Benson et al. *Computational Astrophysics and Cosmology* (2016) 3:3
DOI 10.1186/s40668-016-0016-3

Computational Astrophysics and Cosmology
a SpringerOpen Journal

RESEARCH Open Access

Achieving convergence in galaxy formation models by augmenting N-body merger trees

Andrew J Benson[1*], Chris Cannella[2] and Shaun Cole[3]

Abstract

Accurate modeling of galaxy formation in a hierarchical, cold dark matter universe requires the use of sufficiently high-resolution merger trees to obtain convergence in the predicted properties of galaxies. When semi-analytic galaxy formation models are applied to cosmological *N*-body simulation merger trees, it is often the case that those trees have insufficient resolution to give converged galaxy properties. We demonstrate a method to augment the resolution of *N*-body merger trees by grafting in branches of Monte Carlo merger trees with higher resolution, but which are consistent with the pre-existing branches in the *N*-body tree. We show that this approach leads to converged galaxy properties.

1 Introduction

A commonly used method for generating catalogs of galaxies from theoretical models is to apply semi-analytic galaxy formation models (Baugh 2006) to merger trees of dark matter halos extracted from N-body simulations (Kauffmann et al. 1999; Helly et al. 2003). This approach has been used to both study the physics of galaxy formation (Henriques et al. 2009; Bower et al. 2010; Lu et al. 2011; Mutch et al. 2013; Benson 2014; Ruiz et al. 2015) and to generate mock catalogs for use in determining survey sizes and accuracy (Lemson and The Virgo Consortium 2006; Bernyk et al. 2016) and comparing models and observations on an equal footing (e.g. Farrow et al. 2015).

A limiting factor in this approach is the resolution of the N-body simulation. A dark matter halo must be resolved with of order 100 particles to have robustly determined properties (Velliscig et al. 2015). In a modern cosmological simulation (e.g. the MultiDark Planck simulation; Klypin et al. 2016) this implies a minimum resolved halo mass of $\sim 2 \times 10^{11} \mathrm{M}_\odot$. While this may be sufficiently low to host galaxies of the masses of interest for a given application this does not mean that the physical properties of such galaxies are converged. Galaxy formation in the cold dark matter (CDM) paradigm is inherently hierarchical, meaning that the properties of a galaxy in a given halo may depend on those of galaxies which form in lower mass progenitor halos. If those halos are not resolved, the properties of the primary galaxy may not be converged even though its halo is resolved. This can happen not only because galaxies in unresolved progenitor halos are not available to merge with the primary galaxy, but also because those galaxies will lock up baryonic material into stars (making it unavailable to the primary galaxy), and produce metals which may contaminate the circumgalactic medium (CGM) of the primary galaxy (affecting cooling rates etc.).

For example, Bower et al. (2006) had to introduce a sub-resolution metal enrichment prescription into their semi-analytic model (to account for metals which would have been produced by unresolved galaxies in a higher resolution simulation) when applying it to merger trees from the Millennium simulation (Springel et al. 2005) in order to achieve the same results as when run on higher resolution merger trees to which they had calibrated their model.

Rather than introducing sub-resolution models to counteract the lack of resolution, we propose a method which

*Correspondence: abenson@carnegiescience.edu
[1]Carnegie Observatories, 813 Santa Barbara Street, Pasadena, CA 91101, USA
Full list of author information is available at the end of the article

augments N-body merger trees with high-resolution branches generated using a Monte Carlo approach. The branches generated are consistent with the resolved structure of the N-body tree, but can include much higher resolution structure. While these branches lack positional information about their halos, this is generally not necessary to perform the galaxy formation calculations.

The remainder of this paper is organized as follows. In Section 2 we describe our method for augmenting the resolution of N-body merger trees, and in Section 3 we demonstrate how this approach leads to convergence in galaxy properties (specifically stellar masses are explored in this work, although our approach works for all galaxy properties). Finally, in Section 4 we give our conclusions.

2 Method

Given a merger tree extracted from an N-body simulation, our goal is to augment the resolution of the tree by grafting on higher resolution branches which match the pre-existing tree structure to some given precision. To generate those branches we will use the Parkinson et al. (2008) algorithm, which has been tuned to reproduce the statistical properties of N-body merger trees. Hereafter we refer to trees generated using this algorithm as PCH trees.

We consider the N-body merger tree as a graph consisting of a collection of connected nodes (halos; see Figure 1(a). In this work we will exclusively use halo trees, as opposed to subhalo trees. Our method is applicable to subhalo trees also, but we choose not to explore such

