-linear density field to a future work.

Since we wish to examine the effects of environment on our merger trees, we select for each tree an environmental overdensity (defined within some spherical region of radius  $R_c$ ) in the linear regime extrapolated to the present epoch,  $\delta_c$ . This is drawn from the distribution of Mo & White (1996; their equation 9), that is from a normal distribution conditioned upon the fact that the region has not exceeded the threshold for collapse on any larger scale. This environmental overdensity is assumed to be the same for all progenitor haloes in the given merger tree.

compute the corresponding  $M_{200c}$  mass given the density profile of the halo (which is found iteratively as will be described).

<sup>3</sup>For haloes close to the resolution limit of our merger trees progenitors of mass  $fM$  may not be resolved. In such cases we can simply assign a concentration from a measured concentration–mass–redshift relation. In this work, we do not care about such haloes as we ensure that haloes for which we wish to compute the concentration are sufficiently well resolved that this problem does not occur.

Each merger tree, with final halo of mass  $M_i$  (defined such that  $M_i > M_{i-1}$ ), is considered to have an abundance

$$n_i = \frac{1}{M_i} \int_{\sqrt{M_{i-1}M_i}}^{\sqrt{M_iM_{i+1}}} Mn(M, \delta_c) dM. \quad (7)$$

That is, the abundance is chosen such that the number of such trees per unit volume contains the total mass expected from the halo mass function  $n(M, \delta_c)$  within the mass interval associated with the tree.

To determine the environment-dependent halo mass function,  $n(M, \delta_c)$ , we use the peak-background split approach (Bond et al. 1991; Bond & Myers 1996). This is formally derived within the excursion set theory, and predicts a mass function of the Press & Schechter (1974) form but with the mappings  $\delta_c \rightarrow \delta_c - \delta_e$  and  $\sigma^2(M) \rightarrow \sigma^2(M) - \sigma^2(M_e)$ , where  $M_e$  is the mass corresponding to the spherical region used to define the environment, and  $\sigma^2(M)$  is the fractional variance in the mass in spheres containing mean mass  $M$  in the linear theory density field extrapolated to the present day. In this work, we use the mass function of Sheth & Tormen (2002) with parameter values tuned as described in Section 2.2.3 rather than the Press–Schechter mass function. We retain this same mapping of  $\delta_c$  and  $\sigma^2(M)$  suggested by the peak-background split.

In the case of the Press–Schechter mass function, when the peak-background split mass function is integrated over all environments (weighted by the appropriate distribution function – see Mo & White 1996) the result is the original Press–Schechter mass function (i.e. that unconditioned on environment). For halo masses that exceed the mass of the background, the unconditioned mass function applies. Therefore, in the case of the Press–Schechter mass function, the environmentally averaged mass function is identical to the unconditioned mass function.

The same is not true for the Sheth–Tormen mass function. Instead, we find that averaging this mass function over environment leads to a result slightly larger than the corresponding unconditioned mass function. This overshoot peaks at around 5 percent close to the background mass, and approaches zero at low masses. This overshoot occurs because the distribution function we use for the background overdensity is derived for the constant barrier corresponding to the Press–Schechter mass function. In principle, the correct distribution function could be found for the curved barrier corresponding to the Sheth–Tormen mass function, by numerical solution of the barrier crossing problem. Given the small magnitude of the overshoot we do not seek to find the correct background distribution here, but instead simply recalibrate the parameters of the Sheth–Tormen mass function such that when environmentally averaged it agrees well with  $N$ -body measures of the mass function. Because of the overshoot, there is a discontinuity in the mass function at  $M_e$ . To correct for this we multiply the mass function for  $M > M_e$  by a fixed factor to remove the discontinuity.

While the peak-background split can be used to accurately predict the environmental dependence of the halo mass function, it predicts no environmental dependence in halo formation histories. This is easy to see in excursion set theory, as the behaviour of the trajectory on scales larger than the collapse scale (at variances smaller than the collapse variance) is uncorrelated with the behaviour of the trajectory on smaller scales (larger variances).<sup>4</sup>

In the merger tree algorithm of Parkinson et al. (2008) the quantity  $w(t) = \delta_c(t)/D(t)$ , where  $\delta_c$  is the critical linear theory overdensity for collapse at time  $t$  and  $D(t)$  is the linear theory growth factor,

plays the role of a ‘time’ variable, and sets the threshold which perturbations must reach to collapse and form haloes. Therefore, to introduce an environmental dependence into this algorithm, we make an empirical modification such that

$$w(t; \delta_c) = \delta_c(t) D^{\alpha \delta_c - 1}(t), \quad (8)$$

where  $\alpha$  is a parameter to be determined. This introduces a dependence on the mapping between merger tree ‘time’,  $w(t)$ , and true cosmic time,  $t$ , which depends on the environmental overdensity.

The choice of the radius used to define environment,  $R_e$ , is somewhat arbitrary, but may influence the value of  $\alpha$  in our model. We choose to use  $R_e = 5 h^{-1}$  Mpc (where  $h = H_0/100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ) as this is larger than the collapse scale for the vast majority of haloes,<sup>5</sup> while still being a relatively ‘local’ measure of environment. This point is discussed further in Section 3.2.

### 2.2.3 Calibration

Several predictions of our model – specifically the environment-averaged halo mass function, halo progenitor mass functions, and the distribution of halo concentrations as a function of halo mass – will be altered as a consequence of our introduction of environmental dependences into both the halo mass function,  $n(M; \delta_c)$ , by adopting the peak-background split model, and the halo collapse threshold,  $w(t; \delta_c)$ . Therefore, we recalibrate the parameters of these models to match measurements from  $N$ -body simulations. This calibration is carried out by running a Markov Chain Monte Carlo (MCMC) simulation. Our approach follows that of Benson (2017a) in detail, including utilizing the same MCMC algorithm and convergence criteria. Benson (2017a) constrained the parameters of the Sheth & Tormen (2002) mass function, and the Parkinson et al. (2008) merger tree algorithm to match the mass function and progenitor mass functions measured from the MultiDark Planck  $N$ -body (MDPL2) simulation (Klypin et al. 2016), after having removed splashback haloes to avoid double-counting. We constrain these parameters in the same way here, except that we additionally average halo and progenitor mass functions over environment, and include the parameter  $\alpha$  as an additional parameter in our MCMC simulation. Furthermore, when constructing progenitor mass functions we adopt the error distribution of Trenti et al. (2010) when convolving the intrinsic halo mass function with the expected error distribution. As the Trenti et al. (2010) errors are dominated by the effects of missing structure below the resolution limit, we expect that errors in halo masses will be correlated across time. This correlation should be measured directly from  $N$ -body simulations, using a methodology similar to that adopted by Trenti et al. (2010). Lacking such an analysis, we instead adopt a simple model to describe the covariance between the halo masses of parent and progenitor haloes. Given a parent halo, ‘0’, and progenitor halo, ‘1’ we assume that their masses are drawn from a distribution with covariance matrix  $\mathbf{S}$ . The diagonal elements of this matrix are set

