Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **500,** 493–505 (2021)
Advance Access publication 2020 October 22

doi:10.1093/mnras/staa3261

An efficient hybrid method to produce high-resolution large-volume dark matter simulations for semi-analytic models of reionization

Yisheng Qiu,[1,2]★ Simon J. Mutch,[1,2] Pascal J. Elahi ,[2,3] Rhys J. J. Poulton ,[2,3] Chris Power [2,3] and J. Stuart B. Wyithe[1,2]

[1]*School of Physics, University of Melbourne, Parkville, VIC 3010, Australia*
[2]*ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia*
[3]*International Centre for Radio Astronomy Research, M468, University of Western Australia, 35 Stirling Highway, Perth, WA 6009, Australia*

Accepted 2020 October 16. Received 2020 October 14; in original form 2020 July 29

ABSTRACT

Resolving faint galaxies in large volumes is critical for accurate cosmic reionization simulations. While less demanding than hydrodynamical simulations, semi-analytic reionization models still require very large *N*-body simulations in order to resolve the atomic cooling limit across the whole reionization history within box sizes $\gtrsim 100\,h^{-1}$ Mpc. To facilitate this, we extend the mass resolution of *N*-body simulations using a Monte Carlo algorithm. We also propose a method to evolve positions of Monte Carlo haloes, which can be an input for semi-analytic reionization models. To illustrate, we present an extended halo catalogue that reaches a mass resolution of $M_{\rm halo} = 3.2 \times 10^7\,h^{-1}\,{\rm M}_\odot$ in a $105\,h^{-1}$ Mpc box, equivalent to an *N*-body simulation with $\sim 6800^3$ particles. The resulting halo mass function agrees with smaller volume *N*-body simulations with higher resolution. Our results also produce consistent two-point correlation functions with analytic halo bias predictions. The extended halo catalogues are applied to the MERAXES semi-analytic reionization model, which improves the predictions on stellar mass functions, star formation rate densities, and volume-weighted neutral fractions. Comparison of high-resolution large-volume simulations with both small-volume and low-resolution simulations confirms that both low-resolution and small-volume simulations lead to reionization ending too rapidly. Lingering discrepancies between the star formation rate functions predicted with and without our extensions can be traced to the uncertain contribution of satellite galaxies.

Key words: methods: numerical – galaxies: high-redshift – dark ages, reionization, first stars.

1 INTRODUCTION

Simulating the epoch of reionization is extremely challenging, with different techniques developed to study different aspects of the problem. For example, high-resolution hydrodynamical simulations (e.g. Wise et al. 2012; Johnson, Dalla Vecchia & Khochfar 2013; Ceverino, Glover & Klessen 2017; Rosdahl et al. 2018) can resolve the faintest galaxies with detailed spatial information on the interstellar media (ISM). These faint sources are found to have non-negligible contributions to reionization (Wise et al. 2014; Katz et al. 2020). However, these simulations are limited to a small volume ($\lesssim 10^3\,h^{-3}\,{\rm Mpc}^3$). At the other extreme, Iliev et al. (2014) presented a study in a $425\,h^{-1}$ Mpc box, and pointed out that at least an $\sim 100\,h^{-1}$ Mpc box is required for the convergence of reionization histories. Other studies use seminumerical calculations of reionization to simulate large volumes (e.g. Greig & Mesinger 2015; Hassan et al. 2016; Park et al. 2019). A disadvantage of these approaches is the absence of a detailed galaxy formation model. While large volumes have been achieved by several hydrodynamical simulations (e.g. Feng et al. 2016; Pillepich et al. 2018), they cannot resolve the faintest sources. The Cosmic Reionisation on Computers project (Gnedin 2014; Gnedin & Kaurov 2014; Kaurov & Gnedin 2015) aims to produce hydrodynamical simulations with both large volume and high spatial resolution, with self-consistent treatment of radiative transfer, gas dynamics, and star formation. They reach an ~ 100 pc spatial resolution in an $\sim 80\,h^{-1}$ Mpc box. However, one shortcoming of hydrodynamical simulations is that they are extremely computationally expensive, and therefore cannot be easily used to explore different model variations.

Semi-analytic galaxy formation models (see Baugh 2006; Somerville & Davé 2015, for reviews) provide a good alternative, and can potentially achieve very high mass resolution in large volumes. They take halo merger trees extracted from *N*-body simulations as an input, and evolve several key baryonic components of galaxies within these haloes. They do not consider hydrodynamic forces or the spatial distribution of the ISM, which limits their predictive power but makes them computationally efficient. One example is the MERAXES semi-analytic model (Mutch et al. 2016), which couples galaxy formation with reionization using 21CMFAST (Mesinger & Furlanetto 2007). Predictions for reionization using MERAXES can be found in Geil et al. (2016).

The mass resolution and the simulation volume of semi-analytic models are determined by the input *N*-body simulations. Predictions of cosmic reionization may require a volume greater than $100^3\,h^{-3}\,{\rm Mpc}^3$. For example, Deep Kaur, Gillet & Mesinger (2020) suggested that a $170\,h^{-1}$ Mpc box is needed for a simulation to

★ E-mail: yishengq@student.unimelb.edu.au

© 2020 The Author(s)
Published by Oxford University Press on behalf of the Royal Astronomical Society

Downloaded from https://academic.oup.com/mnras/article/500/1/493/5935244 by guest on 20 April 2024

predict convergent 21 cm power spectra. At the same time, the main contribution of ionizing photons could be from faint sources (e.g. Liu et al. 2016; Finkelstein et al. 2019; Katz et al. 2020; however, see Naidu et al. 2020). In order to resolve all faint sources and examine their contribution to reionization, semi-analytic models require *N*-body simulations with a very large particle number. This work attempts to overcome this challenging task by augmenting *N*-body halo merger trees using Monte Carlo haloes. The first such method was presented in Benson, Cannella & Cole (2016). We extend their study to $z \geq 5$, and introduce an improvement to make the results satisfy the halo mass function (HMF) of the given *N*-body simulation. Detailed reionization calculations require the spatial distribution of haloes. This work also proposes an approach to assign and evolve the position of Monte Carlo haloes, which can reproduce halo clustering predicted by the *N*-body simulation.

This paper is organized as follows: Our methodology of extending *N*-body halo catalogues is presented in Section 2. Specifically, Section 2.1 describes the *N*-body simulations utilized in this work. Section 2.2 introduces the algorithms to augment *N*-body halo merger trees. We populate and evolve the position of Monte Carlo haloes in Section 2.3, and sample their spin parameter in Section 2.4. Then, in Section 3, we apply the extended halo catalogues to the MERAXES semi-analytic reionization model. Finally, this work is summarized in Section 4.

2 METHODOLOGY

2.1 *N*-body simulations

This work utilizes two boxes from the Genesis *N*-body simulations (Elahi et al., in preparation). We focus on extending the mass resolution of L105N2048, which is a $105\,h^{-1}$ Mpc box, containing 2048^3 particles, with $m_{\rm p} = 1.17 \times 10^7\ h^{-1}\,{\rm M}_\odot$. To calibrate and verify our results, we take advantage of L35N2650, which has a much higher resolution. It contains 2650^3 particles in a $35\,h^{-1}$ Mpc box. The particle mass is $m_{\rm p} = 2.00 \times 10^5\,h^{-1}\,{\rm M}_\odot$. All the simulations are run using GADGET-2 (Springel 2005). Haloes in the simulations are identified using VELOCIRAPTOR (Elahi, Poulton & Canas 2019a; Elahi et al. 2019c), which is a six-dimensional friends-of-friends phase space halo finder. Merger trees are constructed using TREEFROG (Elahi, Poulton & Tobar 2019b; Elahi et al. 2019d). Table 3 provides a summary of the *N*-body halo catalogues used in this work. Throughout the paper, we adopt the mass obtained by summing all particles in a friends-of-friends group as halo mass. The Genesis *N*-body simulations use a cosmology with $h = 0.6751$, $\Omega_{\rm m} = 0.3121$, $\Omega_{\rm b} = 0.0491$, $\Omega_\Lambda = 0.6879$, $\sigma_8 = 0.8150$, and $n_{\rm s} = 0.9653$ (fourth column in table 4 of Planck Collaboration XIII 2016). To be consistent, we adopt this cosmology throughout the paper.

2.2 Augmenting *N*-body merger trees

Our approach to augment *N*-body merger trees mainly follows Benson et al. (2016). The basic idea is to generate Monte Carlo merger trees with the desired mass resolution and compare these with an *N*-body merger tree in the mass range where the simulation is fully reliable. If both trees are similar, as determined by several criteria (described below), Monte Carlo haloes with mass below the simulation resolution are attached to the *N*-body merger tree. This results in a hybrid structure, containing both Monte Carlo and *N*-body haloes, but with the same mass resolution as the Monte Carlo tree.

2.2.1 *Generating Monte Carlo trees*

We adopt the Parkinson et al. (2008) algorithm to generate Monte Carlo merger trees. The algorithm is based on binary splits in small internal time-steps. It employs the conditional mass function (CMF)[1] derived from the Extended Press Schechter (EPS) theory (Bower 1991; Bond et al. 1991; Lacey & Cole 1993) with an additional parametrization to take into account the difference between the EPS theory and *N*-body simulations. The CMF is expressed as

$$f(M_1, z_1|M_2, z_2) = G_0 \left(\frac{\sigma_1}{\sigma_2}\right)^{\gamma_1} \left(\frac{\delta_2}{\sigma_2}\right)^{\gamma_2} f_{\rm EPS}(M_1, z_1|M_2, z_2), \quad (1)$$

where $f_{\rm EPS}(M_1, z_1|M_2, z_2)$ is the CMF given by the EPS theory. We denote $\sigma_1 = \sigma(M_1)$ and $\sigma_2 = \sigma(M_2)$, which are the mass variance of the matter density field linearly extrapolated to $z = 0$ and smoothed by a spherical top-hat filter at M_1 and M_2. The density contrast is defined by $\delta_2 = 1.686/D(z_2)$, where $D(z)$ is the linear growth factor. The free parameters are G_0, γ_1, and γ_2. Parkinson et al. (2008) calibrated these free parameters against the Millennium simulation (Springel et al. 2005) in the mass range between 10^{12} and $10^{15}\,h^{-1}\,{\rm M}_\odot$ and from $z = 0$ to 4. However, in this work, we are interested in growing haloes at $z \geq 5$, and require that the mass resolution of the merger trees reaches the atomic cooling threshold ($\sim 10^7$–$10^8\,h^{-1}\,{\rm M}_\odot$) in order to capture the majority of ionizing sources during the epoch of reionization. Therefore, we recalibrate the parameters against our simulations, which also accounts for updated cosmology.

Following Parkinson et al. (2008), the cost function of the calibration is given by

$$\mathcal{C}(G_0, \gamma_1, \gamma_2) = \sum \left[\log_{10} f_{\rm NS} - \log_{10} f_{\rm MC}\right]^2, \quad (2)$$

where $f_{\rm NS}$ and $f_{\rm MC}$ are the CMFs of the *N*-body and Monte Carlo merger trees, respectively. We estimate $\log_{10} f_{\rm NS}$ from L35N2650 and $\log_{10} f_{\rm MC}$ using samples of 300 Monte Carlo merger trees for each descendant mass M_2. The fitting points calculated from the simulation are shown as black dots in Fig. 1. We employ the particle swarm optimization (Shi & Eberhart 1998) to minimize the cost function. The best-fitting parameters are accepted if they do not change for 100 iterations. Their values are given in Table 1, and the fitting results are illustrated in Fig. 1. Our best-fitting parameters improve the cost function by $\Delta\mathcal{C} \approx -0.6$, compared with Parkinson et al. (2008). However, the best-fitting result is still poor at $z = 5.5$ and 10.1. While a potential improvement is to employ weights for different mass or redshift ranges in the cost function, in Appendix A, we show that this approach cannot significantly improve the fitting results.

