Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **467,** 4872–4885 (2017)
Advance Access publication 2017 February 16

doi:10.1093/mnras/stx370

Forward and backward galaxy evolution in comoving cumulative number density space

Paul Torrey,[1]★† Sarah Wellons,[2] Chung-Pei Ma,[3] Philip F. Hopkins[4] and Mark Vogelsberger[1]

[1]*MIT Kavli Institute for Astrophysics & Space Research, Cambridge, MA 02139, USA*
[2]*Harvard–Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA*
[3]*Astronomy Department, University of California at Berkeley, Berkeley, CA 94720, USA*
[4]*TAPIR, Mailcode 350-17, California Institute of Technology, Pasadena, CA 91125, USA*

Accepted 2017 February 9. Received 2017 February 8; in original form 2016 June 23

ABSTRACT

Galaxy cumulative comoving number density is commonly used to forge progenitor/descendant links between observed galaxy populations at different epochs. However, this method breaks down in the presence of galaxy mergers, or when galaxies experience stochastic growth rates. We present a simple analytic framework to treat the physical processes that drive the evolution and diffusion of galaxies within comoving number density space. The evolution in mass rank order of a galaxy population with time is influenced by (1) the non-conservative nature of total galaxy number density driven by galaxies combining in mergers (which we tabulate as a galaxy 'coagulation' rate) and (2) galaxy 'mass rank scatter' driven by stochasticity in stellar-mass growth rates from *in situ* star formation and mergers. We quantify the relative contribution of these two effects to the total mass rank order evolution using the Illustris simulation. We show that galaxy coagulation is dominant at lower redshifts and stellar masses, while scattered growth rates dominate the mass rank evolution at higher redshifts and stellar masses. For a galaxy population at $10^{10}\,\mathrm{M}_{\odot}$, coagulation has been the dominant effect since $z = 2.2$, but a galaxy population at $10^{11}\,\mathrm{M}_{\odot}$ was dominated by mass rank scatter until $z = 0.6$. We show that although the forward and backward median cumulative number density evolution tracks are asymmetric, the backward median cumulative number density evolution can be obtained by convolving the descendant distribution function with progenitor relative abundances. We tabulate fits for the median cumulative number density evolution and scatter that can be applied to improve the way galaxy populations are linked in multi-epoch observational data sets.

Key words: galaxies: abundances – galaxies: evolution – galaxies: formation – galaxies: statistics.

1 INTRODUCTION

Identifying links between progenitor and descendant galaxy populations to empirically infer galaxy evolution tracks is notoriously difficult. Progenitor/descendant links have been forged previously by linking galaxy populations at a constant luminosity (e.g. Wake et al. 2006), constant mass, or by isolating specific galaxy populations such as brightest cluster galaxies (Lidman et al. 2012; Lin et al. 2013; Shankar et al. 2015). Simple linking methods such as these become inaccurate as the galaxy population evolves with time. This inaccuracy results in biased conclusions about the size, morphology, star formation rate, quenched fraction, etc. evolution of galaxies (e.g. van Dokkum & Franx 1996, 2001; Saglia et al. 2010).

A more physically motivated linking method is to forge progenitor/descendant links at a fixed cumulative comoving number density[1] based on the cumulative stellar-mass function (van Dokkum et al. 2010; Bezanson et al. 2011; Brammer et al. 2011; Papovich et al. 2011, 2015; Patel et al. 2013; Ownsworth et al. 2014). In contrast to, e.g. fixed mass linking, the underlying assumption is

★ E-mail: ptorrey@mit.edu.
† Hubble fellow.

[1] Throughout this paper, we frequently refer to cumulative comoving number density simply as number density. However, for clarity, all analyses and plots use cumulative comoving number density $N(> M)$.

that the most massive galaxies at some redshift evolve into the most massive systems at some other redshift. Forging progenitor/descendant links at a constant comoving number density can accommodate evolution in the mass of the galaxy population and is easily performed for any data set where a galaxy stellar-mass function is available. Comoving number density analysis can lead to predictions about the mass (van Dokkum et al. 2010; Marchesini et al. 2014), star formation rate (Ownsworth et al. 2014), size (Ownsworth et al. 2014), gas fraction (Conselice et al. 2013) or morphology evolution of a galaxy population that would not have been possible under more simplistic linking assumptions.

The key feature of comoving number density analysis is that galaxy mass rankings [and hence $N(>M)$, $N(>\sigma)$ or other parameters] are less prone to changes than the galaxy properties itself. That is, galaxy masses or other properties of galaxies can evolve significantly, while assigned number density remains reasonably static among a galaxy population as long as the mass rank order among a galaxy population is preserved. This makes constant comoving number density a better metric or linking progenitor/descendant galaxy populations together at different observational epochs compared to the underlying physical properties themselves.

Although comoving number density analysis provides a good first-order approximation to forge progenitor/descendant links (Leja, van Dokkum & Franx 2013; Torrey et al. 2015), galaxies do not exactly move along constant comoving number density evolution tracks. The galaxy rank order assumption required for constant comoving number density linking to work is broken by (1) galaxy mergers and (2) stochastic growth rates. Galaxy mergers change the total number density of galaxies (e.g. Ownsworth et al. 2014). Galaxies of low mass naturally 'move up' in mass rank and change their assigned number density when higher mass galaxies coagulate during merger events. Stochasticity in growth rates (including star formation rates as well as *ex situ* growth rates from mergers) introduces an element of randomness that violates the assumption that galaxies maintain their relative mass rank order. Stochastic growth rates can lead to a change in the median mass rank of a tracked galaxy population.

Both galaxy merger rates and stochastic growth rates are naturally handled in numerical cosmological simulations. Galaxy comoving number density analysis has been analysed using semi-analytic models (Leja, van Dokkum & Franx 2013; Mundy, Conselice & Ownsworth 2015), abundance matching (Behroozi et al. 2013) and hydrodynamical simulations (Torrey et al. 2015; Clauwens, Franx & Schaye 2016; Jaacks, Finkelstein & Nagamine 2016). These studies have compared the mass and velocity dispersion evolution of galaxy populations using the explicitly tracked simulated galaxy merger trees to compare against the inferred evolution from an assumed constant comoving number density. These studies agree that constant comoving number density analysis recovers the median stellar-mass evolution of a galaxy population within a factor of a few (comoving number density recovers median masses a factor of $\sim$2–3 higher than explicitly tracked galaxies at redshift $\sim$3). However, two issues remain unresolved: (1) the factor of $\sim$2–3 error in the median mass evolution of galaxy populations is driven by a net evolution in the median number density of evolving galaxy populations, and (2) the significant scatter in the number density evolution tracks that initially similar galaxy populations follow.