<sup>5</sup>For haloes whose Lagrangian radius exceeds  $R_e$  we assume no environmental dependence (i.e. we assume  $\delta_e = 0$  for these haloes). We could, instead, select a  $\delta_e$  in this case, conditioned on the fact that the density field exceeded the threshold for collapse,  $\delta_c$ , at some larger mass scale (corresponding to the mass of the halo), and use this as the environmental overdensity in our model. However, this seems unreasonable as it would modify the formation history after the scale used to define the environment had already collapsed. As such, our model makes no statement about the environmental dependence of formation history for haloes in this regime.

<sup>4</sup>This is true under the usual assumption of a sharp- $k$  filter – for other filters the trajectory is non-Markovian (Maggiore & Riotto 2010).

to  $\sigma_0^2$  and  $\sigma_1^2$ , where  $\sigma_0$  and  $\sigma_1$  are the root variances in the parent and progenitor halo mass according to Trenti et al. (2010). For the off-diagonal elements we assume that the covariance is  $C\sigma_0\sigma_1$ , with

$$C = C_0 \left( \frac{M_1}{M_0} \right)^{C_m} \left( \frac{a_1}{a_0} \right)^{C_a}, \quad (9)$$

where  $a$  is expansion factor,  $M$  is halo mass, and  $C_0$ ,  $C_m$ , and  $C_a$  are parameters to be determined. We include these three nuisance parameters in our MCMC simulation, adopting broad, uniform priors in the interval 0–1 for  $C_0$  and 0–2 for both  $C_m$ , and  $C_a$ , and will marginalize over them in our final analysis.

The concentration model of Ludlow et al. (2016) has two free parameters,  $f$  and  $C$ . We include the parameters  $C$  and  $f$  in our MCMC simulation, and constrain the model to match the distribution of concentrations found in the COCO simulations (see Section 2.1.1). Specifically, we construct histograms of the concentrations of haloes in seven logarithmically spaced mass bins spanning the range  $9.4 < \log_{10}(M/M_\odot) < 12.4$ , and adopt a log-likelihood for each mass bin of

$$\log \mathcal{L} = -\frac{1}{2} \Delta \mathbf{C}^{-1} \Delta^T, \quad (10)$$

where  $\Delta$  is a vector of differences between the  $N$ -body and model concentration histograms, and  $\mathbf{C}$  is a covariance matrix. The covariance matrix is taken to be the sum of that of the  $N$ -body halo histogram, and that of the model halo histogram, both assuming Poisson counting statistics. When constructing the histogram of model halo concentrations we exclude haloes that have more than doubled their mass in the last 1.25 crossing times (3.7 Gyr; see Section 2.1.1; which we find excludes 2.6 per cent of haloes), and smooth by a Gaussian with width chosen to match the expected uncertainty in  $N$ -body halo concentration estimates (see equation 5) – this smoothing is taken into accounting when computing the covariance matrix.

For the parameters of the Sheth & Tormen (2002) halo mass function and Parkinson et al. (2008) merger tree algorithm we adopt the same priors as were used by Benson (2017a). For the parameter  $\alpha$  we adopt a uniform prior in the range 0.0–0.4 – we expect this parameter to be positive to induce early formation times in higher density environments (see Section 3), and we find that values larger than 0.4 can lead to  $w(t; \delta_c)$  becoming an increasing function of  $t$ , which would result in progenitor haloes collapsing *after* their descendants and so is physically impossible. For the parameters  $C$  and  $f$  we adopt uniform priors of (100, 800) and (0.01, 0.10), respectively. These priors are broad and include the values found by Ludlow et al. (2016) to match results for both cosmological  $N$ -body haloes,  $(C, f) = (400, 0.02)$ , and simple spherical collapse models,  $(C, f) = (650, 0.02)$ . Furthermore, Ludlow et al. (2016) find that  $C$  and  $f$  correlate with the slope of the  $\langle \rho_{-2} \rangle - \rho_{\text{crit}}(z_{-2})$  relation, when  $f$  approaches the mass  $M_{-2}/M_{200}$  (which is typically 0.1–0.2). This correlation is currently neglected in our modelling, which further motivates our choice to restrict  $f$  to values less than 0.1.  $f$  restricted to small values, e.g.  $f = (0.01, 0.1)$

### 3 RESULTS

The parameter constraints derived from our MCMC simulation are listed in Table 1, along with previously determined values. The maximum likelihood values of the Sheth & Tormen (2002) and Parkinson et al. (2008) model parameters are significantly shifted by the inclusion of environmental dependences, but the resulting halo and progenitor mass functions remain equally good matches

**Table 1.** Maximum posterior probability values and the region containing 68 per cent of the posterior probability for all parameters used in our model (the nuisance parameters  $C_0$ ,  $C_m$ , and  $C_a$  are not listed). For each parameter, we also list the previously determined value where available.