2.2.2 *Augmentation algorithm*

The most important and difficult component of the augmentation is to decide whether a Monte Carlo tree is similar to an *N*-body tree. Instead of comparing entire trees, Benson et al. (2016) decompose an *N*-body merger tree into many subbranches, and match only one subbranch every time with Monte Carlo realizations. A subbranch is comprised of one descendant halo and all haloes that directly merge into it. Hereafter, we refer to this structure as a 'simple branch'.

We denote the mass of each progenitor in an *N*-body simple branch as $M_1, M_2, ..., M_n$ with $M_1 > M_2 > \ldots > M_n$, where n is the number of the progenitors, and let $n_{\rm cut}$ be the number of the progenitors whose

[1]The CMF discussed here is defined by the mass fraction distribution (M_1/M_2) as a function of progenitor mass M_1 given the descendant mass M_2.

Downloaded from https://academic.oup.com/mnras/article/500/1/493/5935244 by guest on 20 April 2024

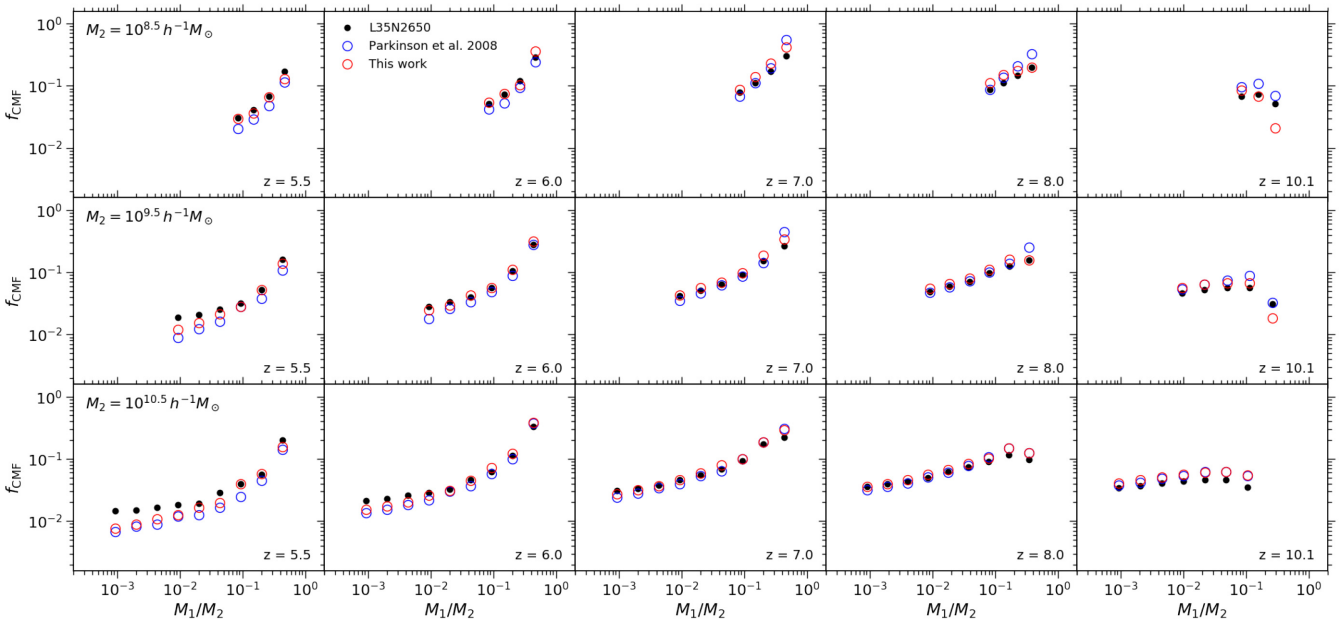

Figure 1. Fitting results of the calibration for the Parkinson et al. (2008) algorithm. The CMFs are defined by $\mathrm{d}f_{\rm CMF}/\mathrm{dln}\,M_1$. Black dots are the fitting data, which are estimated using L35N2650. Red and blue empty circles are the results corresponding to the best-fitting parameters obtained in this work and those used by Parkinson et al. (2008), respectively. The values of the parameters are listed in Table 1.

Table 1. Parameters of the Monte Carlo tree algorithm.

Symbol	Parkinson, Cole & Helly (2008)	This work
G_0	0.57	1.0
γ_1	0.38	0.2
γ_2	− 0.01	− 0.4

mass is above a threshold $M_{\rm cut}$. We use primed symbols for the same quantities of Monte Carlo trees. Benson et al. (2016) match N-body and Monte Carlo simple branches using the following:

(a) $n^{'} \geq n_{\rm cut}$,
(b) for $i = 1, 2, \ldots, n_{\rm cut}$, $|M_i - M_i^{'}| < \xi M_i$,
(c) for $i = n_{\rm cut} + 1, n_{\rm cut} + 2, \ldots, n^{'}$, $M_i^{'} < M_{\rm cut}$,

where ξ is a free parameter and controls the mass precision of the match. Once a match is found, N-body progenitors at $M_{\rm halo} < M_{\rm cut}$ are replaced by Monte Carlo haloes in the same mass range. In the resulting hybrid structure, the descendant halo and progenitors with mass above $M_{\rm cut}$ are from the original simple branch, while progenitors with mass below $M_{\rm cut}$ are additional Monte Carlo haloes from the match.

In practice, relaxing the three matching criteria (a), (b), and (c) is necessary, since there is often no match even for large numbers of Monte Carlo realizations. Benson et al. (2016) increase ξ by a factor of $1 + \epsilon_{\rm mass}$ after $N_{\rm mass}^{\rm limit}$ rejections. However, this only impacts the second condition. We have also found many cases where the first and third conditions are never satisfied. This problem was not reported in Benson et al. (2016), and the reason might be that the mass range investigated in this work is much lower than that in that paper. To address this issue, we increase $M_{\rm cut}$ by a factor of $1 + \epsilon_{\rm cut}$ after $N_{\rm cut}^{\rm limit}$ rejections. We do not allow $M_{\rm cut}$ to be greater than either a maximum mass cut $M_{\rm cut}^{\rm max}$ or the mass of the most massive progenitor. Furthermore, a maximum number of trials $N_{\rm tot}^{\rm limit}$ is employed. Once this number of trials is reached, the algorithm is terminated and returns the input simple branch, with all progenitors below the minimum mass cut $M_{\rm cut}^{\rm min}$ removed. This treatment may remove some N-body haloes without augmentation of Monte Carlo haloes. However, in practice, we find that this situation occurs at a rate that is always smaller than 0.06 per cent for a given snapshot.

N-body merger trees have a special feature that should be taken into account in the comparison with Monte Carlo merger trees. When the halo finder fails to identify the descendant of an N-body halo in the next snapshot, it may try to search for the descendant in later snapshots. Hence, progenitors in an N-body simple branch are not always from the adjacent snapshot. However, this situation never happens for Monte Carlo merger trees. We follow Benson et al. (2016) to resolve the issue. In order to make the trees comparable, for a given N-body simple branch, we manually set all progenitors to be located in the previous snapshot relative to their descendant, and keep their mass unchanged (except for the most massive progenitor, whose mass is interpolated with time).

N-body merger trees typically contain subhaloes, which is an additional feature that Monte Carlo merger trees do not have. Following Benson et al. (2016), we do not consider subhaloes in the tree augmentation. Accordingly, we reconstruct a merger tree that only consists of host haloes from an original N-body tree. The reconstruction proceeds forward with time. If the descendant of an N-body halo is a subhalo, we link it to the host of the subhalo. We neglect the descendant of a subhalo when building the host halo merger trees. We note that the reconstructed trees are only used during the Monte Carlo augmentation. When applying the augmented trees to semi-analytic models, the original links of N-body haloes (including subhaloes) are adopted. We note that these original links may be broken since some N-body haloes are removed by the augmentation algorithm. Section 2.2.3 will discuss the approach to fix the issue.

In reconstructed N-body merger trees, we have found many massive haloes ($M_{\rm halo} \gtrsim 10^{10}\,h^{-1}\,\mathrm{M}_\odot$) that have no progenitors. In the original trees, these haloes only have one subhalo progenitor whose host merges into a different target. When augmenting such haloes, criteria (a) and (b) are automatically satisfied. However,

Downloaded from https://academic.oup.com/mnras/article/500/1/493/5935244 by guest on 20 April 2024

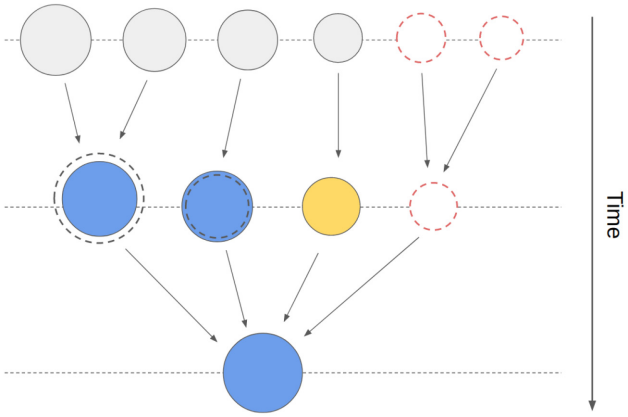

Figure 2. Schematic diagram of augmenting *N*-body halo merge trees. Solid and dashed circles represent *N*-body and Monte Carlo haloes, respectively, with radius proportional to halo mass. The blue and yellow circles form an *N*-body simple branch (defined in Section 2.2.2), which is compared with a Monte Carlo tree. Grey circles also represent *N*-body haloes, but are not considered in this comparison. The algorithm removes haloes with mass below $M_{\rm cut}$, corresponding to the yellow circle. The progenitors of removed haloes will not be taken into account in the next step. Red dashed circles represent Monte Carlo haloes that are added to the *N*-body simulation. The Monte Carlo haloes on the top are grown from its descendant using the Parkinson et al. (2008) algorithm.

we find that forcing criterion (c) overestimates the CMF at $M_{\rm halo} < M_{\rm cut}$. Based on several experiments, we suggest the following modification, which can lead to more consistent CMFs:

(c') if $n > 0$, for $i = n_{\rm cut}+1, n_{\rm cut}+2, \ldots, n'$, $M_i' < M_{\rm cut}$; otherwise, for $i = 1, 2, \ldots, n'$, $M_i' < M_{\rm cut}^{\rm max}$.

Overall, given a simple branch in an *N*-body merger tree, our augmentation algorithm proceeds as follows:

(i) Set $N_{\rm cut}^{\rm trial} = 0$, $N_{\rm mass}^{\rm trial} = 0$, $N_{\rm tot}^{\rm trial} = 0$, $\xi = \xi_0$, and $M_{\rm cut} = M_{\rm cut}^{\rm min}$.
(ii) Whenever a progenitor is at a non-adjacent snapshot of its descendant halo, put it to one previous snapshot of the descendant. If the progenitor is the most massive, interpolate its mass with time.
(iii) Generate a Monte Carlo simple branch using the same configuration as the given *N*-body branch. Increase $N_{\rm tot}^{\rm trial}$ by 1.
(iv) Compare the *N*-body and Monte Carlo simple branches using criteria (a), (b), and (c'). If all three criteria are satisfied, go to step (vii), otherwise, increase the corresponding counters:

(a) If criteria (a) or (c') are false, increase $N_{\rm cut}^{\rm trial}$ by 1.
(b) If criterion (b) is false, increase $N_{\rm mass}^{\rm trial}$ by 1.