Both of these issues can be accounted for using cosmological simulations (Behroozi et al. 2013; Leja et al. 2013; Torrey et al. 2015; Jaacks et al. 2016). In Torrey et al. (2015), we provided fits to the median mass and number density evolution of tracked galaxy populations. We showed that by substituting prescribed evolutionary tracks in comoving number density in place of a constant comoving number density, the merger-tree–defined median mass evolution can be obtained (but see Clauwens et al. 2016, for concerns regarding using a single comoving number density evolution track). Non-constant comoving number density progenitor/descendant links have now been applied to observational data sets (e.g. Marchesini et al. 2014; Papovich et al. 2015; Salmon et al. 2015). However, the non-constant comoving number density evolution tracks do not describe the scatter in number density evolution for tracked galaxy populations and (relatedly) require separate fits to define the forward and backward median number density evolution.

In this paper, we extend the analysis presented in Torrey et al. (2015) by quantifying the scatter in the number density evolution of tracked galaxy populations and show how this scatter links the forward and backward galaxy number density evolution tracks. The primary goals of this paper are (1) to explore the relationship that exists between forward and backward number density evolution rates via their intrinsic scatter (2) to provide tabulated rates for the number density evolution and scatter that can be applied to observational galaxy selection and analysis and (3) to consider the relative importance of galaxy coagulation and scattered/stochastic growth rates in driving number density evolution. We provide general fits to the dispersion of a galaxy population in number density space as a function of time from its initial selection based on the Illustris simulation. We also present an analytic framework that relates the forward and backward evolution of galaxies in number density space based on their scatter rates. We show that using this simple framework, the asymmetry in the forward and backward number density evolution rates can be broadly captured. Fitting functions are provided that can be applied to observational data sets to track galaxy populations, including both the median number density evolution and the scatter. We also address the relative importance of scattered growth rates and galaxy mergers in driving galaxy number density evolution.

The structure of this paper is as follows. In Section 2, we outline a basic formalism for tracking galaxy populations in number density space. This includes a simple relation between the distribution of galaxies in number density space when tracked forward and backward in time. In Section 3, we break down the total/net number density evolution rate in terms of the two underlying processes: galaxy coagulation and scattered growth rates. We quantify the relative impact of coagulation and scatter and explore the galaxy masses and redshift ranges where each effect dominates. In Section 4, we discuss our results including the potential impact for the interpretation of observational data. We conclude and summarize in Section 5.

2 ANALYTIC TREATMENT OF GALAXY NUMBER DENSITY EVOLUTION

Within a comoving volume, V, we rank the galaxies in the order of decreasing stellar mass and assign each galaxy a rank R where $R = 1, 2, 3$, etc. A galaxy with mass M – and associated rank R – will have a cumulative number density $N = R/V$. Mass, rank and cumulative number density are therefore exactly interchangeable. However, we present our analysis in terms of number density arguments because number density is less prone to changes compared to galaxy mass. Rank and number density can also be assigned via dark matter halo mass, velocity dispersion or other appropriate property. However, we have found in previous work (Torrey et al. 2015) that using stellar velocity dispersion or dark matter halo mass to assign

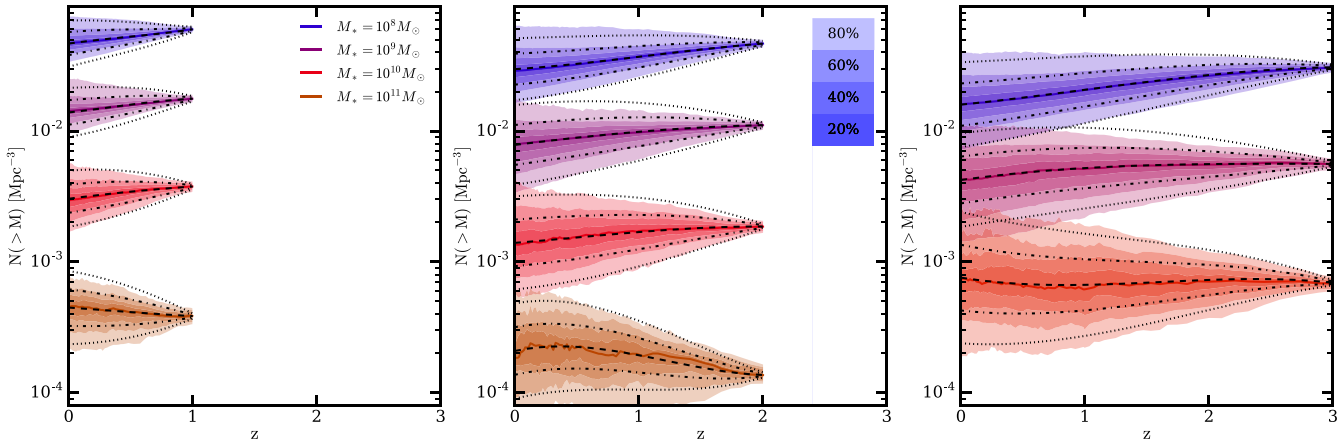

Figure 1. The number density evolution of galaxy populations is shown tracked in time from redshifts $z = 1$, 2 and 3 to redshift $z = 0$, from left-hand panel to right-hand panel, respectively. The coloured shaded bands indicate the number density evolution as directly determined from the Illustris simulation – with the shade of the colour indicating the enclosed galaxy fraction as indicated in the legend. The black dashed lines denote the best fits to the median N provided in equation (A1). The black dot–dashed and dotted lines denote regions of one and two $\sigma_{\log N}$, respectively, given in equation (A2).

rank produces similar number density evolution as stellar mass. We therefore assign rank and number density according to stellar mass throughout this paper.

We consider how a galaxy population selected within a narrow range in number density would evolve with time. Since each of the selected galaxies from the population follows its own distinct evolution in mass ordered rank, the number density of the descendant galaxy population is best described with a distribution function. The probability of a galaxy with initial number density $\mathcal{N}_0 = \log(N_0)$ at redshift z_0 evolving to have a number density $\mathcal{N}_f = \log(N_f)$ after time Δz can be described as $P(\mathcal{N}_f|\mathcal{N}_0, z_0, \Delta z)\mathrm{d}\mathcal{N}_f$. Hereafter, we refer to $P(\mathcal{N}_f|\mathcal{N}_0, z_0, \Delta z)$ as the descendant distribution function (DDF) that describes the distribution of number densities into which an initially homogeneous galaxy population evolves.

Not every galaxy in the initial population will necessarily survive until $z + \Delta z$; some will be consumed in galaxy mergers. The integral of the DDF over all descendant masses therefore yields

$$\int_0^\infty P\left(\mathcal{N}_f|\mathcal{N}_0, z_0, \Delta z\right) \mathrm{d}\mathcal{N}_f = f_\mathrm{s}\left(\mathcal{N}_0, z_0, \Delta z\right), \tag{1}$$

where f_s is the galaxy survival fraction. The mean log number density of the descendant galaxy population is given by

$$\left\langle \mathcal{N}_f \right\rangle = \frac{\int_0^\infty \mathcal{N}_f P\left(\mathcal{N}_f|\mathcal{N}_0, z_0, \Delta z\right) \mathrm{d}\mathcal{N}_f}{f_\mathrm{s}\left(\mathcal{N}_0, z_0, \Delta z\right)}. \tag{2}$$

Evaluating equation (2) requires specifying a form of the DDF that we provide in Section 2.1.