	This work	Previous work	
$a$	+0.791 <sup>+0.074</sup> <sub>−0.023</sub>	+0.874 <sup>+0.005</sup> <sub>−0.005</sub>	Benson (2017a)
$p$	+0.218 <sup>+0.045</sup> <sub>−0.130</sub>	−0.031 <sup>+0.005</sup> <sub>−0.005</sub>	Benson (2017a)
$A$	+0.302 <sup>+0.013</sup> <sub>−0.003</sub>	+0.332 <sup>+0.0002</sup> <sub>−0.0002</sub>	Benson (2017a)
$G_0$	+0.591 <sup>+0.009</sup> <sub>−0.010</sub>	+0.635 <sup>+0.011</sup> <sub>−0.0002</sub>	Benson (2017a)
$\gamma_1$	+0.253 <sup>+0.019</sup> <sub>−0.019</sub>	+0.176 <sup>+0.002</sup> <sub>−0.015</sub>	Benson (2017a)
$\gamma_2$	+0.124 <sup>+0.017</sup> <sub>−0.019</sub>	+0.041 <sup>+0.001</sup> <sub>−0.009</sub>	Benson (2017a)
$\alpha$	+0.077 <sup>+0.017</sup> <sub>−0.009</sub>	–	
$C$	+625.0 <sup>+77.0</sup> <sub>−15.0</sub>	+400.0	Ludlow et al. (2016)
$f$	+0.061 <sup>+0.020</sup> <sub>−0.005</sub>	+0.020	Ludlow et al. (2016)

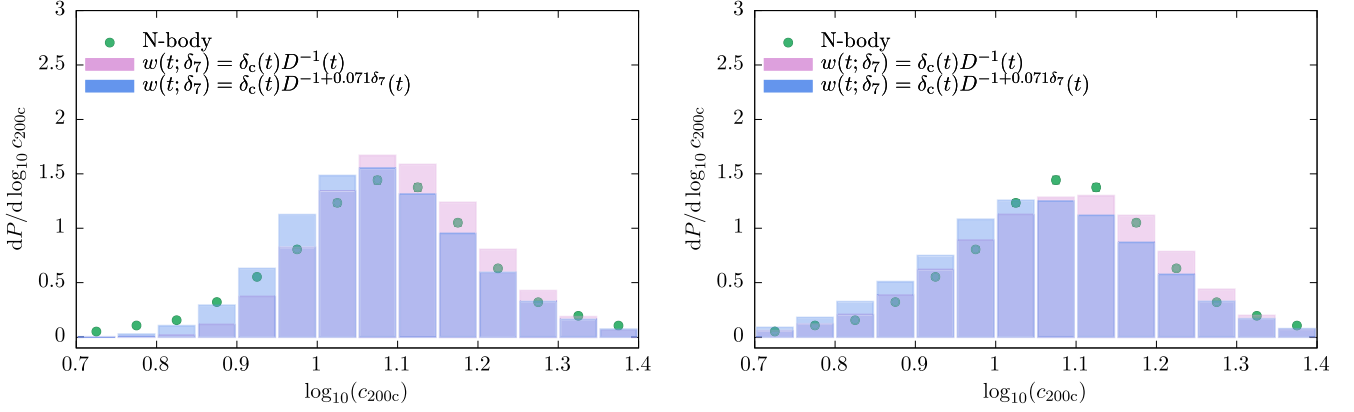
**Table 2.** Mean and scatter in  $\log_{10} c_{200c}$  for different concentration models, and as measured from COCO  $N$ -body haloes, for haloes in the mass range  $9.402 \leq \log_{10}(M_{200c}/M_\odot) < 9.902$  at  $z = 0$ .

Model	$\langle \log_{10} c_{200c} \rangle$	$\sigma_{\log_{10} c_{200c}}$
COCO ( $N$ -body)	1.084	0.152
Ludlow et al. (2016), $\alpha = +0.000$	1.104	0.103
Ludlow et al. (2016), $\alpha = +0.071$	1.077	0.114
Ludlow et al. (2014), $\alpha = +0.000$	1.077	0.133
Ludlow et al. (2014), $\alpha = +0.071$	1.051	0.140

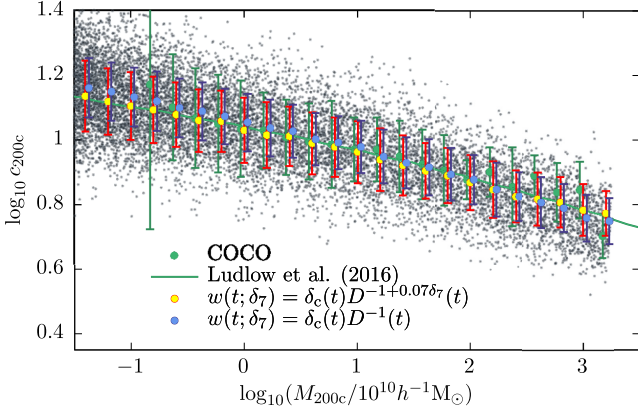
to the MDPL2  $N$ -body measurements. We find that the parameter  $\alpha$  is strongly constrained to be non-zero, indicating that the  $N$ -body data are better fit by a model with environmental dependence in halo merger rates. For the parameters  $(C, f)$  of the Ludlow et al. (2016) model, we find best-fitting values higher than those reported by Ludlow et al. (2016). This may reflect some difference in the structure of our merger trees compared to those extracted from  $N$ -body simulations, but we note that the posterior distribution over  $(C, f)$  shows that these two parameters are strongly degenerate in the direction that includes the  $(C, f) = (400, 0.02)$  results favoured by Ludlow et al. (2016).

### 3.1 Concentrations

Fig. 3 shows the distribution of concentrations,  $c_{200c}$ , for haloes in the mass range  $9.402 \leq \log_{10}(M_{200c}/M_\odot) < 9.902$  at  $z = 0$ . The model of Ludlow et al. (2014) (right-hand panel) produces a distribution that matches  $N$ -body results (shown by green points) quite well when no dependence on environmental density is included in  $w(t; \delta_7)$ . However, as shown by Ludlow et al. (2016), the Ludlow et al. (2014) has significant failings, such as failing to reproduce the concentrations of haloes in WDM models. The model of Ludlow et al. (2016) (left-hand panel) predicts a distribution that is too narrow when  $w(t)$  is independent of environmental density, but matches the  $N$ -body results reasonably well when  $w(t; \delta_7) = \delta_c(t) D^{+0.071\delta_7-1}(t)$ . The improvement is most noticeable in the low-concentration tail of the distribution where the model with no environmental dependence underpredicts the frequency of these low-concentration haloes. To quantify the improvement in the match to  $N$ -body data we can consider the log-likelihood of the COCO concentration distributions used in our calibration. Including the environmental dependence in our model increases the log-likelihood of these distributions by  $\Delta \mathcal{L} \approx 221$ . Table 2



**Figure 3.** Distribution of concentration,  $c_{200c}$ , for haloes in the mass range  $9.402 \leq \log_{10}(M_{200c}/M_{\odot}) < 9.902$  at  $z = 0$ . Green points show results for concentration measured directly from  $N$ -body haloes (see Section 2.1.3). Histograms show results obtained using the models of Ludlow et al. (2016) (left-hand panel) and Ludlow et al. (2014) (right-hand panel). Pink histograms indicate results when the collapse threshold,  $w(t)$ , is independent of environment, while blue histograms indicate results when  $w(t)$  scales with environmental density as indicated in the panels.