(v) Relaxing the criteria when a certain number of rejections is reached:

(a) If $N_{\rm cut}^{\rm trial} = N_{\rm cut}^{\rm limit}$, set $N_{\rm cut}^{\rm trial} = 0$ and increase $M_{\rm cut}$ by a factor of $1+\epsilon_{\rm cut}$. If $M_{\rm cut}$ is greater than $M_{\rm cut}^{\rm max}$ or the mass of the most massive progenitors of the given simple branch, set it to be the minimum of these two values.
(b) If $N_{\rm mass}^{\rm trial} = N_{\rm mass}^{\rm limit}$, set $N_{\rm mass}^{\rm trial} = 0$ and increase ξ by a factor of $1+\epsilon_{\rm mass}$.

(vi) Terminate the algorithm if $N_{\rm tot}^{\rm trial} = N_{\rm tot}^{\rm limit}$, otherwise, go to step (iii).
(vii) Replace progenitors with mass below $M_{\rm cut}$ at the *N*-body simple branch with Monte Carlo haloes in the same mass range.

We apply the augmentation algorithm to every halo in the *N*-body simulation backwards with time, and grow new Monte Carlo haloes using the Parkinson et al. (2008) algorithm. A schematic diagram of the augmentation can be found in Fig. 2.

Table 2. Parameters of the tree augmentation algorithm.

Symbol	Value
ξ_0	0.2
$\epsilon_{\rm mass}$	0.2
$N_{\rm mass}^{\rm limit}$	50
$M_{\rm cut}^{\rm min}$	$100\,m_{\rm p}$ [a]
$M_{\rm cut}^{\rm max}$	$2500\,m_{\rm p}$ [a]
$\epsilon_{\rm cut}$	2.0
$N_{\rm cut}^{\rm limit}$	5
$N_{\rm tot}^{\rm trial}$	1000

Note. [a]For L105N2048, $m_{\rm p} = 1.17\times10^7\,h^{-1}\,{\rm M}_\odot$, which is the particle mass of the simulation.

Free parameters in the algorithm are summarized in Table 2. Ideally, if the CMFs of Monte Carlo merger trees are consistent with the *N*-body simulations, these parameters should primarily affect numerical efficiency and be insensitive to the results. However, as demonstrated in Fig. 1, even with recalibrated parameters, the Parkinson et al. (2008) algorithm is unable to reproduce all parts of the CMFs of the *N*-body merger trees, particularly at the lower mass end and higher redshifts. For this reason, we find that the choice of the algorithm parameters impacts the resulting CMFs. The values listed in Table 2 are chosen based on several experiments in order to obtain better consistency with the *N*-body simulations.

To summarize, our augmentation algorithm builds on the method of Benson et al. (2016) by changing the mass cut $M_{\rm cut}$ dynamically (and introducing the maximum mass cut $M_{\rm cut}^{\rm max}$). When applying the approach of Benson et al. (2016), the result contains only Monte Carlo haloes at $M_{\rm halo} < M_{\rm cut}$ and only *N*-body haloes at $M_{\rm halo} \geq M_{\rm cut}$. In our approach, $M_{\rm cut}$ is not a constant. The minimum and maximum mass cuts become the dividing lines of *N*-body and Monte Carlo haloes. At the mass range in between, halo types are mixed. This modification averages the difference between the merger trees extracted from *N*-body simulations and those generated by the Monte Carlo algorithm.

2.2.3 *Fixing original subhalo trees*

In *N*-body simulations, secondary progenitors may still be self-bound for a certain period after a merger. Such objects are known as subhaloes. During the tree augmentation, we reconstruct *N*-body merger trees that only include host haloes. The reconstructed trees are only used in the comparison of Monte Carlo merger trees. In the application of the extended trees to semi-analytic models, we include subhaloes from the original *N*-body trees. However, the augmentation algorithm removes an *N*-body halo if its mass is below $M_{\rm cut}$, which may break an original subhalo tree. The left-hand panel of Fig. 3 shows such a case, where a subhalo (green circle) merges into a removed halo (dashed circle), and the host of the subhalo merges into a different target. To fix the problem, we redirect the merger target of the subhalo to the descendant of its host halo as shown by the red dashed arrow. An additional case that is worth mentioning is illustrated in the right-hand panel of Fig. 3, where a progenitor host halo of a subhalo is removed during the Monte Carlo augmentation. Consequently, the whole corresponding subhalo tree should also be removed. An easier way to fix the issue is to prevent semi-analytic models from seeding a galaxy in such subhaloes. This treatment is implemented in our application of the extended trees in Section 3.

Downloaded from https://academic.oup.com/mnras/article/500/1/493/5935244 by guest on 20 April 2024

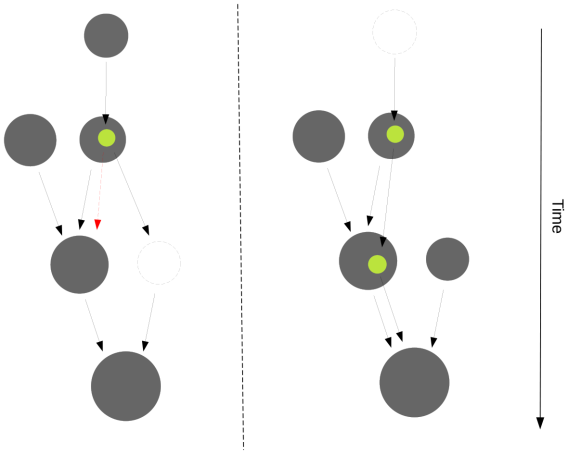

Figure 3. Schematic diagram of fixing subhalo trees. In both panels, black and green circles represent host haloes and subhaloes, respectively. Empty circles correspond to a halo removed by the augmentation algorithm. In the left-hand panel, a subhalo merges into a removed halo, and the host of the subhalo merges into a different target. We fix the problem by redirecting the merger target of the subhalo to the descendant of its host halo as shown by the red dashed arrow. In the right-hand panel, a progenitor host halo of a subhalo is removed by the augmentation algorithm. Consequently, the whole corresponding subhalo tree becomes invalid. This issue can be fixed by preventing semi-analytic models from seeding a galaxy in a subhalo.

2.2.4 *Identifying the complete halo population*

A complete halo population cannot be obtained by applying the augmentation algorithm introduced in Section 2.2.2. The reason is that all Monte Carlo haloes added by the algorithm will eventually merge into an *N*-body halo, while there are unresolved haloes that do not interact with any *N*-body halo at the redshift range covered by the algorithm. This suggests that an additional catalogue of Monte Carlo haloes is required to obtain a complete halo population.

As a specific example in this work, we apply the augmentation algorithm at $z = 5$, adding Monte Carlo haloes to the *N*-body merger trees backwards in time. However, at $z = 5$, the algorithm does not add new haloes that are not resolved (between $M_{\rm cut}^{\rm min}$ and $M_{\rm res}$). In addition, we also miss progenitors of such unresolved haloes in earlier snapshots, resulting in an incomplete halo population. To fix this problem, we create an additional halo catalogue at $z = 5$, using masses and numbers drawn from the HMF of L35N2650. We use interpolation of a histogram instead of a fitting model for the HMF. We then generate trees for these haloes using the Parkinson et al. (2008) algorithm. Hereafter, Monte Carlo haloes generated by the augmentation algorithm are labelled as MC-I, while those in the additional catalogue are referred to as MC-II.

2.2.5 *Applying to N-body simulations*

We apply the approach introduced in the proceeding sections to augment the *N*-body merger trees of L105N2048 from $z = 5$ to 20. We choose three levels of mass resolution: $M_{\rm res} = 1.4 \times 10^8$, 5.7×10^7, and $3.2 \times 10^7\, h^{-1}\,{\rm M}_\odot$, corresponding to the atomic cooling threshold at $z = 5$, 10, and 15, respectively. These three extended halo catalogues are labelled as L105E5, L105E10, and L105E15. Their information is summarized in Table 3.

To test the results, we compare the CMFs of augmented merger trees with our L35N2650 high-resolution simulation in Fig. 4. The upper and lower panels correspond to different descendant halo mass bins. The CMFs of extended trees are shown as dashed lines, which broadly agree with L35N2650. Several discrepancies, e.g. the underestimation at the low-mass end at $z = 5.5$, can be explained by the fact that the CMFs given by the Parkinson et al. (2008) algorithm do not fully agree with the simulation as demonstrated in Fig. 1. However, we find that this overestimation does not affect the stellar mass functions when applying a semi-analytic model to the augmented trees. We show this in Section 3.

The HMFs of the extended trees are demonstrated in the upper panels of Fig. 5. They show excellent agreement with L35N2650. The lower panels of the figure explicitly show the HMFs of *N*-body, MC-I, and MC-II haloes from L105E10. As defined in Section 2.2.4, MC-I haloes augment *N*-body merger trees, while MC-II haloes are added to form a complete sample of haloes, and are independent of *N*-body haloes. While MC-II haloes dominate the population at lower redshifts, MC-I haloes are the main contributor at higher redshifts. Hence, both types of haloes are necessary to calculate the halo abundance across all redshifts.

2.3 Halo positions

When modelling reionization, we require spatial information for haloes within the extended halo catalogues. We aim to assign a position to every Monte Carlo halo and ensure that their two-point statistics agree with *N*-body simulations. Section 2.3.1 discusses a random sampling method for placing MC-II haloes within the simulation in the snapshot that the augmentation of the *N*-body merger trees is started, i.e. at $z = 5$. The method is then also used to verify our approach for evolving the position of Monte Carlo haloes based on the position of their descendant, which is introduced in Section 2.3.2.

2.3.1 *Populating halo positions*

Monte Carlo haloes can be populated into a simulation box using an analytic halo bias to transform the dark matter density field to a halo density field as a function of halo mass (de la Torre & Peacock 2013; Angulo et al. 2014; Neyrinck et al. 2014; Ahn et al. 2015;

Table 3. Information on halo catalogues used in this work.

Name	Type	Box size (h^{-1} Mpc)	Particle mass ($h^{-1}\,{\rm M}_\odot$)	Mass resolution ($h^{-1}\,{\rm M}_\odot$)
L35N2650	*N*-body simulation	35	2.00×10^5	–
L105N2048	*N*-body simulation	105	1.17×10^7	–
L105E5	Hybrid	105	–	1.4×10^8
L105E10	Hybrid	105	–	5.7×10^7
L105E15	Hybrid	105	–	3.2×10^7

Note. The mass resolutions of L105E5, L105E10, and L105E15 correspond to the atomic cooling threshold at $z = 5$, 10, and 15, respectively.