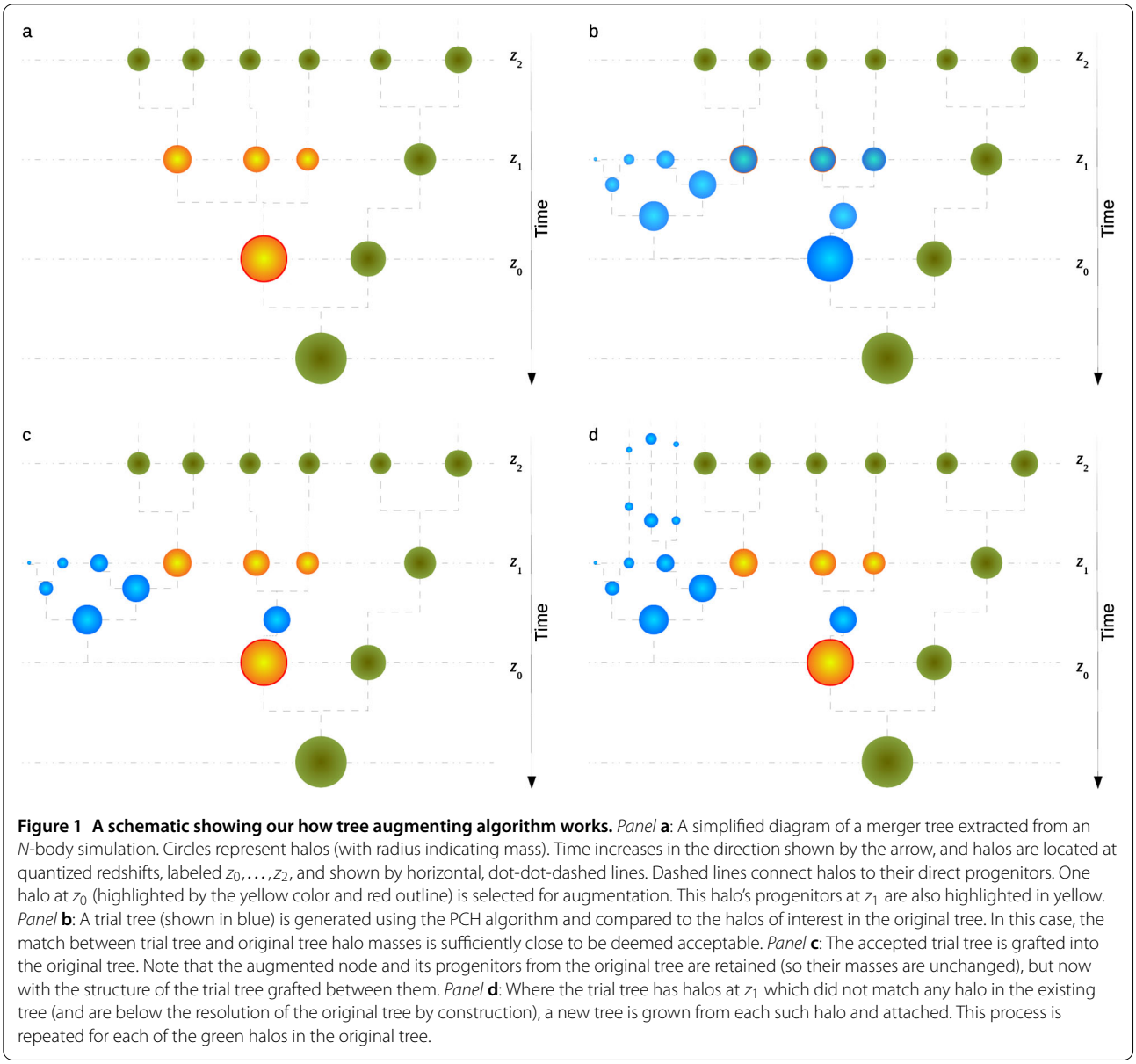

Figure 1 A schematic showing our how tree augmenting algorithm works. *Panel* **a**: A simplified diagram of a merger tree extracted from an N-body simulation. Circles represent halos (with radius indicating mass). Time increases in the direction shown by the arrow, and halos are located at quantized redshifts, labeled $z_0, \ldots, z_2$, and shown by horizontal, dot-dot-dashed lines. Dashed lines connect halos to their direct progenitors. One halo at z_0 (highlighted by the yellow color and red outline) is selected for augmentation. This halo's progenitors at z_1 are also highlighted in yellow. *Panel* **b**: A trial tree (shown in blue) is generated using the PCH algorithm and compared to the halos of interest in the original tree. In this case, the match between trial tree and original tree halo masses is sufficiently close to be deemed acceptable. *Panel* **c**: The accepted trial tree is grafted into the original tree. Note that the augmented node and its progenitors from the original tree are retained (so their masses are unchanged), but now with the structure of the trial tree grafted between them. *Panel* **d**: Where the trial tree has halos at z_1 which did not match any halo in the existing tree (and are below the resolution of the original tree by construction), a new tree is grown from each such halo and attached. This process is repeated for each of the green halos in the original tree.

trees in this work as explained in Section 3.2. Each node is described by a mass and quantized epoch (i.e. a redshift taken from the set of snapshot redshifts, $z_0, \ldots, z_N$ (with $z_0 < z_1 < z_2 \cdots$), output by the N-body simulation), and a collection of zero or more immediate progenitor nodes (i.e. those connected nodes existing at the immediate prior snapshot redshift). In the Parkinson et al. (2008) algorithm the statistics of progenitors of a halo depend only on the mass and redshift of that halo. Therefore, with no loss of generality, for any given node we relabel redshifts such that the halo exists at z_0. Therefore, we label the node with the tuple $(M_0; z_0)$, and the masses and redshift of its immediate progenitors as $(M_1, M_2, \ldots, M_n; z_1)$, where n is the number of progenitors and $M_1 \geq M_2 \geq M_3 \ldots$. Should a halo have one or more immediate progenitors which exist at $z > z_1$ - which may happen if the halo finding algorithm used to build the tree was unable to locate at z_1 a progenitor which *was* found at $z_{>1}$ - we interpolate that progenitor to z_1 (assuming linear growth of mass with time for the primary progenitor, and no mass growth for non-primary progenitors).

Our procedure for augmenting the tree, which is shown schematically in Figure 1, is as follows. We first prune branches of the merger tree which begin in halos with masses below a threshold, $M_{\rm cut}$, chosen such that above this mass the structure of the tree is reliably determined and not affected by the limited resolution of the simulation. Then, we visit each node in turn and apply the following procedure:

1. Grow a merger tree to the desired mass resolution, $M_{\rm res}$, using the Parkinson et al. (2008) algorithm, starting from a base halo of $(M_0; z_0)$ and stopping at z_1 (or with no stopping redshift in the case of a node at the earliest snapshot of the simulation, z_N).
2. We label the masses of the progenitors at z_1 in this trial tree $(M'_1, M'_2, \ldots, M'_{n'})$, with $M'_1 \geq M'_2 \geq M'_3 \ldots$.
3. Accept this trial tree as a match to the original tree structure if:
   (a) $n' \geq n$;
   (b) $|M'_i - M_i| < \epsilon M_i$ for $i = 1, \ldots, n$;
   (c) If $n' > n$, $M'_i < M_{\rm cut}$ for all $i = n+1, \ldots, n'$.
4. If the tree is accepted graft it in to the original tree and proceed to the next node, otherwise, return to step 1 and create a new trial tree.