**Figure 4.** The concentration–mass relation. Points indicate the mean  $\log_{10}c_{200c}$ , while error bars show the root variance in  $\log_{10}c_{200c}$ . Green points are measured from dark matter halo profiles in the COCO-COLD  $N$ -body simulation (see Section 2.1.3), while yellow and blue points are from this work using the algorithm of Ludlow et al. (2016), with and without an environmental dependence in  $w(t)$  (as shown in the figure). Small grey points show individual haloes from this work with  $w(t; \delta_7) = \delta_c(t)D^{+0.071\delta_7-1}(t)$ . The green line indicates the concentration–mass relation computed by Ludlow et al. (2016).

summarizes the mean and scatter in concentration in a narrow range of halo mass as measured from the COCO  $N$ -body haloes, and predicted by the Ludlow et al. (2014, 2016) models applied to our merger trees. Even with an environmental dependence introduced into  $w(t; \delta_7)$  the Ludlow et al. (2016) model is unable to produce a scatter quite as large as that measured for  $N$ -body haloes. This is largely driven by the tail of low-concentration haloes seen in the  $N$ -body simulation, which is not matched by the Ludlow et al. (2016) model applied to our excursion set-derived merger trees.

Fig. 4 shows the concentration–mass relation as measured from  $N$ -body halo profiles, and as predicted by the model of Ludlow et al. (2016).

### 3.2 Formation epochs

$N$ -body haloes show a correlation between environmental overdensity and formation time (defined as the time at which a given halo had first assembled 50 per cent of its final mass into a single

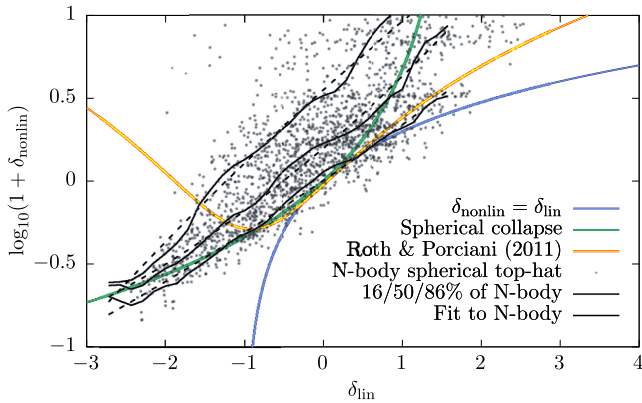
progenitor halo) as shown, for example, by Zehavi et al. (2018). While our model of environmental dependence was constructed to explain the scatter in halo concentrations at fixed mass, we can also ask whether it produces the correct correlation between environment and formation time.

To assess this correlation, and to compare to the measurements of Zehavi et al. (2018), we must account for the fact that our environmental overdensity,  $\delta_e$ , is the linear theory overdensity in the Lagrangian volume surrounding each halo, while that measured by Zehavi et al. (2018) is the non-linear overdensity in an Eulerian volume surrounding each halo. While a fitting function relating these quantities has been proposed by Roth & Porciani (2011) we find that it does not well describe the relation here as we are interested in the relation for special points in the density field, namely the locations of haloes. We therefore measure the linear theory Lagrangian, and non-linear Eulerian overdensities around particles in the Millennium  $N$ -body simulation (as was used by Zehavi et al. 2018).

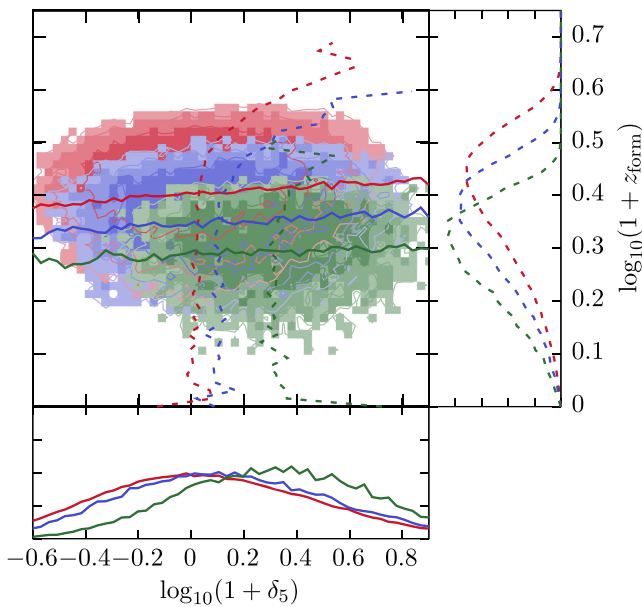
Zehavi et al. (2018) define environment based on a  $5 h^{-1}$  Mpc Gaussian smoothing of the  $N$ -body particle field (computed in  $2 h^{-1}$  Mpc cells), which they label  $\delta_5$ . Based on their fig. 2, the median value of  $1 + \delta_5$  is around 2 for most halo masses. As such, these regions have collapsed in radius by a factor of approximately  $2^{1/3}$  relative to their initial, Lagrangian volume. We therefore choose to use a radius  $R_e = 2^{1/3}5 h^{-1}$  Mpc =  $6.3 h^{-1}$  Mpc to define environment in our model.<sup>6</sup>

Using the particle distribution in the Millennium Simulation we measure the Lagrangian overdensity in spheres of radius  $R_e$  in the initial conditions around each halo’s most-bound particle, and extrapolate this to  $z = 0$  assuming linear perturbation theory. We then determine  $\delta_5$  in the  $z = 0$  density field for the same particles. The results are shown in Fig. 5. We model the distribution of non-linear, Eulerian overdensity at each linear, Lagrangian overdensity

<sup>6</sup>We did not alter the value of  $\alpha$  used in constructing these merger trees, even though it was constrained using  $R_e = 5 h^{-1}$  Mpc rather than  $R_e = 6.3 h^{-1}$  Mpc. Since the dispersion in  $\delta_e$  will depend on  $R_e$  it is possible that  $\alpha$  should be recalibrated for each  $R_e$ . Alternatively, it is possible that one specific value of  $R_e$  is the optimal choice (given some suitable metric to judge optimality) for our model and should always be used to define environment (in which case comparisons to measures using a different definition of environment become more difficult). We leave a full investigation of these issues to a future work.