2.1 Descendant distribution functions

The DDF can be approximated empirically based on numerical simulations. We find functional forms and best fits to the DDF in this section using the Illustris simulation (Genel et al. 2014; Vogelsberger et al. 2014a,b). The Illustris simulation employs a model for galaxy formation (Vogelsberger et al. 2013; Torrey et al. 2014) that is able to broadly reproduce the evolution of the galaxy stellar-mass function out to redshift $z = 6$ (Genel et al. 2014).

Fig. 1 shows the evolution of the distribution of four galaxy populations as they are tracked in time from an initial selection redshift of $z_0 = 1$, 2 and 3 (left- to right-hand panels, respectively). The tracked galaxy populations are selected to have stellar masses of $M_* = 10^8$, 10^9, 10^{10} and $10^{11}\,\mathrm{M}_\odot$ in bins of 0.15 dex width at their initial selection redshift. The adopted bin size has a minimal impact on the resulting number density evolution fits as long as (i) the bin averages over a sufficient number of growth tracks such that noise from individual growth tracks does not affect the result and (ii) the bin size is small compared to the mass scale over which number density evolution tracks change their characteristics. We found that requiring at least 50–100 galaxies in each bin was satisfactory for removing much of the stochastic noise from the number density evolution tracks and that changing our fiducial bin size by 0.05 dex in either direction had negligible impact on our results. The masses of all galaxies are tracked forward in time using the merger trees described in Rodriguez-Gomez et al. (2015). Galaxies that are consumed in a merger event are included until they are consumed. Masses are converted to number densities by inverting the tabulated fitting functions to the cumulative galaxy stellar-mass function from Torrey et al. (2015).

The DDFs presented in Fig. 1 are reasonably well described by a lognormal distribution

$$P\left(\mathcal{N}_f|\mathcal{N}_0, z_0, \Delta z\right) = \frac{f_\mathrm{s}}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\left(\mathcal{N}_f - \left\langle \mathcal{N}_f \right\rangle\right)^2}{2\sigma^2}\right), \tag{3}$$

where $\langle \mathcal{N}_f \rangle$ is the mean number density (in log space) of the descendant population, and σ is the standard deviation. The survival fraction, f_s, mean number density $\langle \mathcal{N}_f \rangle$ and spread, σ, are functions of the initial number density, initial selection redshift and elapsed time. We construct fits to f_s, $\langle \mathcal{N}_f \rangle$, and σ that are presented in Appendix A. Fig. 1 indicates the fits to the median number density evolution track (dashed lines), the $\pm 1\sigma$ fits (dot–dashed lines) and the $\pm 2\sigma$ fits (dotted lines) for the tracked galaxy populations.

These fits accurately capture the median evolution and broadly capture the scatter evolution found in the simulation. The median number density evolution follows non-constant number density evolution tracks with evolution. The magnitude of the change in number density varies based on initial selection redshift and initial selection mass but ranges from roughly constant to changes of $\sim$0.3 dex evolution out to redshift $z = 3$. The scatter grows at an approximate rate of $\sigma \propto 0.2\Delta z$, with more detailed fits given in Appendix A.

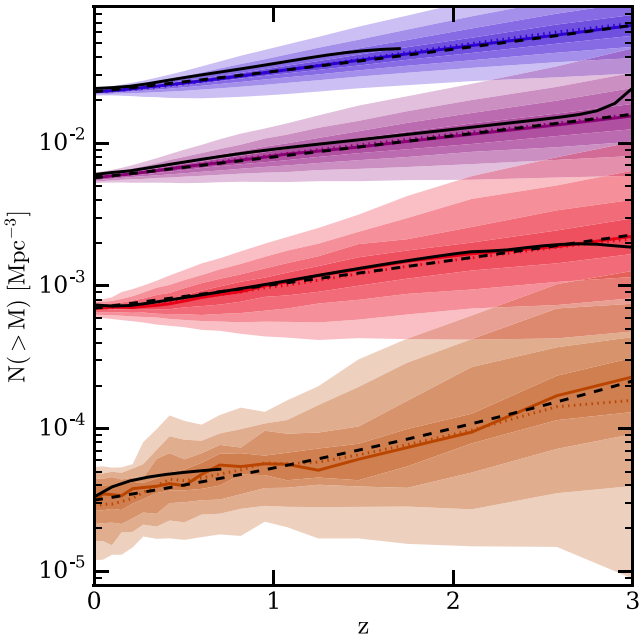

Figure 2. Same as Fig. 1 but for four galaxy populations tracked backward in time. Solid coloured lines indicate the median distribution of explicitly tracked galaxy populations via merger tree. The coloured dotted lines indicate the log-space-mean of the explicitly tracked galaxy populations via merger tree. The black solid lines indicate the inferred log-space-mean PDF based on equation (5). The black dashed lines indicate the best-fitting median progenitor number density based on equation (A4). The broad match between the explicitly tracked and inferred median progenitor number density out to redshift $z \sim 2.5$ validates the relationship between the DDFs and PDFs presented in equation (5).

2.2 Progenitor number density distribution functions

The discussion to this point has been limited to the DDF. We can similarly consider the progenitor distribution function (PDF), which can be examined using the same numerical simulations. Fig. 2 shows the PDF for four galaxy populations selected in thin mass (number density) bins at redshift $z = 0$. Direct fits to the median number density evolution and scatter evolution can be found in Appendix A. The direct fits are shown with dashed solid lines in Fig. 2 that track the simulation median number density evolution very well by construction. Fig. 2 also shows the log-space-mean number density evolution as determined from the galaxy merger trees. The log-space-mean number density evolution (dotted lines) is nearly coincident with the median number density evolution, which would be the case for a lognormal PDF.