In this procedure, ϵ is a parameter which controls the mass precision required in matching the information in the original tree. Once an acceptable tree is found it is grafted into the original tree by simply replacing the base node of the new tree with the node currently being augmented in the original tree, and the progenitor nodes at z_1 in the new tree with the progenitors of the current node in mass-ranked order. Note that we keep the masses of the original tree nodes, we do not replace them with the masses of the matched nodes in the new tree (which will differ from those in the original tree by up to a factor of $1 + \epsilon$). In this way the original information in the tree is preserved while adding in higher mass resolution branches and, potentially, a higher time resolution between redshift snapshots[a] (which can also be important for convergence in galaxy formation models; Benson et al. 2012).

After grafting in the new branches, if $n' > n$ we visit each unmatched progenitor node in the new tree which, by definition, has $M_{\rm res} < M_i < M_{\rm cut}$, and grow a new tree based at $(M_i; z_1)$ with no limit on redshift. This tree is then attached to its progenitor node.

We have implemented this algorithm in the Galacticus galaxy formation code (Benson 2011, 2012) as an operator acting on merger trees which can optionally be applied prior to galaxy formation calculations are performed, and will utilize this code to test the convergence properties of the statistics of merger trees augmented by our algorithm. We also utilize the Galacticus code[b] to predict properties of galaxies which form in augmented and unaugmented merger trees in order to explore the how well convergence in galaxy properties is achieved using our augmenting procedure.

2.1 Application to Millennium simulation trees

To demonstrate our method, and the convergence in galaxy properties achieved, we apply this algorithm to merger trees extracted from the Millennium simulation (Springel et al. 2005). Specifically, we extract merger trees from the `MPAHaloTrees..MR` table of the Millennium database[c] (Lemson and The Virgo Consortium 2006). We choose $M_{\rm cut} = 7.08 \times 10^{10} {\rm M}_\odot$, corresponding to 60 particles. We find that using a lower $M_{\rm cut}$ results in incorrect progenitor mass functions in the augmented trees, suggesting that these masses of lower mass halos are insufficiently reliable for our purposes. We choose an initial value for $\epsilon = \epsilon_0 \equiv 0.15$ by default - we explore in Section 3.1 how sensitive the results are to the choice of this parameter. If a matched tree is not found after $N_{\rm t} = 50$ trials, we increase $\epsilon \rightarrow \epsilon(1 + \epsilon_0)$ and continue until a match is found.[d] The speed of our algorithm for a given level of convergence will be determined by the interplay of ϵ_0 and $N_{\rm t}$. We explore how often ϵ must be increased to find a match in Section 3.1. We augment these trees to a variety of mass resolutions.

Additionally, we also run our galaxy formation model on the Millennium-II (Boylan-Kolchin et al. 2009) merger trees (specifically those from the `MPAHaloTrees..MRII` table). We prune these trees back to $M_{\rm cut} = 5.65 \times 10^8 {\rm M}_\odot$ (corresponding to 60 particles), but do not augment them - using them instead as a reference sample of higher resolution than the Millennium trees. We then augment the Millennium trees to the same resolution of $5.65 \times 10^8 {\rm M}_\odot$ so that they can be compared directly with the Millennium-II trees.

3 Results

To demonstrate our method, we first examine convergence properties in dark matter halo conditional mass functions (Section 3.1). We then explore convergence in galaxy properties as the augmented mass resolution is increased (Section 3.2). Finally, we augment the Millennium simulation to the resolution of the Millennium-II simulation and examine whether consistent results are obtained (Section 3.3).

3.1 Numerical convergence

Since our algorithm utilizes PCH trees to perform augmenting, a simple test of whether our algorithm works as expected is to compare the statistics of PCH trees grown directly to some high resolution with those of PCH trees grown initially to some lower resolution and then augmented to the higher resolution. Figure 2 shows an example of such a test, and clearly shows that the conditional mass function of augmented trees agrees very closely with that of trees grown directly to the same resolution.

Our method has two adjustable, numerical parameters, ϵ_0 and N_t, which control the precision required in matching halo masses between the original and proposed trees and the number of trial trees generated between successive reductions in the match precision required. To explore the sensitivity of our method to ϵ_0 we apply it to merger trees from a subset of the Millennium simulation. Trees are pruned as described above, then augmented to a resolution of $2.5 \times 10^9 \mathrm{M}_\odot$. We repeat this process for $\epsilon_0 = 0.1500, 0.0750$, and 0.0375, and construct conditional mass functions of the resulting merger trees.

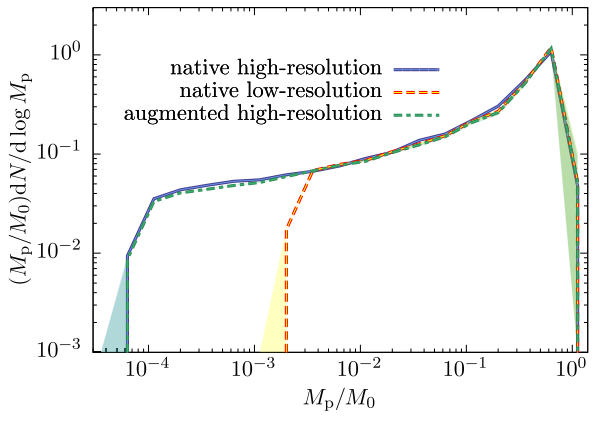

Figure 2 The conditional mass function at $z = 1.08$ of halos in PCH trees which merge to become part of halos of mass $\log_{10}(M_\mathrm{p}/\mathrm{M}_\odot) = 13.42$-$13.66$ by $z = 0$, as a function of the progenitor halo mass normalized to the $z = 0$ halo mass. The blue line indicates trees grown directly to a resolution of $M_\mathrm{res} = 2.5 \times 10^9 \mathrm{M}_\odot$ (with the shaded region indicating the 1-σ Poisson error bars due to the finite number of trees analyzed). The red line shows results for trees grown to a lower resolution of $M_\mathrm{res} = 7.1 \times 10^{10} \mathrm{M}_\odot$. Finally, the green line shows results when these lower resolution trees are augmented to $M_\mathrm{res} = 2.5 \times 10^9 \mathrm{M}_\odot$. The augmenting tolerance parameter was set to $\epsilon_0 = 0.15$.