**Figure 5.** Mapping from Lagrangian, linear overdensity to non-linear Eulerian overdensity. The blue line shows the one-to-one relation, the green curve shows the expectation from spherical collapse, while the yellow line shows the fitting function of Roth & Porciani (2011). Grey points show measurements of overdensity around individual particles in an  $N$ -body simulation. Solid black lines show the 16/50/84 percentiles of the distribution of these points, while the dashed black lines show polynomial fits to these percentiles.



**Figure 6.** Joint and marginalized distributions of non-linear, Eulerian density contrast and formation epoch for haloes in three narrow ranges of  $\log_{10}(M_{\text{FOF}}/h^{-1} M_{\odot})$ : 11.0–11.2 (red), 12.0–12.2 (blue), and 13.0–13.2 (green). Formation epoch is defined as the epoch at which the primary progenitor of the  $z = 0$  halo first reaches half of the final mass. Overlaid on the joint distribution are medians of formation epoch (solid lines) and of density contrast (dashed lines). This figure is intended to be compared directly to that of Zehavi et al. (2018).

as a lognormal distribution, and determine the mean and dispersion of this distribution from the measured points. Then, for each halo in our model, we take its assigned linear, Lagrangian overdensity, and draw a non-linear, Eulerian overdensity from the appropriate lognormal distribution.

Fig. 6 shows joint and marginalized distributions of density contrast and formation epoch in three different halo mass ranges, and can be compared directly to fig. 2 of Zehavi et al. (2018) in which these same distributions were measured directly from

the Millennium Simulation (Springel et al. 2005). To construct these distributions from our model we generated large samples of merger trees as described in Section 2.2, matching the cosmological parameters and power spectrum of the Millennium Simulation. The resulting distributions of overdensity and formation epoch are in qualitative agreement with those found by Zehavi et al. (2018). In particular, the marginalized distributions of overdensity agree well with Zehavi et al. (2018), and are in good quantitative agreement for the position of the mode of the distribution, and the width. Marginalized distributions for formation epoch also agree well. Finally, the joint distribution shows a correlation between formation epoch and environmental overdensity that agrees moderately well with that found by Zehavi et al. (2018) – without our overdensity-dependent modification of the collapse threshold no correlation would exist. A more quantitative comparison of the effects of environment on assembly history in excursion set and  $N$ -body merger trees is currently limited by the different definitions of environment that naturally apply in these two approaches.

### 3.3 Progenitor mass functions

The original Parkinson et al. (2008) algorithm for generating merger trees was calibrated to match progenitor mass functions measured in  $N$ -body simulations. Correctly reproducing progenitor mass functions is crucial for merger trees as this ensures that the evolution of the overall halo mass function is also correctly reproduced. Because of this requirement we included progenitor mass functions in our calibration process as described in Section 2.2.3. Here, we check whether the inclusion of our model for environmental dependence in halo merger rates affects how well our merger trees match progenitor mass functions from  $N$ -body simulations.

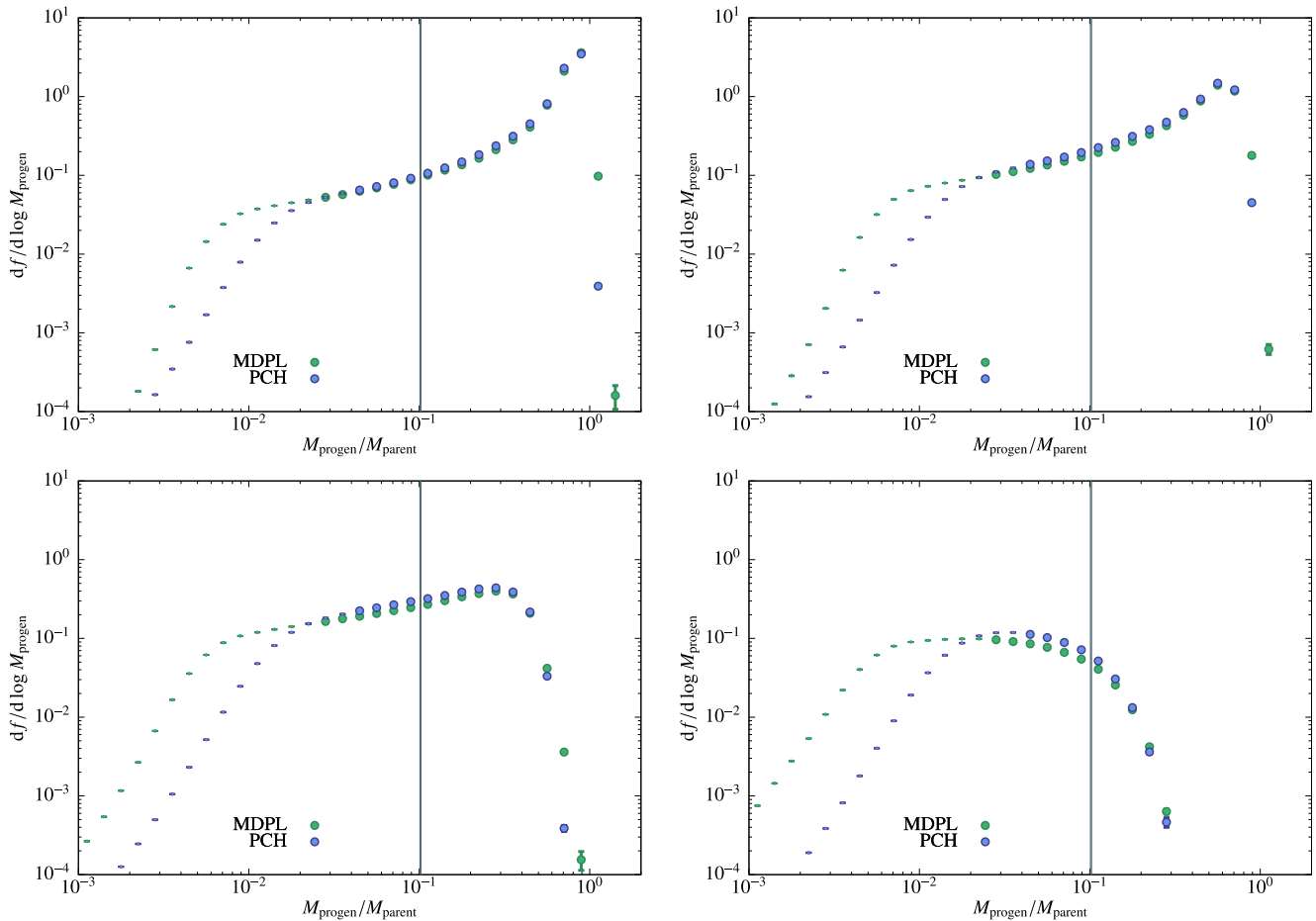
Fig. 7 shows progenitor mass functions for  $z = 0$  haloes of mass  $10^{12.8} M_{\odot}$  averaged across all environments, at four different redshifts ( $z = 0.49$  top-left,  $z = 1.03$  top-right,  $z = 2.03$  bottom-left,  $z = 4.04$  bottom-right). Green points show the progenitor mass function from the MDPL simulation that was used as our calibration data set, while blue points show the results from our model. The good agreement between the  $N$ -body progenitor mass functions and those from our model remains intact after adding in our model of environmental dependence.