Downloaded from https://academic.oup.com/mnras/article/500/1/493/5935244 by guest on 20 April 2024

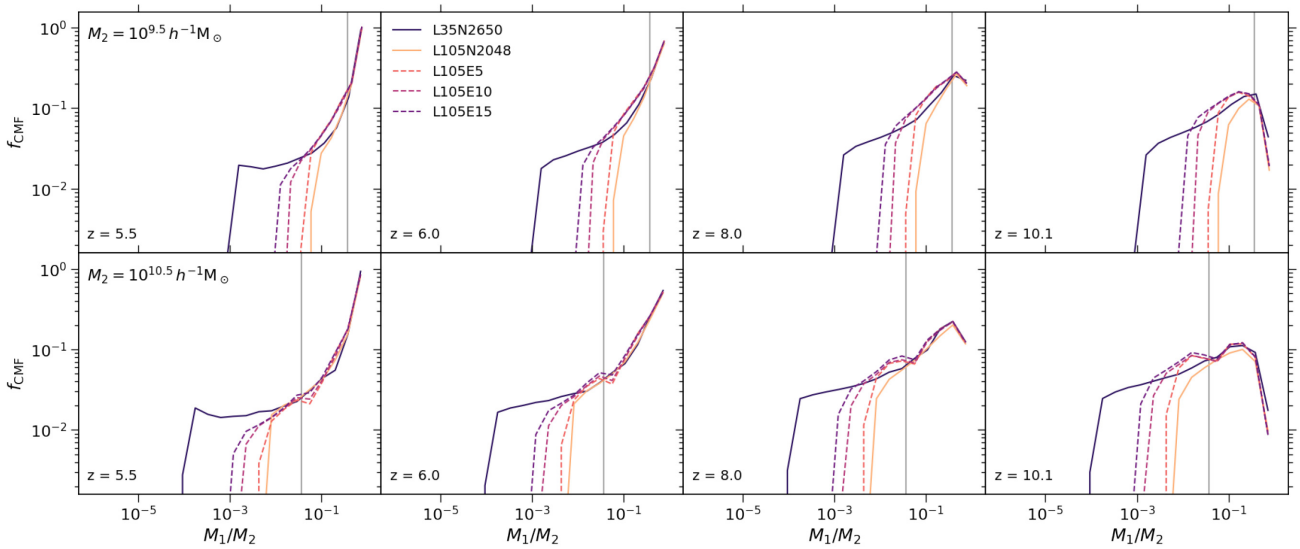

Figure 4. Comparisons of the conditional functions, defined by $\mathrm{d}f_{\rm CMF}/\mathrm{dln}\,M_1$, of *N*-body and augmented merger trees. Solid lines are the results derived using L35N2650 and L105N2048. The information on these two *N*-body simulations can be found in Table 3. Dashed lines are based on augmented halo merger trees, which are obtained by applying the algorithm described in Section 2.2.2 to L105N2048. Darker colours correspond to higher mass resolution. The grey vertical lines show the minimum mass cut of the augmentation algorithm.

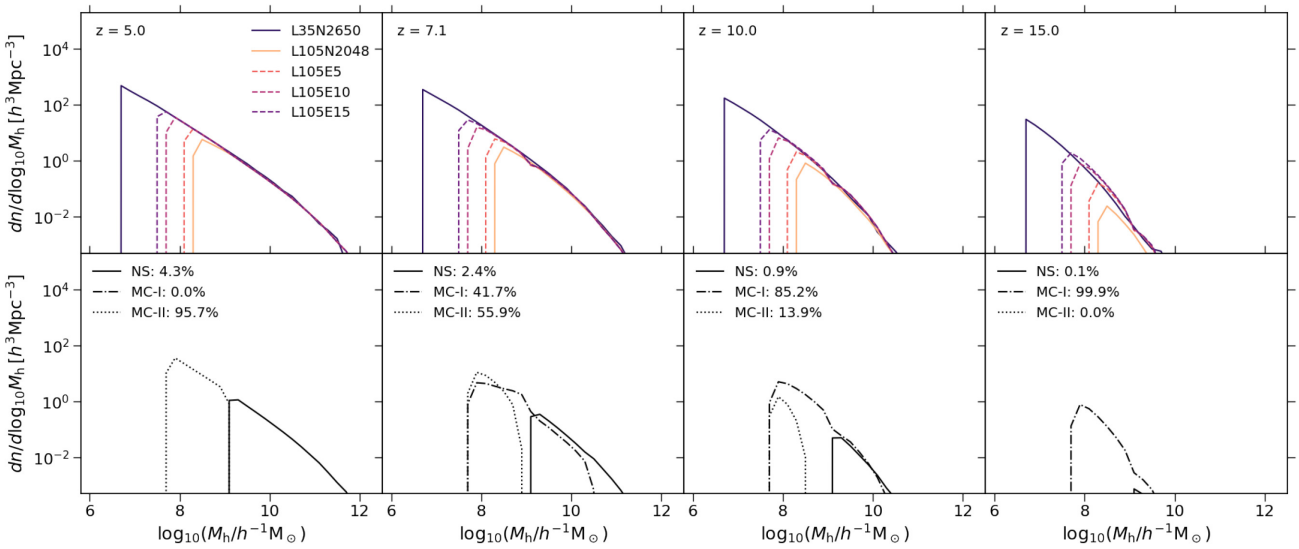

Figure 5. Upper panels: comparisons of the HMFs of *N*-body and extended halo catalogues. Solid lines are estimated from the *N*-body simulations, using L35N2650 and L105N2048. Their information can be found in Table 3. Dashed lines are based on extended halo catalogues, which are obtained by applying the algorithm described in Section 2.2 to L105N2048. The mass resolutions of L105E5, L105E10, and L105E15 correspond to the atomic cooling thresholds at $z = 5$, 10, and 15 respectively. Darker colours correspond to higher mass resolution. Bottom panels: HMFs of *N*-body, MC-I, and MC-II haloes from L105E10. Their mass fractions are labelled in the top left corners. See Section 2.2.4 for the definition of MC-I and MC-II haloes.

Nasirudin, Iliev & Ahn 2020). In this work, the dark matter density field is estimated from the L105N2048 *N*-body simulation using the nearest grid point method. The result is represented as a cubic grid. To estimate the halo density field, we adopt the non-linear halo bias proposed by Ahn et al. (2015), which avoids negative density in underdense regions, and results in better two-point correlation functions on smaller scales. Halo positions are obtained by random sampling. We normalize the halo density field derived from the halo bias, and treat it as a one-dimensional discrete probability distribution. Then, at a given snapshot, we assign every Monte Carlo halo to a cell according to this probability and place it uniformly within the cell so that the number of haloes in each cell follows the Poisson distribution. This approach does not depend on the normalization of the halo density field and can be applied to any given number of Monte Carlo haloes.

To verify this method, we carry out a test within mass ranges that are well resolved by L105N2048. Specifically, we apply this method to 10^5 samples, placing them within an empty box and measuring their two-point correlation functions. Then, we compare the results using *N*-body haloes from L105N2048. We perform the test at $M_{\rm halo} = 10^{9.1}$ and $10^{9.5}\,h^{-1}\,\mathrm{M}_\odot$ from $z = 5$ to 10 with different grid sizes. The results can be found in Fig. 6, which shows good agreement with those estimated from the *N*-body simulation.

Downloaded from https://academic.oup.com/mnras/article/500/1/493/5935244 by guest on 20 April 2024

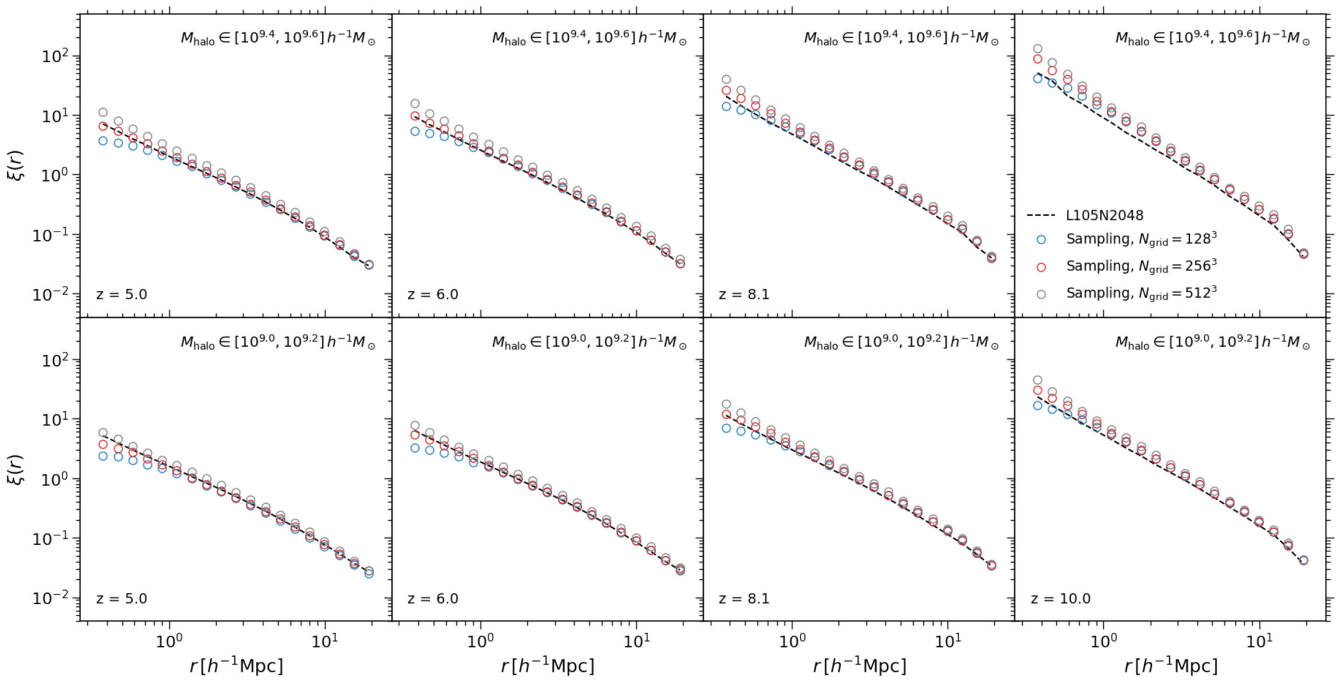

Figure 6. Comparison of two-point correlation functions produced using the random sampling method and estimated from *N*-body simulations. Empty circles are the results based on the random sampling method introduced in Section 2.3.1, with colours corresponding to different grid sizes as labelled on the top rightmost panel. Black dashed lines are estimated from the L105N2048 *N*-body simulations. Each row corresponds to a halo mass bin. These mass ranges are well resolved by L105N2048.

The small-scale clustering predicted by the random sampling method is affected by the choice of grid sizes. Halo positions within a cell of the grid are inaccurate since they are assumed to be uniformly distributed. As expected, the two-point correlations obtained using a 128^3 grid (shown as blue circles in Fig. 6) are underestimated at separations smaller than $0.8\,h^{-1}$ Mpc, which is equal to the cell size of the grid. In terms of the results using a 512^3 grid (grey circles), they have slightly larger clustering amplitudes over all scales than those using a 256^3 grid (red circles). A potential reason could be that the estimation of the dark matter density field becomes noisy when a larger number of cells are used. For the following applications, we adopt a 256^3 grid for the random sampling method. This choice is appropriate since the corresponding cell size ($0.4\,h^{-1}$ Mpc) is smaller than the characteristic size of ionizing regions (e.g. Furlanetto, McQuinn & Hernquist 2006).

Unfortunately, we are unable to do the same test for Monte Carlo haloes in the extended halo catalogues. This is because a complete sample of *N*-body haloes at these mass ranges is only available in L35N2650, for which the box size is not sufficient to estimate two-point statistics. However, we note that the linearity of halo density fields increases towards lower halo mass, implying that the results are likely to be improved at $M_{\rm halo} \lesssim 10^8\,h^{-1}\,{\rm M}_\odot$. This argument indicates that the results in Fig. 6 are conservative for estimating the accuracy of the method. Hence, our method can be safely applied to the mass ranges that we are interested in.