The left-panel of Figure 3 shows convergence in the conditional mass function with the ϵ_0 parameter. Specifically, we show the conditional mass function at $z = 1.08$ of halos which merge to become part of halos of mass $\log_{10}(M_\mathrm{p}/\mathrm{M}_\odot) = 13.42$-$13.66$ by $z = 0$, as a function of the progenitor halo mass normalized to the $z = 0$ halo mass. The blue line indicates the original Millennium trees (with the shaded region indicating the 1-σ Poisson error bars due to the finite number of trees analyzed), while the pale pink line shows results for trees after mass pruning to a resolution of $7.08 \times 10^{10} \mathrm{M}_\odot$. The three green lines indicate results for trees augmented to a resolution of $2.5 \times 10^9 \mathrm{M}_\odot$ using our method for three different values of ϵ_0, as indicated in the panel.

We find that 98.5%, 96.9%, and 92.0% of branches were matched to within the original tolerance for $\epsilon_0 = 0.1500, 0.0750$, and 0.0375 respectively - even though we allow this tolerance to fall off if a matching tree is not found after $N_\mathrm{t} = 50$ trials it is clear that the majority of cases are actually matched at the required initial tolerance. In particular, we find that the conditional mass function is well-converged over the full range of masses as ϵ_0 is varied. This indicates that a value of $\epsilon_0 = 0.15$ is sufficient to ensure convergence.

Considering the parameter N_t, we find that, for $\epsilon_0 = 0.15$, for nodes with only a single progenitor node a match is found after an average of 1.8 trial trees. These cases are typically easy to match as (by selection) these nodes do not have any mergers above the resolution limit over the period being augmented. For nodes with two or more progenitor halos matching becomes more challenging. In those cases we find that a match is found after an average of 20 trial trees. We find that as ϵ_0 is decreased this mean number of trial trees before a match is found increases only slowly (e.g. it is 22 for $\epsilon_0 = 0.0375$). If we increase N_t we find that the mean number of trial trees before a match is found also increases. For example, with $\epsilon_0 = 0.15$, the mean number of trial trees[e] is 19.6, 33.8, 52.3 for $N_\mathrm{t} = 50, 100, 200$ respectively. Since each trial is independent, if the probability for any given tree to match successfully is p, then the probability distribution for a match after n trials is simply $f(n) = p(1-p)^{n-1}$. The mean number of trials, $\langle f(n) \rangle$ is then approximately consistent with the above results for $N_\mathrm{t} = 50, 100$, and 200 if $p = 0.0167$. For $\epsilon_0 = 0.0375$ we find $p \approx 0.0067$. Based on these results, it is clear that N_t could be increased without significant loss of speed, and with some improvement in accuracy. Specifically, for $\epsilon_0 = 0.15$ and the specific tree resolutions considered in these tests, $N_\mathrm{t} = 50$ will result in around 55% of two-or-more progenitor branches being matched at the original tolerance, ϵ_0, while to have 90% of such branches matched at the original tolerance would required $N_\mathrm{t} \approx 140$. As we will show

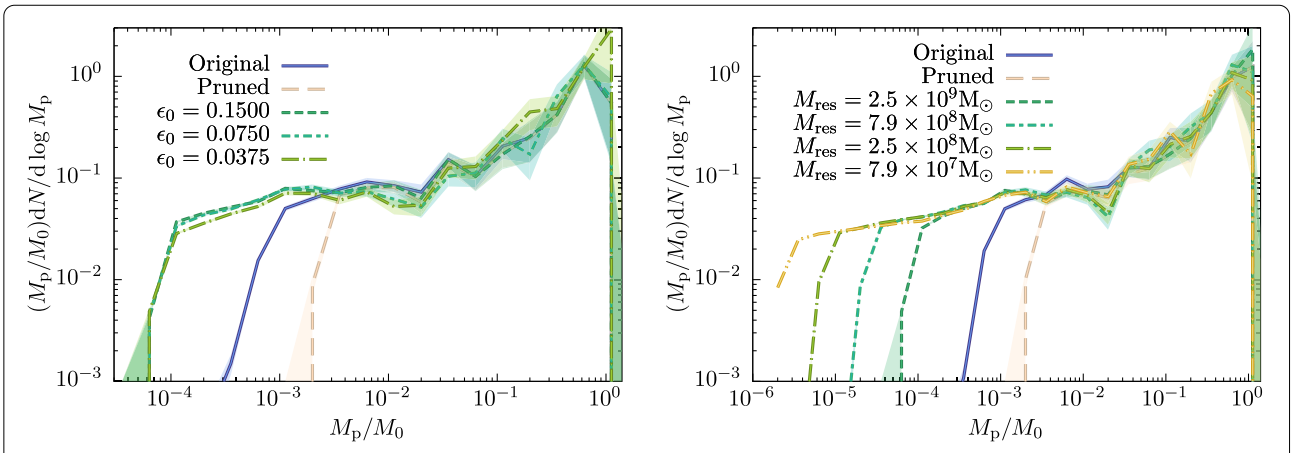

Figure 3 Convergence of the conditional mass function. *Left panel*: Convergence of the conditional mass function with the ϵ_0 parameter. Lines show the conditional mass function at $z = 1.08$ of halos which merge to become part of halos of mass $\log_{10}(M_\mathrm{p}/\mathrm{M}_\odot) = 13.42$-$13.66$ by $z = 0$, as a function of the progenitor halo mass normalized to the $z = 0$ halo mass. The blue line indicates the original Millennium trees (with the shaded region indicating the 1-σ Poisson error bars due to the finite number of trees analyzed), while the pale pink line shows results for trees after mass pruning to a resolution of $7.08 \times 10^{10}\mathrm{M}_\odot$. The three green lines indicate results for trees augmented to a resolution of $2.5 \times 10^9\mathrm{M}_\odot$ using our method for three different values of ϵ_0, as indicated in the panel. *Right panel*: The same conditional mass function as in the left panel, with blue and pale pink lines again showing results for the original and pruned trees. Remaining lines show results after augmenting trees to different mass resolutions as indicated in the panel. All cases use $\epsilon_0 = 0.15$.