As a further test of our model for environmental dependence in merger trees we can examine the progenitor mass functions of haloes selected by environment. Specifically, we compare results from our model to those from the ‘milli-Millennium’ Simulation data base<sup>7</sup> since it provides full particle information allowing us to trace the Lagrangian position of each halo to the initial conditions and measure its Lagrangian, linear density that is directly comparable to the definition of environmental overdensity used in this work. This particle information is not available for the full Millennium Simulation, which limits the statistical power of this comparison. Therefore, we take a relatively broad bin in parent halo mass ( $\pm 0.5$  dex around a central mass), and split the sample into haloes that are in underdense region and those in overdense regions (we also compare results for haloes selected regardless of environment). We then construct the same progenitor mass functions from our model.

We find that the effects of environment on the progenitor mass function are generally quite small in both the  $N$ -body simulation and in our model. The effect of environment is most pronounced

<sup>7</sup><http://gavo.mpa-garching.mpg.de/MyMillennium/MyDB>





**Figure 7.** Progenitor mass functions for  $z = 0$  haloes of mass  $10^{12.8} M_{\odot}$  averaged across all environments, at four different redshifts ( $z = 0.49$  top-left,  $z = 1.03$  top-right,  $z = 2.03$  bottom-left,  $z = 4.04$  bottom-right). Green points show the progenitor mass function from the MDPL simulation that was used as our calibration data set, while blue points show the results from our model.

at high redshifts. Therefore, in Fig. 8 we show the results of this comparison at  $z = 3.87$  of parent haloes in the range  $10^{12} h^{-1} M_{\odot} \leq M_{\text{vir}} < 10^{13} h^{-1} M_{\odot}$ , selected at  $z = 0$ .  $N$ -body results are shown by points with error bars, while results from our model are shown by lines. Yellow points and line correspond to the progenitor mass function for parent haloes selected independent of environment, while green/blue points and lines correspond to the progenitor mass function for parent haloes selected to be in under/overdense regions, respectively. The vertical grey line corresponds to 20 particles in the Millennium Simulation for  $M_{\text{parent}} = 10^{12.5} h^{-1} M_{\odot}$  – results will be affected by the simulation resolution close to this line.

While the  $N$ -body results are noisy, the trend with environmental overdensity shows a direction and magnitude that closely matches that predicted by our model.

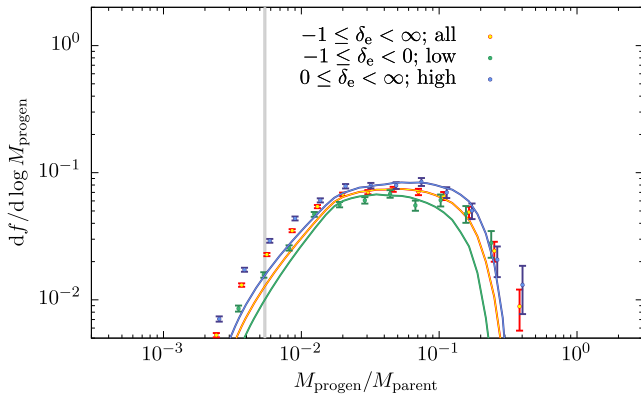
#### 4 DISCUSSION AND SUMMARY

We have applied the model of Ludlow et al. (2016) to merger trees generated via the algorithm of Parkinson et al. (2008) to predict the concentration of haloes in those merger trees from their formation histories. We find that, with the original Parkinson et al. (2008) algorithm this model results in insufficient scatter in concentration at fixed halo mass. Since the Ludlow et al. (2016) model predicts the correct amount of scatter when applied to merger trees extracted

from  $N$ -body simulations this suggests that the failure lies within the merger tree construction algorithm itself. We hypothesize that the missing ingredient is the effects of environment on halo formation times, and introduce a simple, empirical model for this effect to the Parkinson et al. (2008) algorithm. With this modification our model correctly predicts the mean concentration as a function of halo mass, and comes closer to matching the measured scatter, while simultaneously matching the statistics of progenitor halo mass functions.

We also show that this simple model for the influence of environment on halo formation reproduces, at least qualitatively, the correlation between environment and formation epoch measured in  $N$ -body simulations. This brings some of the important effects of assembly bias into the extended Press–Schechter framework for merger tree construction.

Our model is based solely upon the linear density field, and so has no knowledge of the non-linear density field. It may not therefore capture any effects related to the non-linear density field (e.g. in cases where a single halo dominates the local density field). The degree to which this approximation (i.e. considering only the linear density field) is reasonable can be judged by the ability of our model to capture environmental correlations and trends – some of which we have demonstrated in this work. We expect that there will be effects related to the non-linear collapse of the primary halo, and possibly other details of the local density field, which we do