2.3.2 *Evolving halo positions*

Evolution in the clustering of haloes is influenced by their peculiar motions. Our approach of evolving halo positions is based on the linear continuity equation. We again divide the L105N2048 box into a cubic grid with 256^3 cells. For Monte Carlo trees at t_1, the first step is to place the haloes into the same cell as their direct descendant at t_2. We assume that the spatial distribution of the haloes at t_1 can be described by a halo density field denoted as $\mathcal{D}(\vec{x}, t_1)$. The idea is to move these haloes using a velocity field such that their spatial distribution becomes a desired halo density field denoted as $\mathcal{D}(\vec{x}, t_2)$. We assume that this process can be described by the linear continuity equation. If $\Delta t = t_2 - t_1$ is small, the velocity field can be obtained by

$$\nabla\vec{v}(\vec{x}, t_2) = -\frac{1}{\Delta t}\left[\mathcal{D}(\vec{x}, t_1) - \mathcal{D}(\vec{x}, t_2)\right]. \quad (3)$$

In the linear regime, we want

$$\mathcal{D}(\vec{x}, t_1) = b(M_1, t_1)\delta_{\rm DM}(\vec{x}, t_1), \quad (4)$$

where M_1 is the mass of the Monte Carlo haloes and $b(M, t)$ is the linear halo bias. After a forward evolution, the change of the density field for haloes at t_1 with mass M_1 is contributed from both the variation of the background dark matter density field and local interactions such as smooth mass accretion and mergers. Although a detailed model that considers all the effects is complicated, we find that evolving halo positions using the following expression for $\mathcal{D}(\vec{x}, t_2)$ can lead to reasonable two-point statistics.

$$\mathcal{D}(\vec{x}, t_2) = b(M_1/\bar{\mu}_{\rm R}, t_2)\delta_{\rm DM}(\vec{x}, t_2), \quad (5)$$

where $\bar{\mu}_{\rm R}$ is the mean mass ratio between the progenitor and descendant haloes.

Then, it is straightforward to compute the velocity field using the Fourier transform. The velocity field in *k*-space can be written as

$$\vec{v}(\vec{k}, t_2) = b(M_1/\bar{\mu}_{\rm R}, t_2)\vec{u}(\vec{k}, t_2) - b(M_1, t_1)\vec{u}(\vec{k}, t_1) \quad (6)$$

with

$$\vec{u}(\vec{k}, t) = \frac{i\vec{k}}{\Delta t k^2}\delta_{\rm DM}(\vec{k}, t), \quad (7)$$

Downloaded from https://academic.oup.com/mnras/article/500/1/493/5935244 by guest on 20 April 2024

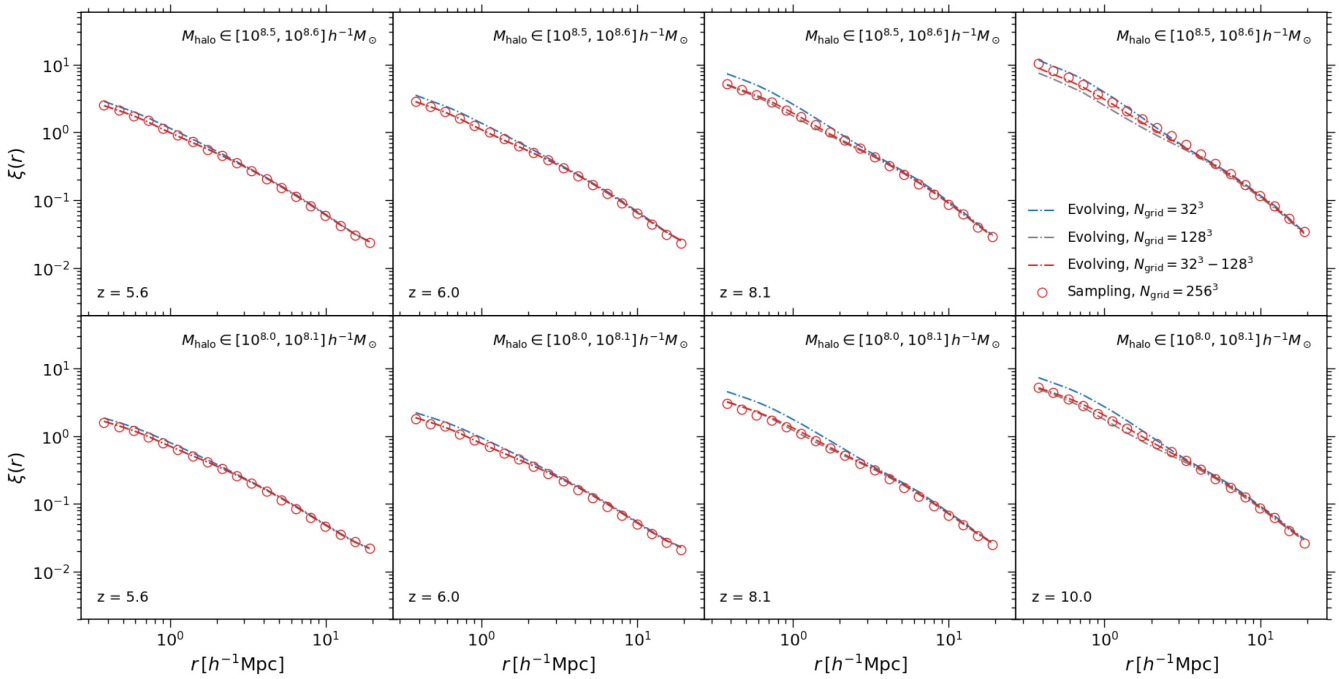

Figure 7. Comparison of two-point correlation functions produced using the evolving method and estimated from *N*-body simulations. Dash–dotted lines are the results based on the evolving method introduced in Section 2.3.2. For blue and grey lines, grids with 32^3 and 128^3 cells are used to calculate the velocity field, respectively, while for red lines, the adopted grid size varies with redshift, with 128^3 cells at $z = 5$–6, 64^3 cells at $z = 6$–8, and 32^3 cells at $z > 8$. Red empty circles are the results obtained using the sampling method described in Section 2.3.1, which can be used to check the accuracy of the evolving method.

The real space velocity field then can be obtained using the inverse Fourier transform. Since $\vec{u}(\vec{k}, t)$ is independent of halo mass, we only need to perform the Fourier transform once per snapshot, and the velocity can be calculated per halo, without any mass bins. This advantage is only available when the halo bias and the dark matter density field are separable. For the linear halo bias, we adopt the fitting model given by Tinker et al. (2010).

We apply this method to all extended halo catalogues and find that the choice of grid sizes to calculate $\vec{u}(\vec{k}, t)$ can affect the results. In Fig. 8, we show that the median velocity of Monte Carlo haloes is underestimated at $z \sim 5$ using a 32^3 grid and is overestimated at $z \sim 10$ using a 128^3 grid. This trend is expected. The density field should not be over smoothed, as this loses the information on density peaks. On the other hand, the halo bias increases rapidly with redshift, in which case the halo density field cannot be described by the linear bias. Smoothing the density field over larger regions can increase the linearity.

To verify the two-point correlation functions predicted by the evolving method, we have to use the sampling method introduced in the previous section. A direct comparison with L35N2650 is not feasible due to its limited box size, and the accuracy of this indirect approach is confirmed in the previous section. Fig. 7 compares the two-point correlation functions obtained using the sampling and evolving methods. Since halo positions are evolved backwards with time, when a 32^3 grid is used, the errors due to the underestimation of the halo velocity accumulate towards higher redshifts, which results in the overestimation of the two-point correlation functions at $z \gtrsim 6$ (see blue dash–dotted lines). Overall, we find good agreement between the results based on both methods, particularly on large scales.

Based on the discussion above, we have decided to vary the grid size with redshift when evolving halo positions. Specifically, we use a 128^3 grid at $z = 5$–6, a 64^3 grid at $z = 6$–8, and a 32^3 grid at $z > 8$. This treatment results in both consistent two-point correlation functions and velocity distributions, which are shown as red dash–dotted lines and red histograms in Figs 7 and 8, respectively.

2.4 Spin parameters

Many semi-analytic models use the halo spin parameter (defined by Bullock et al. 2001) to compute quantities including disc size and star formation rate. To facilitate this, we sample the spin parameter of Monte Carlo haloes using the spin distributions estimated from the *N*-body simulation. At $z \geq 5$, negligible dependence on halo mass is found in the spin distributions of our simulations, which is consistent with Knebe & Power (2008) and Angel et al. (2016). The mass independent spin distributions can be described by a lognormal distribution (e.g. van den Bosch 1998; Knebe & Power 2008) or a modified profile taking into account the long tail of low spins (e.g. Bett et al. 2007; Angel et al. 2016). In this work, we adopt a non-parametric approach. We train a Gaussian kernel density estimator (see e.g. Scott 2015) using samples from our *N*-body simulations (in $\log_{10}\lambda$ space), and assign the spin of Monte Carlo haloes by resampling from the density estimator. We choose the bandwidth of the density estimator according to Scott's Rule (Scott 2015).

In Fig. 9, black and red histograms are the spin distributions based on *N*-body and Monte Carlo haloes, respectively. When assembling *N*-body haloes to estimate the spin distributions, we only include haloes comprised of at least 100 particles and exclude all subhaloes. Our results illustrate excellent agreement between the resampled and original distributions by construction. We note that our approach can be generalized to the case where spin parameter is tightly correlated with halo mass by splitting the total sample into several mass bins and applying the kernel density estimator to each subsample.

Downloaded from https://academic.oup.com/mnras/article/500/1/493/5935244 by guest on 20 April 2024

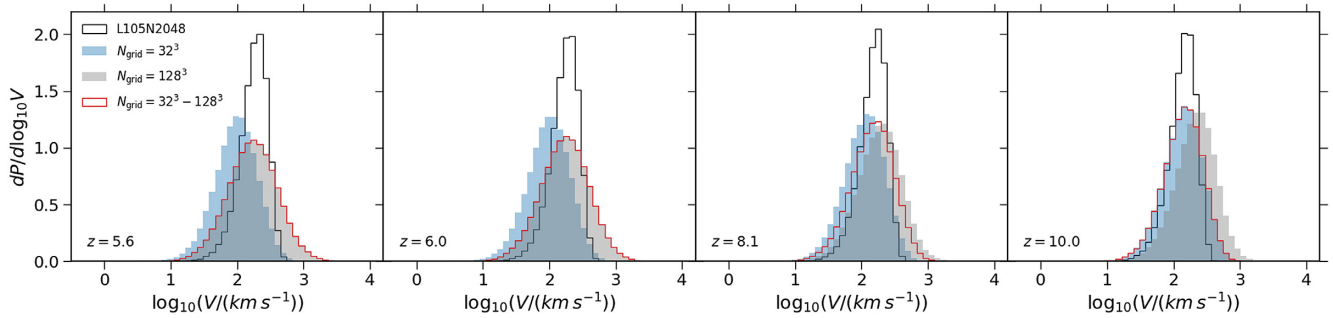

Figure 8. Peculiar velocity distributions of *N*-body and Monte Carlo haloes. The velocities of Monte Carlo haloes are derived using the method introduced in Section 2.3.2. For blue and grey histograms, grids with 32^3 and 128^3 cells are used in the calculations, while for red histograms, the adopted grid size varies with redshift, with 128^3 cells at $z = 5$–6, 64^3 cells at $z = 6$–8, and 32^3 cells at $z > 8$. The distributions of *N*-body haloes are shown as black histograms.