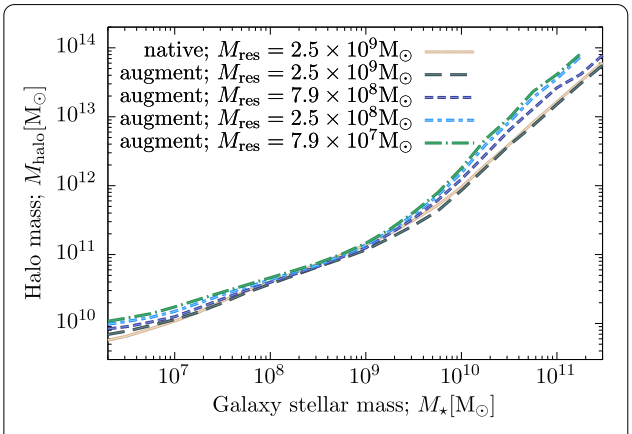

Figure 4 The stellar mass vs. halo mass relation as a function of augmented tree mass resolution. The 'native' line shows results for trees built to a resolution of $2.5 \times 10^9\mathrm{M}_\odot$ entirely using the PCH algorithm. Other lines show results for Millennium simulation merger trees augmented to different mass resolutions as indicated in the figure.

below, $N_\mathrm{t} = 50$ is sufficient to achieve good convergence in all tests that we consider.

The right-panel of Figure 3 shows the same conditional mass function, but now explores convergence as the augmented mass resolution is changed at fixed $\epsilon_0 = 0.15$. As mass resolution in the augmented branches is increased the conditional mass function clearly converges at fixed mass ratio (of course, at higher resolution the conditional mass function is populated down to lower mass ratio). This demonstrates that our method is stable as mass resolution is changed.

3.2 Galaxy property convergence

To explore convergence in galaxy properties we examine the stellar mass vs. halo mass relation. As mentioned in Section 2 we use halo trees in this work, as opposed to subhalo trees. The Galacticus model used in this work is configured to not require any information from N-body subhalos - merging times for subhalos (whether these were part of the original trees, or high resolution subhalos grafted into those trees) are computed using the fitting function of Jiang et al. (2008) with orbital parameters drawn from the distribution of Benson (2005). Our augmenting method will also work with subhalo trees. However, if N-body subhalo information (such as positions, orbital parameters, etc.) from such trees were used by a semi-analytic model, then equivalent information for the high resolution grafted subhalos would have to be treated via (semi-)analytic methods. Such an approach will be the subject of a future work.

Figure 4 shows the stellar mass vs. halo mass relation as computed by the Galacticus model. Note that we are not interested in whether this relation (which depends on the specific implementation of baryonic physics in our model) agrees well with observations, but merely whether it converges as we augment trees to higher resolution. Of course, the details of convergence will depend on the details of the galaxy formation physics modeled. However, providing that galaxy formation becomes inefficient in some sufficiently low mass halos convergence should

always be reachable. The 'native' lines shows results for merger trees built entirely via the PCH algorithm, with $M_{\rm res} = 2.5 \times 10^9 \mathrm{M}_\odot$. Other lines show results from Millennium merger trees augmented to various different resolutions as indicated in the figure.

When augmented to a mass resolution of $2.5 \times 10^9 \mathrm{M}_\odot$ the Millennium trees produce results in good agreement with the 'native' trees. As the resolution is increased there is a clear systematic shift in the stellar mass vs. halo mass relation - with the halo mass at fixed stellar mass increasing at both low and high masses, while remaining almost constant around the bend in the relation at a few times $10^9 \mathrm{M}_\odot$ in stellar mass. Importantly though, it is clear that the relation is converging as the tree resolution is increased. This is expected - in this particular model infall of gas into halos from the intergalactic medium is inhibited in halos with circular velocities at their virial radius below 35 km s^{-1} (corresponding to halo masses of $2.05 \times 10^{10} \mathrm{M}_\odot$ at $z = 0$). Once the tree resolution is high enough to resolve all such halos we expect convergence in the results.

3.3 Millennium vs. Millennium-II

As a further demonstration of our method, we use it to augment merger trees from the Millennium simulation (mass resolution $7.08 \times 10^{10} \mathrm{M}_\odot$ after pruning) to match the mass resolution of the Millennium-II simulation ($5.65 \times 10^8 \mathrm{M}_\odot$ after pruning). We augment roughly 3.5% of the volume of the Millennium simulation. Since the Millennium-II simulation has a total volume 125 times smaller than the Millennium simulation this means our augmented 3.5% of the Millennium volume corresponding to roughly 439% (=125 × 3.5%) of the Millennium-II simulation volume. We then run our galaxy formation model on these augmented trees to predict galaxy properties.