Examination of Figs 1 and 2 shows that the median evolution of the PDF and the DDF has different slopes. This is a real consequence of the median number density evolution rate being different for galaxy populations that are tracked forward and backward in time (see fig. 6 of Torrey et al. 2015). However, the PDF and the DDF can be related by considering the mean mass rank of the progenitor galaxies that will grow into galaxies with mass rank $\mathcal{N}_f$. To achieve this, we need to consider the galaxies that will scatter/evolve into a particular descendant bin given the relative abundance of progenitors and the DDF. Specifically, the mean log number density of the progenitor galaxy population that will evolve to have a number density $\mathcal{N}_f$ after a time Δz is

$$\langle \mathcal{N}_0 \rangle = \frac{\int_0^\infty \mathcal{N}_0 P\left(\mathcal{N}_f | \mathcal{N}_0, z_0, \Delta z\right) \frac{\mathrm{d}n}{\mathrm{d}\mathcal{N}_0} \mathrm{d}\mathcal{N}_0}{\int_0^\infty P\left(\mathcal{N}_f | \mathcal{N}_0, z_0, \Delta z\right) \frac{\mathrm{d}n}{\mathrm{d}\mathcal{N}_0} \mathrm{d}\mathcal{N}_0}, \quad (4)$$

where the integration is over all possible progenitor galaxies $\mathrm{d}\mathcal{N}_0$ and $P(\mathcal{N}_0|\mathcal{N}_f, z_0, \Delta z)$ is *the same* DDF described in the previous subsection. The factor of $\mathrm{d}n/\mathrm{d}\mathcal{N}_0$ describes the relative abundance of the possible progenitor galaxies. Specifically, in analogy to the galaxy stellar-mass function – where the distribution of galaxies in stellar mass is described via $\phi = \mathrm{d}n/\mathrm{dlog}M$ – the distribution of galaxies in number density space can be described as $\mathrm{d}n/\mathrm{dlog}N$. However, unlike the stellar-mass function, galaxies are trivially distributed in number density space such that $\mathrm{d}n/\mathrm{dlog}N = N = 10^{\mathcal{N}}$. We can therefore express equation (4) as

$$\langle \mathcal{N}_0 \rangle = \frac{\int_0^\infty \mathcal{N}_0 P\left(\mathcal{N}_f | \mathcal{N}_0, z_0, \Delta z\right) 10^{\mathcal{N}_0} \mathrm{d}\mathcal{N}_0}{\int_0^\infty P\left(\mathcal{N}_f | \mathcal{N}_0, z_0, \Delta z\right) 10^{\mathcal{N}_0} \mathrm{d}\mathcal{N}_0}. \quad (5)$$

The integrands of equation (5) are performing a convolution of (a) the probability that a galaxy with initial number density $\mathcal{N}_0$ evolves into a galaxy with number density $\mathcal{N}_f$ with (b) the relative abundance of galaxies with initial number density $\mathcal{N}_0$. The relative abundance factor was not present in the descendant average mass rank calculation and is the cause of the offset between the forward and backward number density evolution tracking.

The physical asymmetry is driven by the relative overabundance of fast growth tracks owing to the higher abundance of low-mass galaxies. The absolute contribution of a progenitor bin must take into account both the likelihood that a galaxy would scatter into the desired descendant bin and the relative abundance of the progenitor galaxy population. Using equations (3) and (5), the average mass rank of the progenitor galaxy population can be inferred. The resulting inferred number density evolution trajectories are indicated with black solid lines in Fig. 2. The inferred number density evolution is shown only where the fit is well constrained. Specifically, we require that the progenitor median number density plus 2σ distribution is within the validity bounds of the single forward number density evolution fit. The inferred number density evolution is within $\sim$0.05 dex of the merger tree evolution tracks for all redshifts and mass bins where these criteria are met (the agreement deviates more significantly when contributions to the progenitor galaxy population from outside of the fitting function validity region are allowed). The inferred number density for the two intermediate-mass bins (purple and red) remains in broad qualitative agreement with the simulation data out to $z = 3$.

The errors in the inferred number density evolution are driven largely by deviations from the DDF from a strict lognormal distribution. Although the lognormal DDF assumption is in broad agreement with the empirically derived DDF from the simulations, there are asymmetric features of the DDF that become increasingly prominent as the galaxy population is tracked further in redshift. The offsets seen in Fig. 2 are small enough to give confidence that our method of linking PDFs and DDFs is correct, but will always be less accurate than the direct fits provided in Appendix A.

Finally, we note that demonstrating the time reversibility of descendant and progenitor tracking is made possible here by the definition of a single, continuous fitting function in $\mathcal{N}_0$, z_0 and Δz for the median and scatter of the DDF. Such continuous fits have not been previously provided.

3 RELATING NUMBER DENSITY EVOLUTION TO PHYSICAL PROCESSES

In this section, we discuss the physical processes that drive the evolution in the number density tracks shown previously. Specifically, we quantify the importance of galaxy coagulation and stochastic galaxy growth rates towards number density evolution and compare the relative importance of these two mechanisms.

The average (median) number density evolution discussed at length in the previous section can be described as

$$\langle N(z)\rangle = N(z_0) + \int_{z_0}^{z} \left\langle \frac{\mathrm{d}N}{\mathrm{d}z} \right\rangle \mathrm{d}z, \tag{6}$$

where $\langle \mathrm{d}N/\mathrm{d}z\rangle$ is the rate of change of the average number density in linear (not log) space. Using this form, we consider the physical processes that drive the average number density evolution. Populations of galaxies change their median mass rank either (1) because a galaxy moves up in mass rank (in order of decreasing mass) by one when two galaxies above its mass merge to form a single galaxy or (2) by rapid or slow mass growth, such that the galaxies change their mass rank relative to their peers. These two processes provide a complete basis that can capture all of the galaxy mass assigned rank order evolution. Moreover, these two processes mirror the breakdown of the two fundamental assumptions of constant comoving number density analysis: that galaxy mergers do not significantly impact the total number density of galaxies and that galaxies preserve their mass-assigned rank order in time. In the previous section, we showed how the average number density evolves for tracked galaxy populations in time. In this section, we discuss the relative importance that galaxy coagulation and scattered growth rates have on comoving number density evolution.

The median number density rate of change for a galaxy population can be expressed with explicit dependence on the two number density evolution channels

$$\left\langle \frac{\mathrm{d}N}{\mathrm{d}z} \right\rangle = \left\langle \frac{\mathrm{d}N}{\mathrm{d}z} \right\rangle_{\mathrm{c}} + \left\langle \frac{\mathrm{d}N}{\mathrm{d}z} \right\rangle_{\mathrm{s}}, \tag{7}$$

where $\langle \mathrm{d}N/\mathrm{d}z\rangle_{\mathrm{c}}$ gives the rate of change of cumulative number density owing to changes in the total galaxy number density via mergers above the mass scale of interest (hereafter: the 'galaxy coagulation' rate) and $\langle \mathrm{d}N/\mathrm{d}z\rangle_s$ gives number density rate of change from scattered growth. Equation (7) provides a rigorous and clear breakdown of the total number density evolution rate that facilitates the analysis in the subsequent subsections. However, it does not constitute a unique breakdown of the total number density evolution rate. For example, while galaxy coagulation is driven by galaxy mergers above a given mass scale (as described in detail below), the galaxy scatter rate also contains a contribution from *ex situ* mass growth driven by galaxy mergers. We therefore emphasize that our primary motivation for adopting the breakdown in equation (7) is that the two terms mirror the two fundamental assumptions of constant comoving number density analysis. In what follows, we adopt

$$\frac{\mathrm{d}N_{\mathrm{c}}}{\mathrm{d}z} \equiv \left.\frac{\mathrm{d}N}{\mathrm{d}z}\right|_{\mathrm{c}}; \quad \frac{\mathrm{d}N_{\mathrm{s}}}{\mathrm{d}z} \equiv \left.\frac{\mathrm{d}N}{\mathrm{d}z}\right|_{\mathrm{s}} \tag{8}$$

for simplicity of notation. However, we note that only the time derivatives of number density can be split into coagulation and scatter components, not number density itself.