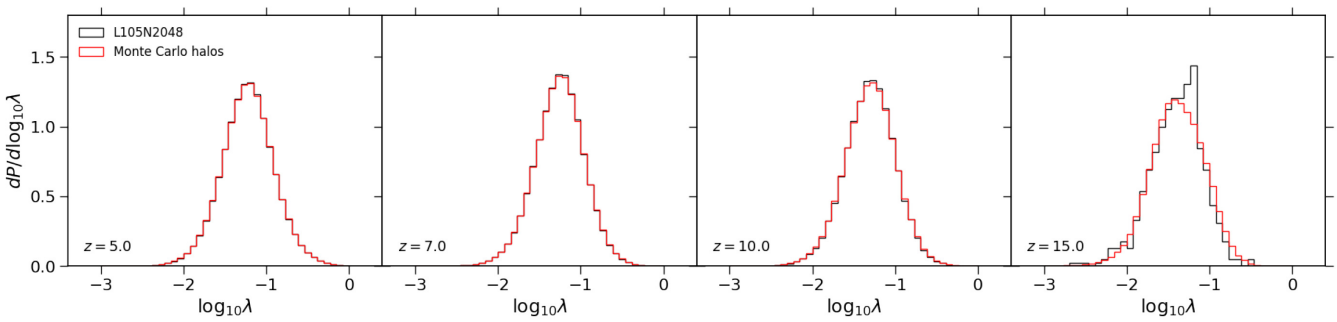

Figure 9. Spin distributions of *N*-body and Monte Carlo haloes, plotted as back and red histograms, respectively. The spin parameters of Monte Carlo haloes are resampled from the distributions of *N*-body haloes using a Gaussian kernel density estimator (see e.g. Scott 2015). We only include *N*-body haloes comprised of at least 100 particles to estimate their spin distributions, with subhaloes excluded.

3 APPLICATION TO MERAXES

We apply both the *N*-body and extended halo catalogues to the MERAXES semi-analytic model (Mutch et al. 2016). In addition to the implementation of several key galaxy formation processes including radiative cooling, star formation, and supernova feedback, the MERAXES model is coupled with 21CMFAST (Mesinger & Furlanetto 2007) to realize inhomogeneous reionization feedback and to predict reionization related properties such as the global neutral fraction and 21 cm power spectra. The MERAXES model only seeds galaxies in haloes whose mass is above the atomic cooling threshold. We adopt the same parameters as Mutch et al. (2016) but note that the model predictions can be different from Mutch et al. (2016) due to the use of different halo merger trees. However, the main focus of this work is to demonstrate the consistency between the *N*-body and extended halo catalogues and to illustrate the consequences of adopting different halo mass resolutions rather than to present a model that satisfies all current observational constraints.

3.1 Galaxy properties

Figs 10, 11, and 12 demonstrate the stellar mass functions, star formation rate functions, and star formation rate densities predicted by MERAXES, respectively. L35N2650 is a small-volume *N*-body simulation with very high mass resolution, which is used to verify the results based on the extended halo catalogues. The predicted galaxy properties using L35N2650 and extended trees are shown as purple solid and dashed lines, respectively. We find a difference in the peaks of both the stellar mass and star formation rate functions, which may result from the fact that Monte Carlo merger trees do not contain subhaloes. This point is illustrated in the lower panels of Figs 10 and 11, where we show that L35N2650 provides significantly higher satellite fractions than the extended halo catalogues, particularly at the low-stellar mass and low-star formation rate ends. In MERAXES, all gas infalling into a friends-of-friends group is assumed to be accreted on to the central galaxy. Therefore, satellite galaxies have less fuel to form stars. Despite this disagreement, we find excellent agreement between the cosmic star formation rate densities obtained using L35N2650, L105E10, and L105E15 at $z < 10$. The result based on L105E15 shows higher star formation rate density than L35N2650 at $z > 10$. However, L35N2650 has a higher mass resolution. This is likely due to the overestimation of the HMFs at these redshifts as illustrated in Fig. 5.

An additional finding is that the effect of mass resolution does not seem to be cumulative. While the mass resolutions of L105E5, L105E10, and L105E15 are different (and all above the atomic cooling threshold at $z = 5$), in Fig. 10, their corresponding stellar mass functions overlap at $z = 5$. Fig. 12 also shows that the star formation rate densities predicted by the extended trees converge towards $z = 5$. These findings are non-trivial. We note that even if a halo is below the atomic cooling threshold at a given redshift, it can still host a galaxy. The reason is that the atomic cooling threshold increases with redshift, and as long as any progenitor of a halo is above the cooling limit, the halo will contain a galaxy. Therefore, we should not expect that halo catalogues with different mass resolutions produce similar stellar mass and star formation rate functions towards $z = 5$. On the contrary, our results indicate that if all haloes above the atomic cooling threshold at a given redshift are resolved, an ability to resolve less massive haloes at an earlier time has little effect on predicted galaxy properties such as the stellar mass and star formation rate functions at the given redshift.

Downloaded from https://academic.oup.com/mnras/article/500/1/493/5935244 by guest on 20 April 2024

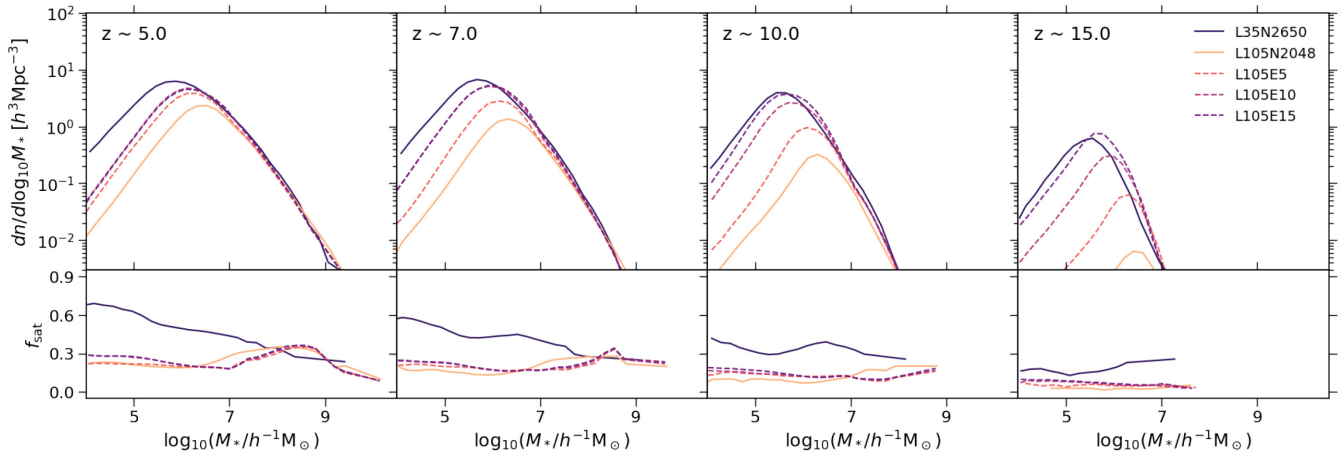

Figure 10. Upper panels: stellar mass functions predicted by the MERAXES semi-analytic model. Lower panels: satellite fractions as a function of stellar mass. For all panels, solid lines use the original halo merger trees from our *N*-body simulations. Dashed lines are the results based on extended catalogues, which consist of both *N*-body and Monte Carlo haloes. Darker colours correspond to higher mass resolution. The information on each halo catalogue as labelled in the top right corner can be found in Table 3.

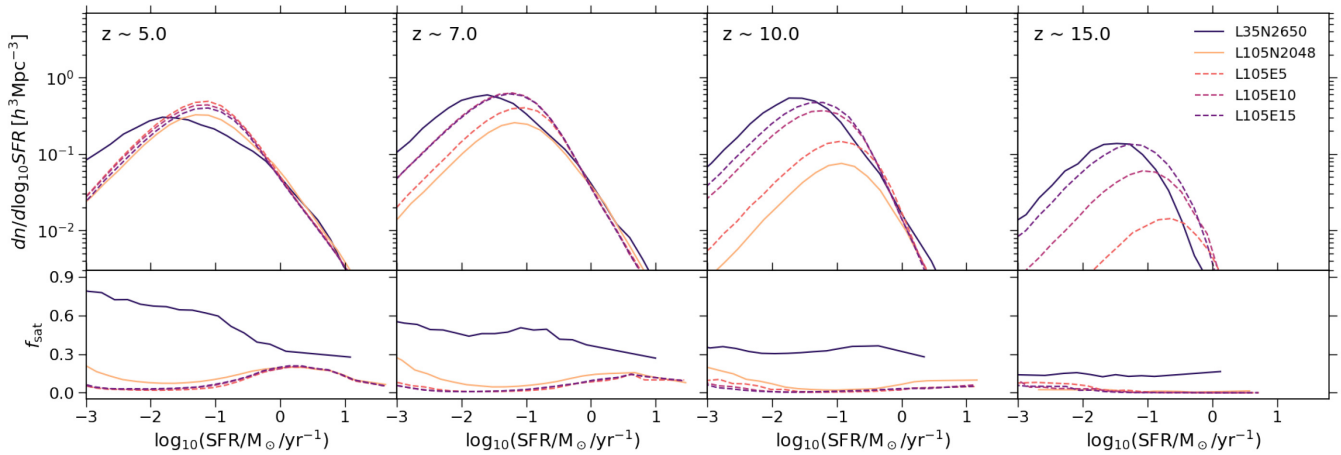

Figure 11. Upper panels: star formation rate functions predicted by the MERAXES semi-analytic model. Lower panels: satellite fractions as a function of star formation rate. For all panels, solid lines use the original halo merger trees from our *N*-body simulations. Dashed lines are the results based on extended catalogues, which consist of both *N*-body and Monte Carlo haloes. Darker colours correspond to higher mass resolution. The mass resolutions of L105E5, L105E10, and L105E15 are the atomic cooling thresholds at $z = 5$, 10, and 15, respectively. The information on each halo catalogue as labelled in the top right corner can be found in Table 3.

3.2 Reionization histories

Having demonstrated the galaxy properties based on both *N*-body and extended halo catalogues, we now focus on the predictions of cosmic reionization. The end of reionization is known to be too rapid in simulations that do not resolve all faint galaxies or do not have a sufficiently large volume (Barkana & Loeb 2004; Iliev et al. 2014). We therefore expect that the predictions of the reionization history are sensitive to both halo mass resolution and simulation volume.

Fig. 13 illustrates the effect of halo mass resolution on the predicted volume-weighted neutral fractions. We see a difference between results using direct *N*-body merger trees (from L35N2650 and L105N2048). However, it is not straightforward to interpret this due to the different simulation volumes. Our extended halo catalogues (L105E10 and L105E15) have the same volume as L105N2048 and produce consistent star formation rate densities with L35N2650. Fig. 13 shows that the end of reionization occurs earlier in L105E10 and L105E15 than in L105N2048, which confirms that mass resolution has an impact on the reionization history. This is expected since reionization is sensitive to cumulative star formation. A similar result was previously obtained by Finlator et al. (2018). We note that our results only indicate a minimum requirement of the mass resolution for predicting convergent reionization histories, since our model neglects star formation below the atomic cooling threshold, which may also provide a non-negligible contribution to reionization (e.g. Wise et al. 2014).