In Figure 5 we show stellar mass functions of galaxies in the Millennium and Millennium-II simulations, nor-

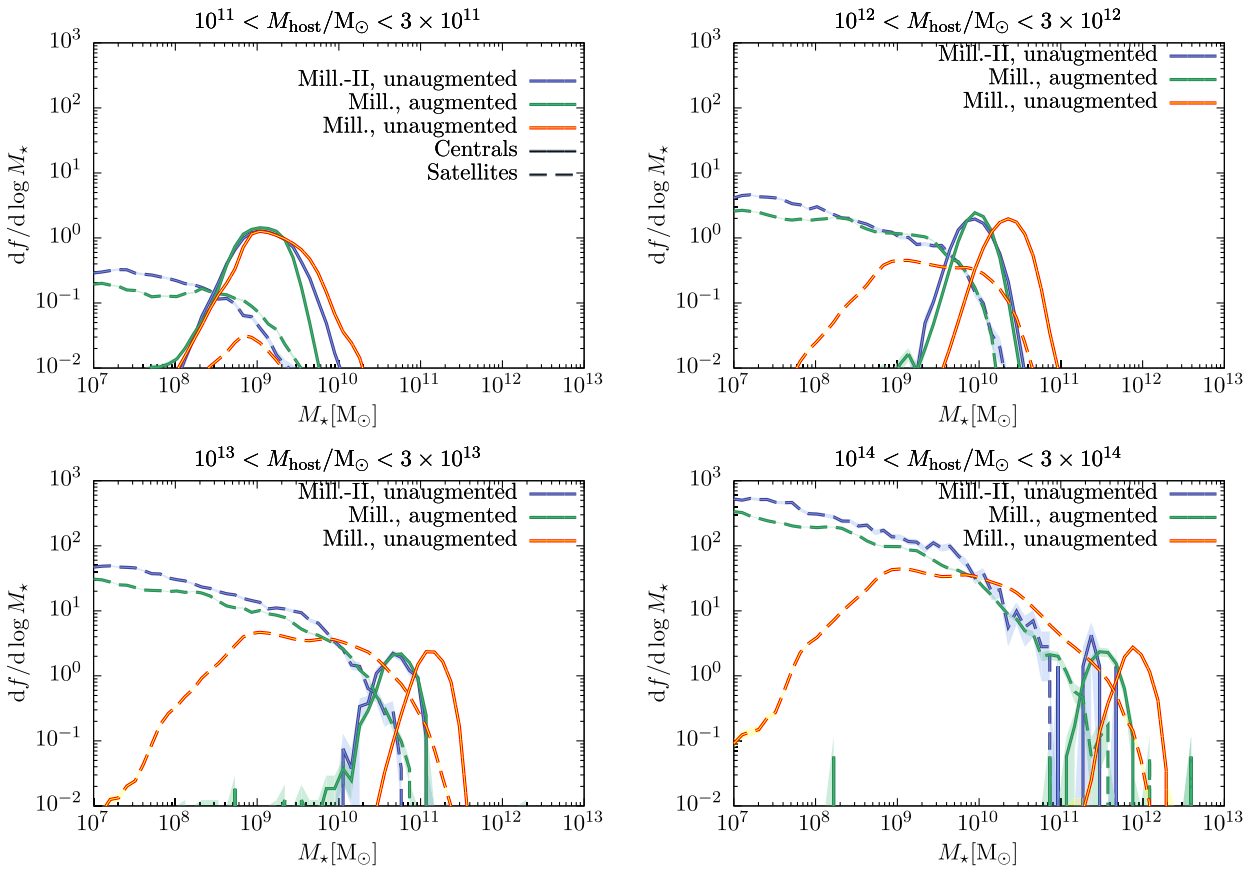

Figure 5 Stellar mass functions of galaxies in the Millennium and Millennium-II simulations. Each panel shows the distribution of galaxy stellar masses in $z = 0$ halos in a narrow range of masses (as shown above each panel), normalized to show the average number of galaxies within a single $z = 0$ halo. Solid lines indicate masses of central galaxies, while dashed lines indicate satellite galaxies. Red lines show results for higher resolution, unaugmented (but pruned) Millennium simulation trees, blue lines show results for unaugmented (but pruned) Millennium-II simulation trees, and green lines show results for Millennium trees augmented to match the resolution of the Millennium-II trees. Shaded bands indicate 1-σ Poisson errors on the mass functions arising from the finite number of merger trees analyzed.

malized to show the average number of galaxies within a single $z = 0$ halo. Each panel shows the distribution of galaxy stellar masses in $z = 0$ halos in a narrow range of masses (as shown above each panel). Solid lines indicate masses of central galaxies, while dashed lines indicate satellite galaxies. Red lines show results for unaugmented (but pruned) Millennium simulation trees, blue lines show results for unaugmented (but pruned) Millennium-II simulation trees, and green lines show results for Millennium trees augmented to match the resolution of the Millennium-II trees. The unaugmented Millennium and Millennium-II trees show clear differences in these mass functions, which must arise due to their different resolutions. The augmented Millennium trees produce results clearly much closer to those from Millennium-II trees, although some differences remains. Before assessing the success of the augmenting algorithm in this case we will first examine whether the residual effects of mass resolution are plausibly affecting these results.

To assess the expected effects of resolution on this plot we identify galaxies in halos of mass between 1 and 2 times the resolution limit of the pruned Millennium-II trees. We then find the distribution of galaxy stellar masses in such halos and identify the 84th percentile of that distribution. We find this to be approximately at $M_\star = 3.7 \times 10^6 \mathrm{M}_\odot$. As defined, this resolution limit therefore lies below the range plotted in Figure 5. Furthermore, we can select galaxies in the stellar mass range $M_\star = 1\text{-}2 \times 10^7 \mathrm{M}_\odot$ (i.e. the lowest masses shown in Figure 5) and examine the distribution of halo masses. We find that the 16th percentile of this distribution is $M_\mathrm{halo} = 8.7 \times 10^9 \mathrm{M}_\odot$ - over 10 times higher than the resolution of our pruned Millennium-II trees. Nevertheless, this does not guarantee that the mass functions shown in Figure 5 will be completely free from resolution artifacts, since these $8.7 \times 10^9 \mathrm{M}_\odot$ halos are, of course, built from yet smaller halos closer to the resolution threshold.