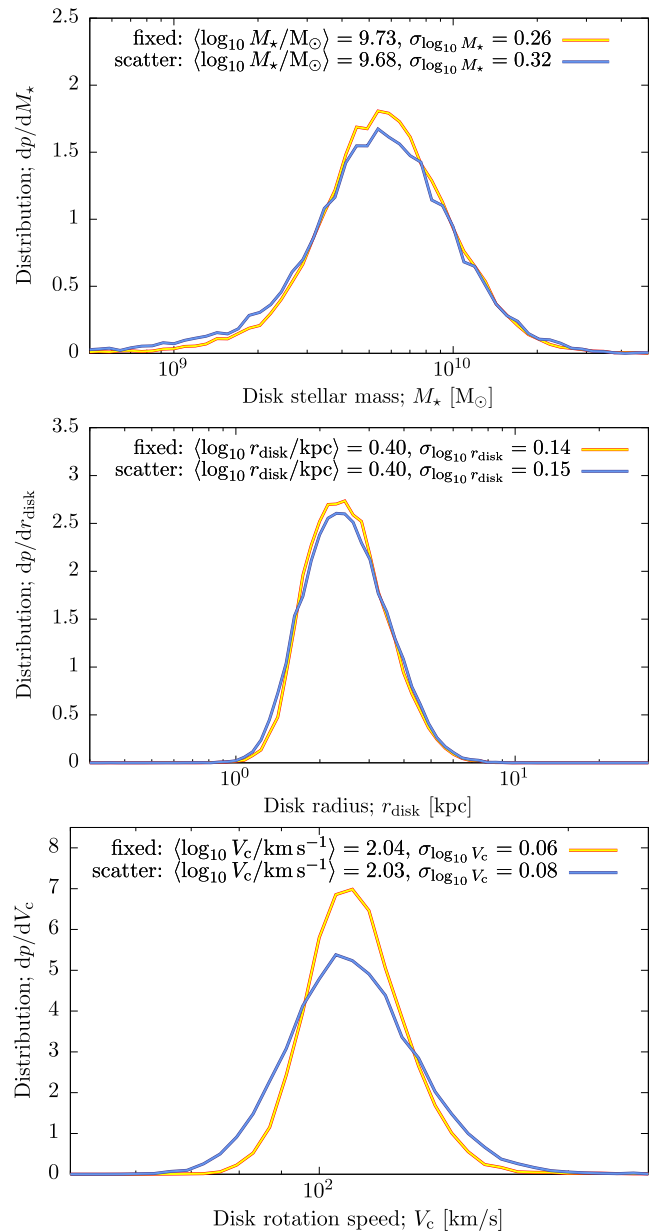


**Figure 8.** The progenitor mass function at  $z = 3.87$  of parent haloes in the range  $10^{12} h^{-1} M_{\odot} \leq M_{\text{vir}} < 10^{13} h^{-1} M_{\odot}$ , selected at  $z = 0$ .  $N$ -body results are shown by points with error bars, while results from our model are shown by lines. Yellow points and line correspond to the progenitor mass function for parent haloes selected independent of environment, while green/blue points and lines correspond to the progenitor mass function for parent haloes selected to be in under/overdense regions, respectively. The vertical grey line corresponds to 20 particles in the Millennium Simulation for  $M_{\text{parent}} = 10^{12.5} h^{-1} M_{\odot}$  – results will be affected by the simulation resolution close to this line.

not capture in our model. Nevertheless, we have shown that this simple model does capture some key aspects of the environmental dependence of halo assembly.

This model has important consequences for semi-analytic models of galaxy formation built on such merger trees. The concentrations of haloes are a key factor in determining the sizes of galaxies (Jiang et al. 2018), and determine key observable properties such as rotation curves. To illustrate the effects of this work we utilize the GALACTICUS model (Benson 2012) to predict the distribution of disc stellar masses, scale lengths, and rotation speeds<sup>8</sup> for central, disc-dominated galaxies occupying haloes in the mass range  $1-3 \times 10^{12} M_{\odot}$  at  $z = 0$ . The resulting distributions are shown in Fig. 9, which also shows the mean and standard deviation of each distribution. The yellow line shows the result obtained when the concentration of haloes does not include any scatter (i.e. it is set to the mean concentration as a function of halo mass and redshift, specifically using the fitting formula given by Ludlow et al. 2016). The blue line shows the result when scatter in concentration is included using the model of Ludlow et al. (2016) with parameter values as calibrated in this work. Incorporating scatter into halo concentrations has little effect on the mean of the distributions but does add to the scatter by a small but non-negligible increase. This is most readily apparent in the case of rotation speeds, in which the scatter is increased from 0.06 to 0.08 dex. Semi-analytic

<sup>8</sup>In GALACTICUS galaxy sizes and rotation speeds are solved for assuming each disc is in rotationally supported equilibrium in the combined gravitational potential of the dark matter halo and the baryonic content of the galaxy itself. The sizes and rotation speeds will therefore depend on the halo concentration through its effect on the dark matter density profile and potential. Stellar masses of galactic discs will also be affected by concentration since they depend on the sizes and rotation speeds of those discs (via the dependence of the algorithms for star formation and feedback on those quantities). See Benson (2012) for further details.



**Figure 9.** The distribution of galactic disc stellar masses (upper), scale lengths (middle), and rotation speeds (at one scale length; lower) for disc-dominated, central galaxies occupying haloes in the mass range  $1-3 \times 10^{12} M_{\odot}$  at  $z = 0$ . The yellow line shows the result obtained when the concentration of haloes does not include any scatter (i.e. it is set to the mean concentration as a function of halo mass and redshift, specifically using the fitting formula given by Ludlow et al. 2016). The blue line shows the result when scatter in concentration is included using the model of Ludlow et al. (2016) with parameter values as calibrated in this work. The mean and standard deviation of the distribution is shown for each case.

models that assign concentrations based on the mean concentration–mass relation will therefore miss a significant contribution to the scatter in these observables – as far as we are aware all other semi-analytic models either utilize a mean (or median) concentration–mass–redshift relation to assign concentrations to haloes, or else do not make use of concentrations in deriving galaxy properties. Models that attempt to include this scatter by randomly drawing concentrations from the measured distribution of concentrations

will, by construction, incorporate the scatter in observable quantities, but will miss any correlation with other galaxy properties that may be influenced by formation history.

## ACKNOWLEDGEMENTS

AL acknowledges financial support from the Australian Research Council (project number FT160100250). We thank Nick Crook, whose initial investigation of applying the Ludlow et al. (2016) model to PCH merger trees motivated this work. SC acknowledges the support of the Science and Technology Facilities Council (ST/P000541/1).