Galaxy coagulation drives intuitive net changes in the number density of a galaxy population by making total galaxy number density a non-conserved quantity. It is also fairly intuitive that an individual galaxy can undergo a change in its mass ordered *rank* (and therefore can adjust its assigned number density) by growing much faster or slower than its peers. It is somewhat less intuitive to predict the impact that relative galaxy growth rates have on the mass rank order evolution when averaged over a galaxy population. However, as shown below, the contribution of scatter to the median mass rank of a galaxy population is significant when compared to the coagulation rate for a range of galaxy masses and redshifts.

3.1 Galaxy coagulation

A galaxy of mass M_i will necessarily change mass rank if two galaxies both with masses $M > M_i$ merge together. When this happens, the number of galaxies with mass $M > M_i$ will decrease by exactly one, forcing all lower mass galaxies to move up in mass rank. We therefore define the galaxy coagulation rate as the rate at which galaxies with mass rank higher than i are being swallowed by mergers (i.e. undergoing mergers with larger systems).[2]

We calculate the galaxy coagulation rate in the simulation by identifying all galaxies with initial mass $M > M_i$ that are not the main progenitor of a halo in a subsequent snapshot – indicating that the galaxy has been consumed by a larger system in a merger event. This number is converted to the coagulation rate using a first-order finite difference scheme with a target redshift step size of $\Delta z \approx 0.1$. The resulting coagulation rate is shown in Fig. 3. The coagulation rate is lower for high-mass systems compared to their low-mass counterparts owing simply to the lower abundance of high-mass systems; the effect of galaxy coagulation rate on number density is cumulative.

The coagulation rate $\mathrm{d}N_c/\mathrm{d}z$ is equal to the galaxy–galaxy merger rate integrated over appropriate parameters. Traditionally, the galaxy–galaxy merger rate is tabulated as the number of mergers per unit redshift per unit mass ratio per halo, $\mathrm{d}N_m/\mathrm{d}\xi\mathrm{d}z$ (e.g. Fakhouri & Ma 2008; Rodriguez-Gomez et al. 2015). The galaxy coagulation rate can be obtained via the galaxy–galaxy merger rate by integrating over the appropriate limits

$$\frac{\mathrm{d}N_c}{\mathrm{d}z}(M,z) = \int_{2M}^{\infty} \mathrm{d}M_{\mathrm{d}} \int_{\frac{M}{M_{\mathrm{d}}-M}}^{1} \mathrm{d}\xi \frac{\mathrm{d}N_m}{\mathrm{d}\xi\mathrm{d}z} \frac{\mathrm{d}n}{\mathrm{d}M_{\mathrm{d}}}, \tag{9}$$

where M_{p} is the primary galaxy mass, $\xi = M_{\mathrm{s}}/M_{\mathrm{p}}$ is the mass ratio of the secondary to the primary and $M_{\mathrm{d}} = M_{\mathrm{s}} + M_{\mathrm{p}}$ is the merger descendant mass. The additional mass function dependence in the integrand is used to convert the 'per halo' into a 'per volume'. The lower limit of $2M$ on the first integral represents the lowest possible descendant mass for two merging systems with masses $M_{\mathrm{p}} \gtrsim M_{\mathrm{s}} \gtrsim M$. Adopting the tabulated and parametrized merger rates of Rodriguez-Gomez et al. (2015) the best-fitting galaxy stellar-mass functions from the appendix of Torrey et al. (2015), and integrating equation (9) yields an agreeable result to the direct measurement presented in Fig. 3. Equation (9) reduces the three-parameter dependence of the total merger rate expression (M_{p}, ξ, z) to a two-parameter dependence on mass M and redshift z. This parameter reduction is associated with the limitations in the mass ratio that restrict mergers to only include systems that are both more massive than mass M.

[2] We briefly note that galaxy mergers also change galaxy mass, which can cause mass rank/number-density evolution. However, in our analytic framework, we explicitly separate changes in the total number density of galaxies (strictly driven by galaxy mergers) from mass rank order evolution driven by stochastic growth rates (which has contributions from *in situ* and *ex situ*/merger growth).

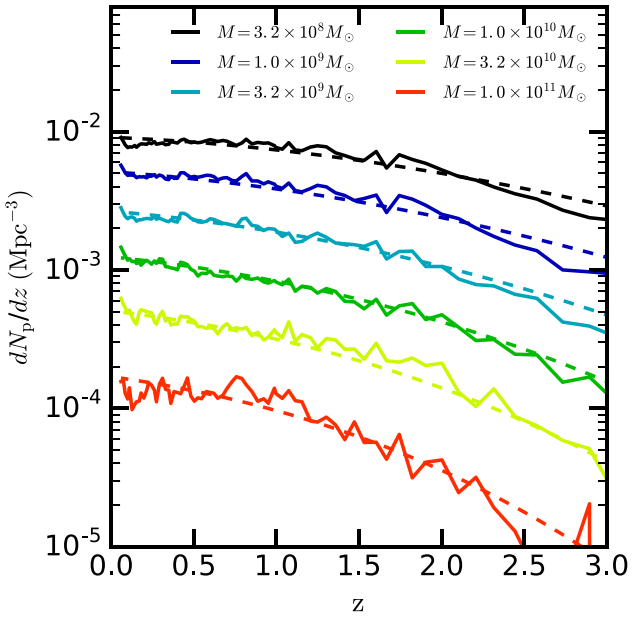

Figure 3. The galaxy coagulation rate as a function of time for several galaxy mass bins. The galaxy coagulation rate describes the rate at which a galaxy's mass rank is expected to change owing to the coagulation of galaxies with higher mass rank (higher mass). Unlike Figs 1 and 2 where we tracked galaxy population evolution, here we show the galaxy coagulation rate for the same fixed M_* at all redshifts, as indicated in the legend. Dashed lines indicate the fits used in Appendix A to assess the relative importance of galaxy coagulation and scattered growth rates. The galaxy coagulation rate can be compared against the competing galaxy scatter rate presented in Fig. 4.

3.2 Scatter rate

Variable/stochastic growth rates of galaxies can drive mass rank order changes among a galaxy population. In the ideal picture of constant comoving number density evolution, all galaxies grow in mass with time at a rate that maintains mass rank ordering (i.e. no galaxies are allowed to 'pass' similar-mass systems in their mass rank order). This assumption breaks down for real galaxy populations that experience stochastically varying growth rates owing to, e.g. the spread in the star formation main sequence and/or stellar-mass growth via galaxy mergers. Real galaxy populations will change mass rank when growth rates of similar-mass galaxies are not homogeneous – driven by either scatter in the star formation main sequence or *ex situ* growth via stochastic galaxy mergers.