Focusing first on central galaxies we see that such galaxies in unaugmented Millennium trees are biased to higher stellar mass (at fixed halo mass) than those in (unaugmented) Millennium-II trees. In this specific model, this occurs because baryons are locked up as stars and interstellar medium in lower masses galaxies in the high resolution Millennium-II trees, making them unavailable in the formation of the central galaxy. In the lower resolution Millennium trees, these baryons remain available for incorporation into the central galaxy since the low mass galaxies never form. However, after augmenting the Millennium trees to match the resolution of the Millennium-II trees we see that central galaxy stellar mass distributions are almost identical in Millennium and Millennium-II.

For satellite galaxies we see a similar behavior - in unaugmented Millennium trees satellite galaxies tend to be biased to higher masses than their counterparts in Millennium-II trees. After augmenting the satellite mass functions agree closely at high masses in Millennium and Millennium-II, although we see some differences between the satellite mass functions, even after augmenting, at lower masses. This is particularly noticeable in the top-left panel for higher mass central galaxies, and is apparent in all panels for low mass satellite galaxies. This may be due to limitations of our approach, but may also simply reflect differences in how resolution effects depend on mass in N-body and PCH trees. To investigate these possibilities we repeat this study but instead of using trees extracted from the Millennium and Millennium-II simulation we generate a set of PCH trees matched to the resolutions of those two simulations (referring to these as 'pseudo-Millennium' and 'pseudo-Millennium-II' trees) and then attempted to augment the lower resolution trees to match the higher resolution trees. In this case, by construction, the mass dependence of the resolution cut-off in tree branches is identical between the two sets of trees. We find that the satellite mass functions agree very well between augmented pseudo-Millennium and pseudo-Millennium-II trees. As such, differences in the mass-dependence of the resolution cut-off between PCH and N-body trees seem to be the cause of the differences in the low mass regions of satellite mass functions in Figure 5 - in the top left panel the mass to which the Millennium trees are pruned before augmenting, $7.08 \times 10^{10} \mathrm{M}_\odot$, is very close to the range of halo masses plotted, $1\text{-}3 \times 10^{11} \mathrm{M}_\odot$, so it is not too surprising that resolution has some residual effects.

In any case, we expect that in typical applications of our approach the important issue will be that galaxies in halos above the resolution limit of the original simulation (corresponding to the low mass turnovers in the unaugmented Millennium mass functions) have converged properties - since it is these halos for which spatial information is available. Clearly, in this example such convergence is achieved.

4 Conclusions

We have described a simple yet powerful way to augment the resolution of merger trees extracted from N-body simulations of structure formation by grafting in high resolution branches generated using the PCH algorithm. These grafted branches are chosen to be consistent with the existing halo masses of the tree, while providing statistically representative structure and halo masses beyond the resolution of the original tree.

We have demonstrated that our method produces results that are converged with respect to its numerical tolerance parameter, and with the mass resolution to which trees are augmented. Additionally, by applying a semi-analytic galaxy formation model to the augmented trees we have shown that galaxy properties converge as tree resolution is increased, and that we can successfully augment trees from a low resolution simulation to match the resolution of a higher resolution simulation and that the resulting

galaxy properties are in excellent agreement between the two cases.

The approach of applying Markov Chain Monte Carlo techniques to semi-analytic models of galaxy formation in order to constrain their parameters and match the statistical properties of the observed galaxy population (Henriques et al. 2009; Bower et al. 2010; Lu et al. 2011; Mutch et al. 2013; Benson 2014; Ruiz et al. 2015) has become invaluable in allowing the construction of accurate galaxy catalogs, and for exploring the physics of galaxy formation. Such calibrations should of course be carried out using merger trees of sufficiently high resolution that the galaxy properties are converged. Unfortunately, if the constrained model is then to be applied to a cosmological N-body simulation, typically with lower resolution, the resulting galaxy properties may shift away from their constraints. Our approach solves this problem by allowing sufficiently high resolution to be attained even in low resolution cosmological simulations.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
AJB assisted in development of the implementation of the algorithm within GALACTICUS, performed the Millennium simulation tests, and wrote the majority of the text. CC developed the code to implement the augmenting algorithm within the GALACTICUS code, carried out tests of convergence of the algorithm, and contributed to the text of the paper. SC originated the idea of our algorithm, and contributed to the text of the paper.

Author details
[1]Carnegie Observatories, 813 Santa Barbara Street, Pasadena, CA 91101, USA. [2]California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA. [3]Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham, DH1 3LE, UK.

Acknowledgements
The Millennium Simulation databases used in this paper and the web application providing online access to them were constructed as part of the activities of the German Astrophysical Virtual Observatory (GAVO).
SC acknowledges support by the Science and Technology Facilities Council [ST/L00075X/1]. AJB acknowledges valuable discussion with Yu Lu. We acknowledge the support of the Ahmanson Foundation through the provision of computational resources used in this work.

Endnotes
[a] This higher time resolution is achieved across all halo masses in the tree, not just in the new high resolution branches. The trial trees created in our augmenting procedure contain high time resolution structure across all halo mass scales - from the mass resolution of our augmenting procedure up to the mass of the halo being augmented. When an accepted trial tree is grafted into the original tree the full structure is retained, including halos above the resolution limit of the original simulation but which exist in between snapshots of that simulation.

[b] Specifically for this work we used GALACTICUS revision `e7e891a6b00c740322d3131c31af818ad1e8686e` which can be obtained from the GALACTICUS repository at https://bitbucket.org/abensonca/galacticus.

[c] http://gavo.mpa-garching.mpg.de/MyMillennium.

[d] As trial trees are generated, we keep a copy of the trial tree with the best match to the original tree masses found so far. If, after increasing ϵ this tree becomes a sufficiently good match we use it instead of generating any further trial trees.

[e] These means include only cases where a successful match was found within N_t trials.