Small box simulations are known to suffer from both cosmic variance and lack of large-scale modes (e.g. Barkana & Loeb 2004). We demonstrate this effect using subvolumes of the L105E10 extended halo catalogue. In the left-hand and middle panels of Fig. 14, we show reionization histories in two different sizes of subvolumes, having $L_{\rm sub} = 35$ and $21\ h^{-1}$ Mpc. The former has the same volume as L35N2650, while the latter is roughly equal to the maximum bubble size that we choose in the 21CMFAST algorithm within MERAXES. Each subvolume contains different amounts of large-scale power, leading to a rapid end of reionization in each case,

Downloaded from https://academic.oup.com/mnras/article/500/1/493/5935244 by guest on 20 April 2024

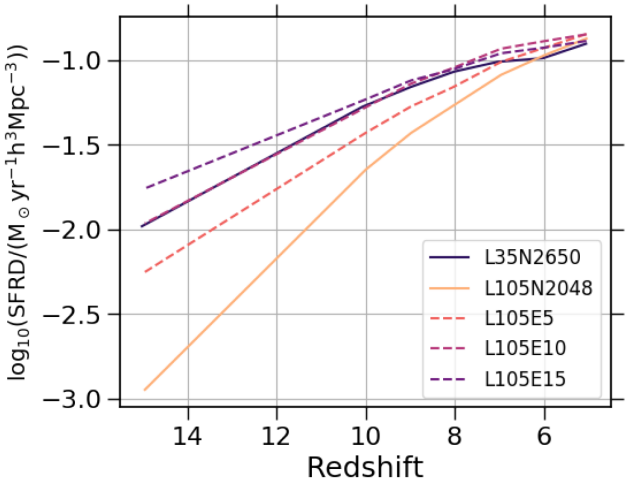

Figure 12. Star formation rate density predicted by the MERAXES semi-analytic model. Solid lines use the original halo merger trees from our *N*-body simulations. Dashed lines are the results based on extended catalogues, which consist of both *N*-body and Monte Carlo haloes. Darker colour corresponds to higher mass resolution. The mass resolutions of L105E5, L105E10, and L105E15 are the atomic cooling thresholds at $z = 5$, 10, and 15, respectively. The information on each halo catalogue as labelled in the bottom right corner can be found in Table 3.

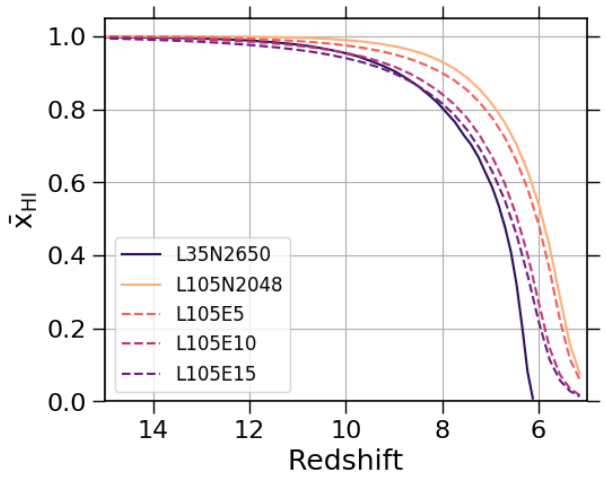

Figure 13. Volume-weighted neutral fractions predicted by the MERAXES model. Solid lines and dashed lines are the results based on *N*-body and extended halo catalogues, respectively. Darker colours correspond to higher mass resolution. The mass resolutions of L105E5, L105E10, and L105E15 are the atomic cooling thresholds at $z = 5$, 10, and 15, respectively. See Table 3 for the information on these halo catalogues.

but at a range of redshifts. This explains the deviation of the shape of the late-time reionization history in L35N2650 from the predictions based on L105E10 and L105E15. The large-volume simulations average cosmic variance shown within subvolumes in Fig. 14.

In the right-hand panel of Fig. 14, we compare the standard deviation of redshift at fixed neutral fractions in the subvolumes (solid lines) with the analytic prediction of Barkana & Loeb (2004) (dashed lines). They pointed out that the difference of the collapse fraction in random regions of the Universe can be interpreted as an offset in redshift with respect to the cosmic mean. The scatter of the offset can be calculated from the critical collapse fraction, and be related to the width or duration of the reionization history by equating it to the size of a particular reionization region (Wyithe & Loeb 2004). Despite the complexities in MERAXES, the analytic prediction provides a reasonable estimation of cosmic variance. Overall, our results reinforce the importance of a large volume for cosmic reionization simulations, which has also been highlighted by previous studies (e.g. Iliev et al. 2006, 2014; Deep Kaur et al. 2020).

In addition, our results show that resolving all haloes above the atomic cooling threshold across whole cosmic reionization is important for calculating a converged reionization history. Robertson et al. (2015) analysed the joint observational constraints of Thomson scattering optical depth measured by Planck Collaboration XIII (2016) and cosmic star formation rate density estimated by Madau & Dickinson (2014), suggesting that cosmic reionization happens at $6 \lesssim z \lesssim 10$. Our results imply that simulations should reach at least the atomic cooling threshold at $z = 10$ in order to explore such reionization scenarios. The decrease of the atomic cooling threshold with increasing redshift places constraints on the required halo mass resolution of simulations towards the beginning of reionization.

4 SUMMARY

In this paper, we present a hybrid method to compute high-resolution halo merger trees within large-volume *N*-body simulations for semi-analytic reionization models, which is based on the work of Benson et al. (2016). As an application, we extend the mass resolution of halo merger trees extracted from the Genesis *N*-body $105\,h^{-1}$Mpc simulation box at $z \geq 5$. We verify the results using a small *N*-body simulation with very high resolution, and find good agreement for the HMFs. We also introduce a method to assign and evolve the position of Monte Carlo haloes. The resulting two-point correlation functions are consistent with *N*-body simulations at separations greater than $0.4\,h^{-1}$ Mpc. In the application to the MERAXES semi-analytic model, the extended halo catalogues provide significant improvements on the predicted galaxy properties and reionization history.

(i) The decreasing atomic cooling threshold requires simulations to have higher mass resolution towards higher redshifts. Our model confirms that the faint sources at the beginning of reionization can have a significant impact on the reionization history, and therefore resolving the atomic cooling threshold throughout reionization is necessary for reliable calculations of the reionization history.

(ii) The end of reionization is predicted to be too rapid in simulations that either fail to resolve all faint galaxies or have a too small volume, putting demands on halo mass resolution and simulation volume. Using our extended tree algorithm, we show that the convergent predictions of the late-stage reionization history need both large volumes ($L_{\rm box} \gtrsim 100\,h^{-1}$ Mpc) and resolution of the atomic cooling threshold across the whole reionization history.

(iii) If all haloes above the atomic cooling threshold at a given redshift are resolved, resolving even smaller haloes at higher redshifts has negligible effect on predictions of galaxy population properties from the MERAXES semi-analytic model at the given redshift.

Our methodology provides a powerful tool to achieve desired mass resolution in large volumes. The largest extended halo catalogue obtained in this work has the mass resolution of $M_{\rm halo} = 3.2 \times 10^7\,h^{-1}\,{\rm M}_\odot$ in a $105\,h^{-1}$ Mpc box, equivalent to an *N*-body simulation with $\sim 6800^3$ particles. Given the efficiency of the Monte Carlo algorithms, our approach can be applied to larger volumes (several hundred Mpc on each side), which are necessary for studying the statistics of reionization including X-ray heating and global 21 cm signal during cosmic dawn.

Downloaded from https://academic.oup.com/mnras/article/500/1/493/5935244 by guest on 20 April 2024

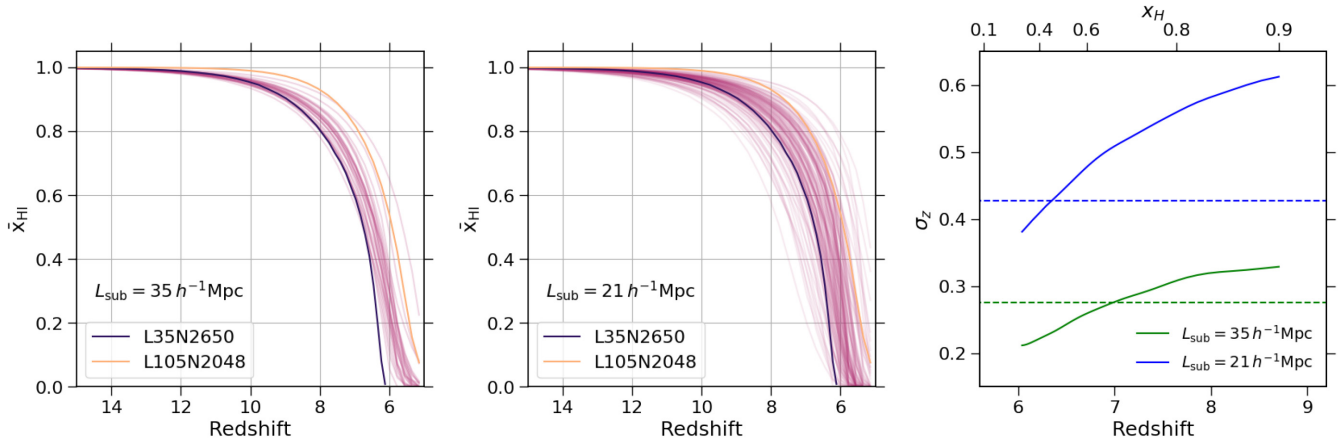

Figure 14. The left-hand and middle panels show the reionization histories in subvolumes with side lengths of 35 and 21 h^{-1} Mpc, respectively. The latter is roughly equal to the maximum bubble size that we choose for 21CMFAST. These results are based on L105E10. In the right-hand panel, solid lines show the standard deviations of redshift in subvolumes at fixed neutral fractions. Redshifts on the bottom axis are converted using the mean relation of the entire volume. The deviations are compared with the analytic predictions of Barkana & Loeb (2004), which are shown as dashed lines.

ACKNOWLEDGEMENTS

We thank the anonymous referee for providing a detailed report to improve the quality of the paper. This research was supported by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013. This work was performed on the OzSTAR national facility at Swinburne University of Technology. OzSTAR is funded by Swinburne University of Technology and the National Collaborative Research Infrastructure Strategy (NCRIS). YQ thanks Yuxiang Qin and Bradley Greig for useful discussions.

We acknowledge the use of the following software: ASTROPY[2] (Astropy Collaboration 2013, 2018), CORRFUNC (Sinha & Garrison 2019; Sinha & Garrison 2020), CYTHON (Behnel et al. 2011), HMF (Murray, Power & Robotham 2013), IPYTHON (Perez & Granger 2007), MATPLOTLIB (Hunter 2007), NUMPY (van der Walt, Colbert & Varoquaux 2011), PANDAS (McKinney 2010), SEABORN,[3] and SCIPY (Jones et al. 2001).