Unlike galaxy coagulation, the sense of mass rank order change owing to scattered growth rates can be either positive or negative. If a tracked galaxy population passes a large number of galaxies in mass rank order, fewer systems will remain with higher masses that results in a number density evolution rate of the same sign as galaxy mergers (positive,[3] according to equation 7). Conversely, if a tracked galaxy population is passed by more systems, the number of systems with larger masses will grow with time leading to a number density evolution rate of the opposite sign as galaxy mergers.

[3] A tracked galaxy population growing rapidly with time will move towards *smaller* number density values with time. However, the sign of the rate of change is positive since equation (7) expresses the number density evolution rate with redshift.

To handle these two cases, we identify the median mass of a tracked galaxy population as being $\langle M_i \rangle$ at time t and $\langle M_i' \rangle$ at some later time t'. We can then define

(1) *Galaxy mass rank demotion*: the number of galaxies with mass $M < \langle M_i \rangle$ at time t that have mass $M' > \langle M_i' \rangle$ at t'.
(2) *Galaxy mass rank promotion*: the number of galaxies with mass $M > \langle M_i \rangle$ at time t that have mass $M' < \langle M_i' \rangle$ at t'.

We adopt the terms 'galaxy demotion' and 'galaxy promotion' since the tracked galaxy population is being demoted or promoted in mass rank ordering relative to the rest of the galaxy population. We tabulate the galaxy promotion/demotion rates in the simulation using finite differencing with spacing of $\Delta z \approx 0.1$. In practice, we find all galaxies with $M > 10^8\,\mathrm{M}_\odot$ at redshift z and use the merger tree to identify all descendants at redshift $z - \Delta z$. The galaxy promotion rate has the same sign as the galaxy merger rate (positive), while the galaxy demotion rate has the opposite. Note that both *in situ* mass growth and growth via mergers are important to the scatter rate. However, galaxies that are not the main progenitor at time t of their descendent at time t' are considered to have been 'consumed' and therefore do not contribute to the scatter rate. The forward promotion and demotion mass rank scatter rates are shown in the left- and right-hand panels of Fig. 4 as solid lines for several mass bins.

The relative values of the galaxy promotion and demotion rates are set by the average mass growth rates of the tracked galaxy populations, the average mass growth rates of galaxies of neighbouring mass (number density) bins, as well as the relative abundance of galaxies in the tracked and neighbouring mass growth bins. At redshift $z = 0$, the galaxy promotion rate is larger than the galaxy demotion rate by a factor of $\sim$2 for $M \approx 3 \times 10^8\,\mathrm{M}_\odot$ systems but smaller by factor of $\sim$1.5 for $M \approx 10^{11}\,\mathrm{M}_\odot$ systems. The point of equality (i.e. no net change from scatter in the median number density of the tracked galaxy population) between the promotion and demotion rates is around $M = 3 \times 10^{10}\,\mathrm{M}_\odot$.

We explore the relative importance of the galaxy promotion, demotion and coagulation rates in more depth in the following subsection.

3.3 Relative importance of galaxy coagulation and scattered growth rates

To compare the relative importance of the galaxy coagulation and scattered growth rates, we adopt simple polynomial fits that are shown as dashed lines in Figs 3 and 4. The form of the fits takes the same fitting formula as the scatter and survival fraction fits discussed in Appendix A.

The top panel of Fig. 5 shows the net number density rate of change for galaxies for four mass bins (as indicated in the legend). Most galaxies experience positive number density evolution rates that are driven by the combined dominance of mergers and mass rank promotion – both of which drive galaxies to lower number densities as they move to lower redshift. However, there are periods at early times where mass rank demotion dominates for massive galaxies (i.e. lower mass galaxies are passing this bin), leading to a galaxy population moving to lower mass rank (higher number density) with time. These periods of time are indicated with dashed lines in the top panel of Fig. 5.

The net number density evolution rate can drop to zero and/or change signs depending on the relative importance of the galaxy coagulation, promotion and demotion rates. We therefore address the relative contribution of coagulation and scatter by considering

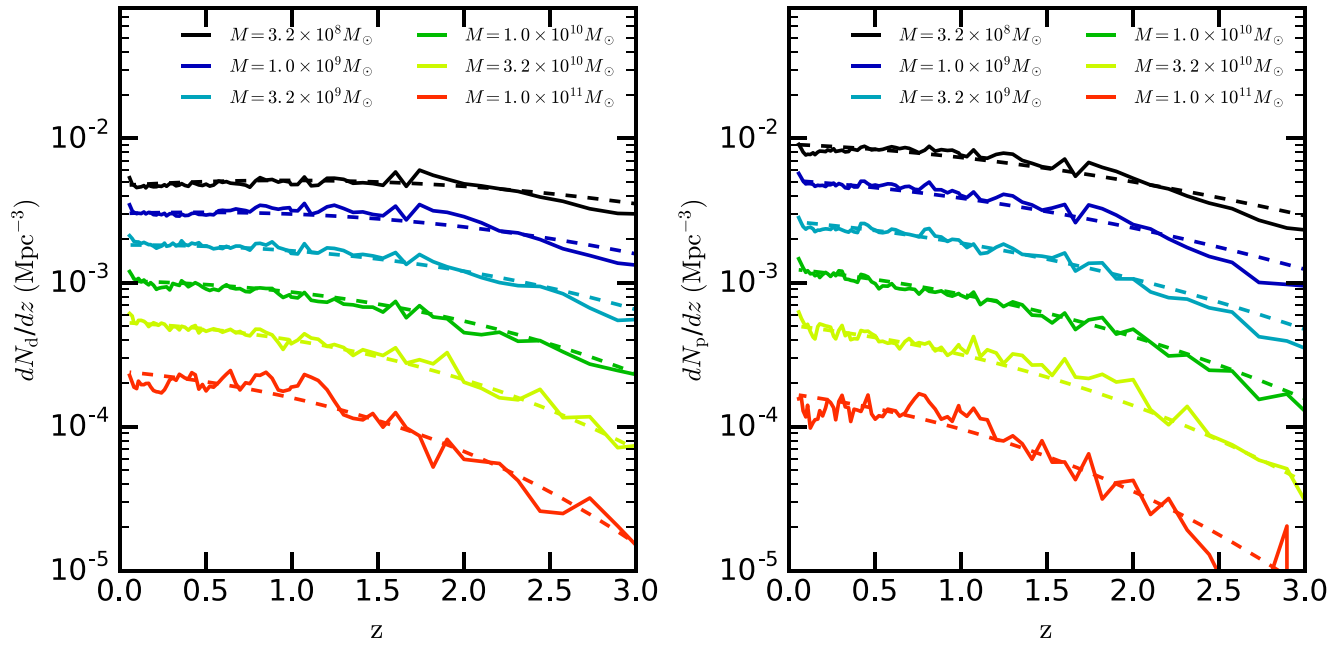

Figure 4. The promotion (left) and demotion (right) forward scatter rates. Solid lines indicate the promotion/demotion scatter rates as directly tabulated from the simulation; as in Fig. 3, the promotion/demotion rates are shown for the same fixed M_* at all redshifts, as indicated in the legends. Dashed lines indicate the best fits of the form used in Appendix A to assess the relative importance of galaxy coagulation and scattered growth rates.

the magnitude of the promotion and demotion rates (each shown individually) normalized by the galaxy coagulation rate as shown in the middle panel of Fig. 5. With the exception of the demotion rates of the two lowest mass bins, we find that the promotion- and demotion-scatter rates tend to be larger than the galaxy coagulation rate. The galaxy scatter rates are an order of magnitude larger than the coagulation rates for high-mass galaxies at high redshift. However, since the promotion and demotion rates partially cancel, it is possible for the coagulation rates to still dominate the net median mass rank order evolution.