## REFERENCES

- Benson A. J., 2012, *New Astron.*, 17, 175  
 Benson A. J., 2017a, *MNRAS*, 467, 3454  
 Benson A. J., 2017b, *MNRAS*, 471, 2871  
 Benson A. J., Borgani S., De Lucia G., Boylan-Kolchin M., Monaco P., 2012, *MNRAS*, 419, 3590  
 Benson A. J. et al., 2013, *MNRAS*, 428, 1774  
 Bond J. R., Myers S. T., 1996, *ApJS*, 103, 1  
 Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, *ApJ*, 379, 440  
 Borzyszkowski M., Porciani C., Romano-Díaz E., Garaldi E., 2017, *MNRAS*, 469, 594  
 Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001, *MNRAS*, 321, 559  
 Castorina E., Sheth R. K., 2013, *MNRAS*, 433, 1529  
 Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, *MNRAS*, 319, 168  
 Correa C. A., Wyithe J. S. B., Schaye J., Duffy A. R., 2015, *MNRAS*, 452, 1217  
 Dalal N., White M., Bond J. R., Shirokov A., 2008, *ApJ*, 687, 12  
 Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, *ApJ*, 292, 371  
 Desjacques V., 2008, *MNRAS*, 388, 638  
 Einasto J., 1965, *Trudy Astrofizicheskogo Inst. Alma-Ata*, 5, 87  
 Gao L., Springel V., White S. D. M., 2005, *MNRAS*, 363, L66  
 Gao L., Navarro J. F., Cole S., Frenk C. S., White S. D. M., Springel V., Jenkins A., Neto A. F., 2008, *MNRAS*, 387, 536  
 Hahn O., Porciani C., Dekel A., Carollo C. M., 2009, *MNRAS*, 398, 1742  
 Hellwing W. A., Frenk C. S., Cautun M., Bose S., Helly J., Jenkins A., Sawala T., Cytowski M., 2016, *MNRAS*, 457, 3492  
 Jiang F. et al., 2018, preprint ([arXiv:1804.07306](https://arxiv.org/abs/1804.07306))  
 Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, *MNRAS*, 457, 4340  
 Lacey C., Cole S., 1993, *MNRAS*, 262, 627  
 Ludlow A. D., Angulo R. E., 2017, *MNRAS*, 465, L84  
 Ludlow A. D., Navarro J. F., Li M., Angulo R. E., Boylan-Kolchin M., Bett P. E., 2012, *MNRAS*, 427, 1322  
 Ludlow A. D. et al., 2013, *MNRAS*, 432, 1103  
 Ludlow A. D., Navarro J. F., Angulo R. E., Boylan-Kolchin M., Springel V., Frenk C., White S. D. M., 2014, *MNRAS*, 441, 378  
 Ludlow A. D., Bose S., Angulo R. E., Wang L., Hellwing W. A., Navarro J. F., Cole S., Frenk C. S., 2016, *MNRAS*, 460, 1214  
 Maggiore M., Riotto A., 2010, *ApJ*, 711, 907  
 Mo H. J., White S. D. M., 1996, *MNRAS*, 282, 347  
 Musso M., Sheth R. K., 2012, *MNRAS*, 423, L102  
 Navarro J. F., Frenk C. S., White S. D. M., 1996, *ApJ*, 462, 563  
 Navarro J. F., Frenk C. S., White S. D. M., 1997, *ApJ*, 490, 493  
 Navarro J. F. et al., 2010, *MNRAS*, 402, 21

- Paranjape A., Hahn O., Sheth R. K., 2018, *MNRAS*, 476, 3631  
 Parkinson H., Cole S., Helly J., 2008, *MNRAS*, 383, 557  
 Power C., Navarro J. F., Jenkins A., Frenk C. S., White S. D. M., Springel V., Stadel J., Quinn T., 2003, *MNRAS*, 338, 14  
 Press W. H., Schechter P., 1974, *ApJ*, 187, 425  
 Roth N., Porciani C., 2011, *MNRAS*, 415, 829  
 Sheth R. K., Tormen G., 2002, *MNRAS*, 329, 61  
 Sheth R. K., Tormen G., 2004, *MNRAS*, 350, 1385  
 Springel V. et al., 2005, *Nature*, 435, 629  
 Trenti M., Smith B. D., Hallman E. J., Skillman S. W., Shull J. M., 2010, *ApJ*, 711, 1198  
 Zehavi I., Contreras S., Padilla N., Smith N. J., Baugh C. M., Norberg P., 2018, *ApJ*, 853, 84

## APPENDIX A: OPTIMIZATION OF THE MERGER TREE CONSTRUCTION ALGORITHM

We use the merger tree construction algorithm of Cole et al. (2000; see also Parkinson et al. 2008). In that algorithm, small steps in ‘time’ (actually in  $w(t) \equiv \delta_c(t)/D(t)$ , where  $\delta_c$  is the critical threshold for spherical collapse and  $D(t)$  is the linear growth factor, which serves the role of a time variable) are taken. These steps,  $\delta w$ , are required to be sufficiently small that the probability,  $P = R\delta w$  (where  $R$  is the rate of branching events per unit interval of  $w$ ), of a branching event in any time-step is small, typically  $P < \epsilon$  with  $\epsilon = 0.1$ . Additionally, the time-steps must be small enough to ensure that subresolution accretion on to the halo is not too large during the time-step, and that the approximations of the Cole et al. (2000) merger rate are valid. In the limit of high branching rate (which occurs when the mass of the branch in question is much larger than the mass resolution), the step will be limited by branching.

We make a small modification to this algorithm. We remove the limit on the time-step due to branching rate, retaining the other limits on the time-step. Then, in each time-step, we find the interval to the next branching event by drawing a value at random from a negative exponential distribution with rate parameter  $R$ . If this interval exceeds the maximum allowed time-step, no branching occurs, and the time-step proceeds. If the interval is less than the maximum allowed time-step, branching occurs at that point. In the regime of high branching rates this approach allows for larger (by a factor  $1/\epsilon$  on average) time-steps to be taken. Note that we do not have to concern ourselves in the subsequent time-step with the fact that no branching occurred in the previous time-step because of the memorylessness nature of the negative exponential distribution. That is, the distribution of branching intervals conditioned on the fact that no branching occurred in the previous time-step is just the same negative exponential distribution.

The increase in speed of the tree building will depend on both the algorithm implementation, and the tree resolution. In the case of the specific implementation of the Cole et al. (2000) algorithm, with a mass resolution equal to  $5 \times 10^{-6}$  times the final halo mass, this optimization results in speed-ups by a factor of 1.2.

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.