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Ahn K., Iliev I. T., Shapiro P. R., Srisawat C., 2015, MNRAS, 450, 1486
Angel P. W., Poole G. B., Ludlow A. D., Duffy A. R., Geil P. M., Mutch S. J., Mesinger A., Wyithe J. S. B., 2016, MNRAS, 459, 2106
Angulo R. E., Baugh C. M., Frenk C. S., Lacey C. G., 2014, MNRAS, 442, 3256
Astropy Collaboration, 2013, A&A, 558, A33
Astropy Collaboration, 2018, AJ, 156, 123
Barkana R., Loeb A., 2004, ApJ, 609, 474
Baugh C. M., 2006, Rep. Prog. Phys., 69, 3101
Behnel S., Bradshaw R., Citro C., Dalcin L., Seljebotn D. S., Smith K., 2011, Comput. Sci. Eng., 13, 31
Benson A. J., Cannella C., Cole S., 2016, Comput. Astrophys. Cosmol., 3, 3
Bett P., Eke V., Frenk C. S., Jenkins A., Helly J., Navarro J., 2007, MNRAS, 376, 215
Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440
Bower R. G., 1991, MNRAS, 248, 332
Bullock J. S., Dekel A., Kolatt T. S., Kravtsov A. V., Klypin A. A., Porciani C., Primack J. R., 2001, ApJ, 555, 240
Ceverino D., Glover S. C. O., Klessen R. S., 2017, MNRAS, 470, 2791
de la Torre S., Peacock J. A., 2013, MNRAS, 435, 743
Deep Kaur H., Gillet N., Mesinger A., 2020, MNRAS, 495, 2354
Elahi P. J., Poulton R., Canas R., 2019a, VELOCIraptor-STF: Six-Dimensional Friends-of-Friends Phase Space Halo Finder, Astrophysics Source Code Library, record ascl:1911.020
Elahi P. J., Poulton R., Tobar R., 2019b, TreeFrog: Construct Halo Merger Trees and Compare Halo Catalogs, Astrophysics Source Code Library, record ascl:1911.021
Elahi P. J., Cañas R., Poulton R. J. J., Tobar R. J., Willis J. S., Lagos C. d. P., Power C., Robotham A. S. G., 2019c, PASA, 36, e021
Elahi P. J., Poulton R. J. J., Tobar R. J., Cañas R., Lagos C. d. P., Power C., Robotham A. S. G., 2019d, PASA, 36, e028
Feng Y., Di-Matteo T., Croft R. A., Bird S., Battaglia N., Wilkins S., 2016, MNRAS, 455, 2778
Finkelstein S. L. et al., 2019, ApJ, 879, 36
Finlator K., Keating L., Oppenheimer B. D., Davé R., Zackrisson E., 2018, MNRAS, 480, 2628
Furlanetto S. R., McQuinn M., Hernquist L., 2006, MNRAS, 365, 115
Geil P. M., Mutch S. J., Poole G. B., Angel P. W., Duffy A. R., Mesinger A., Wyithe J. S. B., 2016, MNRAS, 462, 804
Gnedin N. Y., 2014, ApJ, 793, 29
Gnedin N. Y., Kaurov A. A., 2014, ApJ, 793, 30
Greig B., Mesinger A., 2015, MNRAS, 449, 4246
Hassan S., Davé R., Finlator K., Santos M. G., 2016, MNRAS, 457, 1550
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Iliev I. T., Mellema G., Pen U. L., Merz H., Shapiro P. R., Alvarez M. A., 2006, MNRAS, 369, 1625
Iliev I. T., Mellema G., Ahn K., Shapiro P. R., Mao Y., Pen U.-L., 2014, MNRAS, 439, 725
Johnson J. L., Dalla Vecchia C., Khochfar S., 2013, MNRAS, 428, 1857
Jones E. et al., 2001, SciPy: Open Source Scientific Tools for Python. Available at: http://www.scipy.org/
Katz H. et al., 2020, MNRAS, 494, 2200
Kaurov A. A., Gnedin N. Y., 2015, ApJ, 810, 154
Knebe A., Power C., 2008, ApJ, 678, 621
Lacey C., Cole S., 1993, MNRAS, 262, 627
Liu C., Mutch S. J., Angel P. W., Duffy A. R., Geil P. M., Poole G. B., Mesinger A., Wyithe J. S. B., 2016, MNRAS, 462, 235

[2]http://www.astropy.org
[3]https://github.com/mwaskom/seaborn

Downloaded from https://academic.oup.com/mnras/article/500/1/493/5935244 by guest on 20 April 2024

McKinney W., 2010, in van der Walt S., Millman J., eds, Proc. 9th Python in Sci. Conf. p. 51. Available at: http://conference.scipy.org/proceedings/scipy2010/mckinney.html
Madau P., Dickinson M., 2014, ARA&A, 52, 415
Mesinger A., Furlanetto S., 2007, ApJ, 669, 663
Murray S. G., Power C., Robotham A. S. G., 2013, Astron. Comput., 3, 23
Mutch S. J., Geil P. M., Poole G. B., Angel P. W., Duffy A. R., Mesinger A., Wyithe J. S. B., 2016, MNRAS, 462, 250
Naidu R. P., Tacchella S., Mason C. A., Bose S., Oesch P. A., Conroy C., 2020, ApJ, 892, 109
Nasirudin A., Iliev I. T., Ahn K., 2020, MNRAS, 494, 3294
Neyrinck M. C., Aragón-Calvo M. A., Jeong D., Wang X., 2014, MNRAS, 441, 646
Park J., Mesinger A., Greig B., Gillet N., 2019, MNRAS, 484, 933
Parkinson H., Cole S., Helly J., 2008, MNRAS, 383, 557
Perez F., Granger B. E., 2007, Comput. Sci. Eng., 9, 21
Pillepich A. et al., 2018, MNRAS, 475, 648
Planck Collaboration XIII, 2016, A&A, 594, A13
Robertson B. E., Ellis R. S., Furlanetto S. R., Dunlop J. S., 2015, ApJ, 802, L19
Rosdahl J. et al., 2018, MNRAS, 479, 994
Scott D. W., 2015, Multivariate Density Estimation: Theory, Practice, and Visualization. John Wiley Sons, Hoboken, NJ
Shi Y., Eberhart R., 1998, 1998 IEEE Int. Conf. Evol. Comput. Proc., IEEE World Congress on Computational Intelligence (Cat. No. 98TH8360). IEEE, Piscataway, NJ, p. 69
Sinha M., Garrison L., 2019, in Majumdar A., Arora R., eds, Software Challenges to Exascale Computing. Springer, Singapore, p. 3
Sinha M., Garrison L. H., 2020, MNRAS, 491, 3022
Somerville R. S., Davé R., 2015, ARA&A, 53, 51
Springel V., 2005, MNRAS, 364, 1105
Springel V. et al., 2005, Nature, 435, 629
Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, ApJ, 724, 878
van den Bosch F. C., 1998, ApJ, 507, 601
van der Walt S., Colbert S. C., Varoquaux G., 2011, Comput. Sci. Eng., 13, 22
Wise J. H., Turk M. J., Norman M. L., Abel T., 2012, ApJ, 745, 50
Wise J. H., Demchenko V. G., Halicek M. T., Norman M. L., Turk M. J., Abel T., Smith B. D., 2014, MNRAS, 442, 2560
Wyithe J. S. B., Loeb A., 2004, Nature, 432, 194

APPENDIX: ADDITIONAL CALIBRATIONS OF THE PARKINSON ALGORITHM

In Section 2.2.1, we do not employ any weights for different mass and redshift ranges in the cost function for calibrating the Parkinson et al. (2008) algorithm. In this appendix, we present two additional calibrations of the algorithm to show the potential bias of this treatment. In Fig. A1, the calibration results that use $z = 5.5$ data only and $z = 10.1$ data only are shown as yellow and purple empty circles, respectively. The corresponding parameters are listed in Table A1. The result that uses $z = 5.5$ data only is improved at $z = 5.5$ but becomes significantly poorer at higher redshifts. In terms of the purple empty circles, the fitting is improved at $M_2 = 10^{10.5}\,h^{-1}\,\mathrm{M}_\odot$, $z = 10.1$ and is similar or slightly poorer at other mass and redshift ranges. These results suggest that the employment of weighting may only provide moderate improvements on the calibration of the Parkinson et al. (2008) algorithm, which, however, is purely artificial. Therefore, we do not employ any weights on the calibration and adopt the parameters obtained in Section 2.2.1 as the fiducial model in this work.

Table A1. Results of two additional calibrations for the Parkinson et al. (2008) algorithm.

Symbol	All	$z = 5.5$ only	$z = 10.1$ only
G_0	1.0	0.7	0.6
γ_1	0.2	0.2	0.5
γ_2	− 0.4	0.4	− 0.1

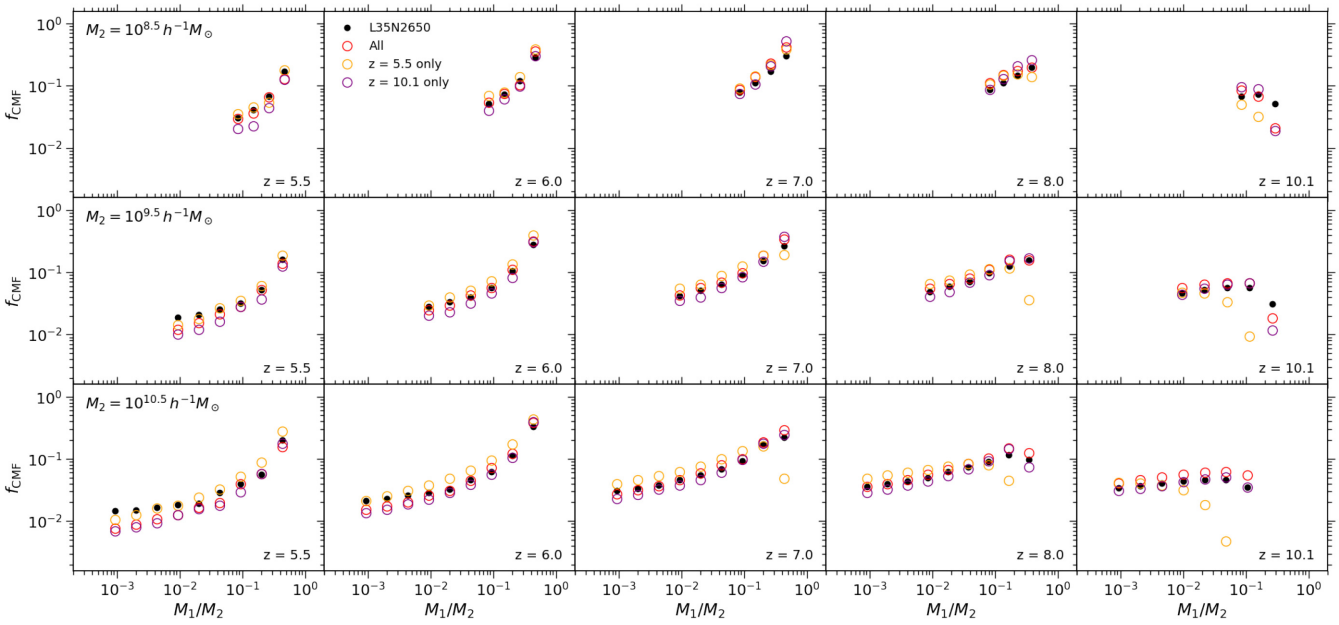

Figure A1. Fitting results of two additional calibrations for the Parkinson et al. (2008) algorithm. The CMFs are defined by $\mathrm{d}f_{\rm CMF}/\mathrm{dln}\,M_1$. Black dots are the fitting data, which are estimated using L35N2650. Red empty circles are the same as those in Fig. 1. Yellow and purple empty circles are the results that use $z = 5.5$ data only and $z = 10.1$ data only, respectively. Their corresponding parameters are listed in Table A1.

This paper has been typeset from a TEX/LATEX file prepared by the author.

Downloaded from https://academic.oup.com/mnras/article/500/1/493/5935244 by guest on 20 April 2024