The bottom panel of Fig. 5 shows the fractional contribution of the galaxy coagulation rate to the total number density evolution rate. Here, we have specifically defined the galaxy coagulation rate contribution to the total number density evolution rate as

$$f_c = \frac{\langle dN_c/dz \rangle}{|\langle dN_c/dz \rangle| + |\langle dN_s/dz \rangle|}. \quad (10)$$

Galaxy coagulation is the dominant contributor to the net number density evolution rate for galaxies with stellar masses $M_* < 10^{9.5}\,M_\odot$ out to redshift $z = 3$. For galaxies that are more massive than $10^{9.5}\,M_\odot$, the galaxy coagulation will be subdominant to scatter at early times, but eventually comes to dominate. However, for the most massive systems, scatter naturally dominates the mass rank order evolution budget owing to (i) the paucity of more massive systems that are able to drive large galaxy coagulation rates and (ii) the dominance of *ex situ* stellar-mass growth driven by stochastic merger events with lower mass galaxies for massive, quenched systems.

Massive galaxies in the Illustris simulation grow predominantly through galaxy mergers (Rodriguez-Gomez et al. 2015). Yet, scatter in their rank order dominates over galaxy coagulation when considering their mass rank order evolution budget. These two claims are consistent because the *ex situ* growth of massive galaxies is driven by mergers with less massive systems. Mergers with less massive systems – even if they are abundant and significantly impact the galaxy's mass evolution – do not influence the galaxy coagulation rate. Instead, the stochastic growth histories of massive galaxies associated with galaxy mergers act only to increase the influence of mass rank order scatter on the total mass rank order evolution budget.

A converse argument is true for low-mass galaxies that build up the majority of their mass through *in situ* star formation while their total mass rank order evolution budget is dominated by galaxy coagulation. Specifically, the galaxy mergers that lead to the large galaxy coagulation rate for low-mass galaxies do not involve the low-mass galaxies themselves. Instead, the galaxy coagulation rate is set by the merger rate between two larger galaxies and the galaxy scatter rate is set by the scatter in the *in situ* mass growth rates.

4 DISCUSSION

Forging progenitor/descendant links between observed galaxy populations at different redshifts using a constant cumulative number density is widely used and reasonably physically justified. It has been shown using abundance matching models, semi-analytic models and hydrodynamical simulations that one can recover, e.g. the mass evolution of Milky Way (MW) progenitors while only introducing errors of the order of ∼0.3–0.5 dex. However, the sense of this error is systematic and partially correctable. In Torrey et al. (2015) and this paper, we have argued for a modified method of forging progenitor/descendant links using relaxed assumptions about the evolution of galaxies in comoving number density space based on the results of numerical simulations. We have advocated that progenitor/descendant galaxy populations can be statistically linked based on their comoving number density. However, we also argue that this link must include both the median evolution of the galaxy

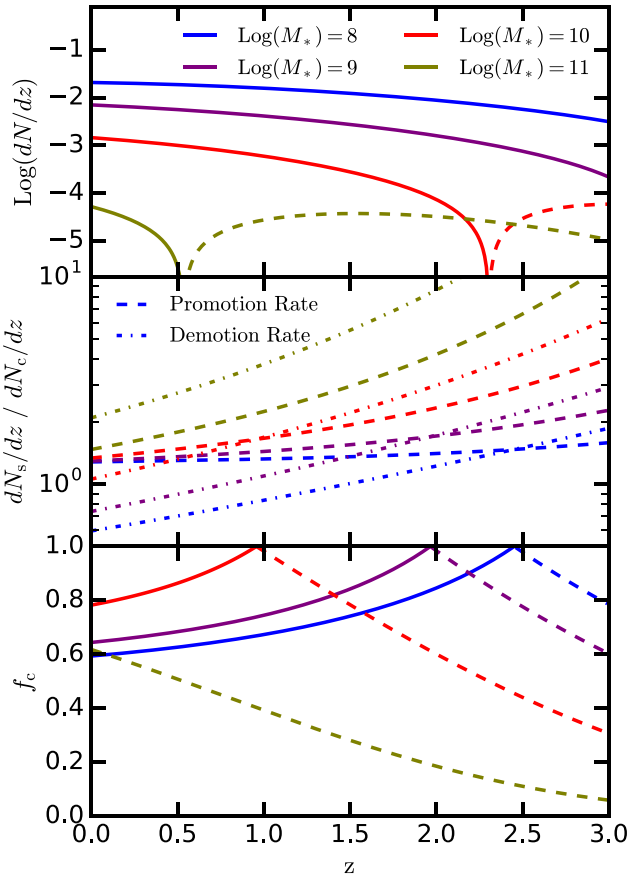

Figure 5. Top: the net median rate of change of four galaxy populations of fixed stellar mass with time. Dashed lines indicate negative values – where the galaxy demotion rate dominates over the combined promotion rate and coagulation rate. Negative values occur only for galaxies that are at the upper end of the redshift-dependent mass function. Middle: the magnitude of the galaxy promotion and demotion rate relative to the galaxy coagulation rate. Bottom: the fractional importance of galaxy coagulation to the net median number density evolution rate. Dashed lines indicate that the net scatter rate has the opposite sign to the galaxy coagulation rate. At late times (where the lines are solid), the net scatter rate and galaxy coagulation rate both have the same sign. The transition to constructive addition of the net scatter and coagulation rates occurs when the net scatter rate is zero owing to equal (but non-zero) contributions from the galaxy promotion and demotion rates.

population in comoving number density space (Torrey et al. 2015) and the significant intrinsic scatter (this paper).

In this paper, we analysed galaxy evolution in number density space using basic analytic arguments and empirical fits derived from the Illustris simulation. The empirical fits include a direct measure of the intrinsic scatter that makes identifying *direct* progenitor/descendant links impossible. Accounting for the intrinsic scatter makes it possible to draw conclusions about the statistical distribution of possible progenitor/descendant properties that would not be captured using a single comoving number density track. This is critical for describing the full range of formation histories that galaxies may evolve through. Recognizing the significant impact of scattered growth rates on the formation of galaxy populations is important for avoiding a secondary form of progenitor bias: considering only the median growth track to the exclusion of all others.