Received: 7 April 2016 Accepted: 17 August 2016
Published online: 22 August 2016

References
Baugh, CM: A primer on hierarchical galaxy formation: the semi-analytical approach. Rep. Prog. Phys. **69**, 3101-3156 (2006)
Benson, AJ: Orbital parameters of infalling dark matter substructures. Mon. Not. R. Astron. Soc. **358**, 551-562 (2005)
Benson, A: Galacticus: a semi-analytic model of galaxy formation. Astrophysics Source Code Library (2011). acsl:1108.004
Benson, AJ: Galacticus: a semi-analytic model of galaxy formation. New Astron. **17**, 175-197 (2012). doi:10.1016/j.newast.2011.07.004
Benson, AJ: Building a predictive model of galaxy formation - I. Phenomenological model constrained to the $z = 0$ stellar mass function. Mon. Not. R. Astron. Soc. **444**, 2599-2636 (2014). doi:10.1093/mnras/stu1630
Benson, AJ, Borgani, S, De Lucia, G, Boylan-Kolchin, M, Monaco, P: Convergence of galaxy properties with merger tree temporal resolution. Mon. Not. R. Astron. Soc. **419**, 3590-3603 (2012). doi:10.1111/j.1365-2966.2011.20002.x
Bernyk, M, Croton, DJ, Tonini, C, Hodkinson, L, Hassan, AH, Garel, T, Duffy, AR, Mutch, SJ, Poole, GB, Hegarty, S: The theoretical astrophysical observatory: cloud-based mock galaxy catalogs. Astrophys. J. **223**(1), 9 (2016). doi:10.3847/0067-0049/223/1/9
Bower, RG, Benson, AJ, Malbon, R, Helly, JC, Frenk, CS, Baugh, CM, Cole, S, Lacey, CG: Breaking the hierarchy of galaxy formation. Mon. Not. R. Astron. Soc. **370**, 645-655 (2006)
Bower, RG, Vernon, I, Goldstein, M, Benson, AJ, Lacey, CG, Baugh, CM, Cole, S, Frenk, CS: The parameter space of galaxy formation. Mon. Not. R. Astron. Soc. **407**, 2017-2045 (2010). doi:10.1111/j.1365-2966.2010.16991.x
Boylan-Kolchin, M, Springel, V, White, SDM, Jenkins, A, Lemson, G: Resolving cosmic structure formation with the Millennium-II simulation. Mon. Not. R. Astron. Soc. **398**, 1150-1164 (2009)
Farrow, DJ, Cole, S, Norberg, P, Metcalfe, N, Baldry, I, Bland-Hawthorn, J, Brown, MJI, Hopkins, AM, Lacey, CG, Liske, J, Loveday, J, Palamara, DP, Robotham, ASG, Sridhar, S: Galaxy and mass assembly (GAMA): projected galaxy clustering. Mon. Not. R. Astron. Soc. **454**, 2120-2145 (2015). doi:10.1093/mnras/stv2075
Helly, JC, Cole, S, Frenk, CS, Baugh, CM, Benson, A, Lacey, C: Galaxy formation using halo merger histories taken from N-body simulations. Mon. Not. R. Astron. Soc. **338**, 903-912 (2003)
Henriques, BMB, Thomas, PA, Oliver, S, Roseboom, I: Monte Carlo Markov chain parameter estimation in semi-analytic models of galaxy formation. Mon. Not. R. Astron. Soc. **396**, 535-547 (2009). doi:10.1111/j.1365-2966.2009.14730.x
Jiang, CY, Jing, YP, Faltenbacher, A, Lin, WP, Li, C: A fitting formula for the merger timescale of galaxies in hierarchical clustering. Astrophys. J. **675**, 1095-1105 (2008)
Kauffmann, G, Colberg, JM, Diaferio, A, White, SDM: Clustering of galaxies in a hierarchical universe - I. Methods and results at $z = 0$. Mon. Not. R. Astron. Soc. **303**, 188-206 (1999). doi:10.1046/j.1365-8711.1999.02202.x
Klypin, A, Yepes, G, Gottlöber, S, Prada, F, Heß, S: MultiDark simulations: the story of dark matter halo concentrations and density profiles. Mon. Not. R. Astron. Soc. **457**, 4340-4359 (2016). doi:10.1093/mnras/stw248
Lemson, G, The Virgo Consortium: Halo and galaxy formation histories from the Millennium simulation: public release of a VO-oriented and SQL-queryable database for studying the evolution of galaxies in the LambdaCDM cosmogony (2006). arXiv:astro-ph/0608019
Lu, Y, Mo, HJ, Weinberg, MD, Katz, N: A Bayesian approach to the semi-analytic model of galaxy formation: methodology. Mon. Not. R. Astron. Soc. **416**, 1949-1964 (2011). doi:10.1111/j.1365-2966.2011.19170.x
Mutch, SJ, Poole, GB, Croton, DJ: Constraining the last 7 billion years of galaxy evolution in semi-analytic models. Mon. Not. R. Astron. Soc. **428**, 2001-2016 (2013). doi:10.1093/mnras/sts182
Parkinson, H, Cole, S, Helly, J: Generating dark matter halo merger trees. Mon. Not. R. Astron. Soc. **383**, 557-564 (2008)
Ruiz, AN, Cora, SA, Padilla, ND, Domínguez, MJ, Vega-Martínez, CA, Tecce, TE, Orsi, A, Yaryura, Y, García Lambas, D, Gargiulo, ID, Muñoz Arancibia, AM: Calibration of semi-analytic models of galaxy formation using particle

swarm optimization. Astrophys. J. **801**, 139 (2015). doi:10.1088/0004-637X/801/2/139

Springel, V, White, SDM, Jenkins, A, Frenk, CS, Yoshida, N, Gao, L, Navarro, J, Thacker, R, Croton, D, Helly, J, Peacock, JA, Cole, S, Thomas, P, Couchman, H, Evrard, A, Colberg, J, Pearce, F: Simulations of the formation, evolution and clustering of galaxies and quasars. Nature **435**, 629-636 (2005). doi:10.1038/nature03597

Velliscig, M, Cacciato, M, Schaye, J, Crain, RA: The alignment and shape of dark matter, stellar, and hot gas distributions in the EAGLE and cosmo-OWLS simulations. Mon. Not. R. Astron. Soc. **453**, 721-738 (2015). doi:10.1093/mnras/stv1690