We have made an attempt in this paper to reconcile the asymmetry that is encountered when tracking galaxy populations forward and backward in time. It is well established that tracking galaxies forward and backward in time yields median number density evolution tracks that are not symmetric. We have shown that using the single DDF prescribed in Section 2.1, one can broadly recover forward evolution of galaxies in number density space and approximately recover the backward/PDF. This is important because it allows us to view the number density evolution of galaxies as a single coherent process, regardless of whether one is tracking galaxies forward or backward in time.

4.1 Dependence on baryon physics

In this section, we consider possible uncertainty in the galaxy number density evolution tracks derived from Illustris owing to uncertainty in the baryon physics modelling used in the Illustris simulation. In particular, we aim to disentangle the impact of (i) discrepancies between the observed and Illustris galaxy cumulative stellar-mass functions from (ii) uncertainty in the derived Illustris comoving number density evolution tracks that are the main focus of this paper.

4.1.1 Comparison of prescribed number density evolution tracks from Illustris, Millennium and Bolshoi

Uncertainty in the prescribed number density evolution tracks cannot be assessed through comparison with observational data because information about how galaxies evolve in comoving number density space is not available in observations. Instead, the best method for understanding the uncertainty associated with the prescribed number density evolution tracks in this paper is to consider how prescriptions from Illustris differ from those obtained through alternative simulations. To achieve this, we compare the number density evolution tracks of haloes obtained from the Millennium-II simulation (Boylan-Kolchin et al. 2009) and Bolshoi simulation (Klypin, Trujillo-Gomez & Primack 2011) against those prescribed in this paper from Illustris. Since Millennium-II and Bolshoi are both dark-matter-only simulations, the number density evolution is driven strictly by the stochastic assembly of dark matter haloes. Comparing the evolution of galaxies in comoving cumulative number density space from Millennium and Bolshoi against Illustris provides insight into the extent to which galaxy number density evolution is driven by the specific baryon physics included in Illustris versus the dark matter halo assembly.

A comparison of the galaxy cumulative number density evolution is shown in Figs 6 and 7 in the forward and backward tracked directions, respectively. The number density evolution for Bolshoi was obtained using the best fits provided in Behroozi et al. (2013). Behroozi et al. (2013) assigned rank order/number density to haloes at any snapshot based on their peak historical halo mass. Peak historical halo mass was used because dark matter is more easily stripped from haloes when compared with stars (Reddick et al. 2013). We have applied the same procedure described in Behroozi et al. (2013) to Millennium-II using the halo merger trees (Springel et al. 2005) that we obtained from the Millennium simulation public data base (Lemson & Springel 2006). Hence, the number density evolution found in Millennium and Bolshoi is driven strictly by halo assembly using a simple abundance matching prescription (i.e. no semi-analytic model is used in our analysis).

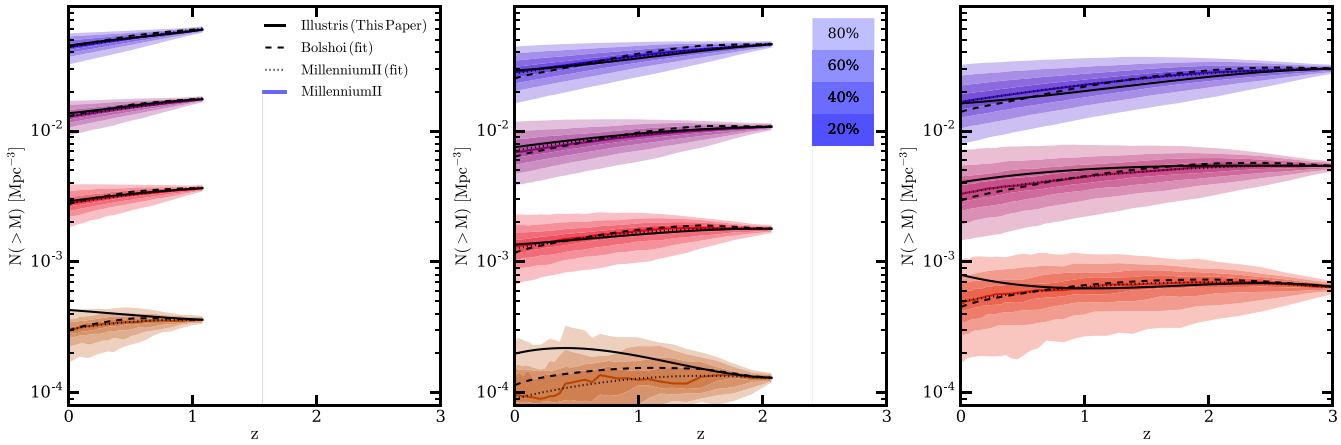

Figure 6. A comparison of the number density evolution of galaxy populations is shown tracked in time from redshifts $z = 1$, 2 and 3 to redshift $z = 0$, from left- to right-hand panel, respectively. The coloured shaded bands indicate the number density evolution as directly determined from the Millennium-II simulation – with the shade of the colour indicating the enclosed galaxy fraction as indicated in the legend. The black solid lines denote the best fits to the median N provided in equation (A1) from the Illustris simulation. The black dashed lines denote the best fits to the median N provided in equation (A1) from the Bolshoi simulation as obtained through Behroozi et al. (2013).

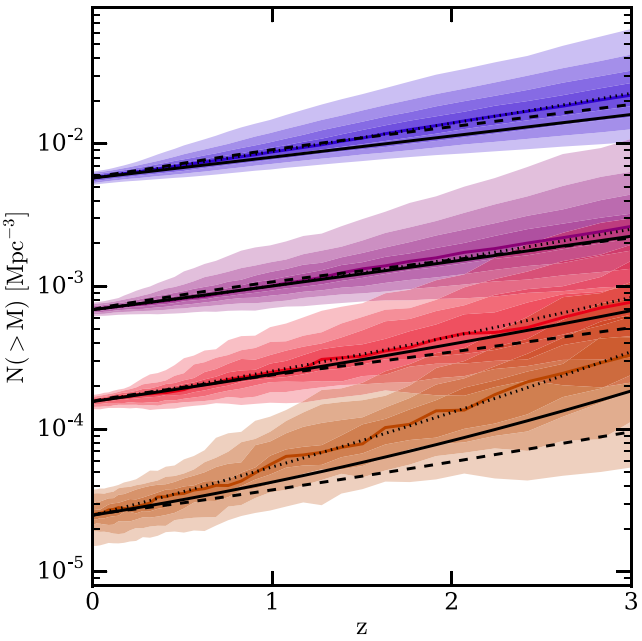

Figure 7. Same as Fig. 6 but for galaxy populations tracked backward in time. Coloured bands indicated Millennium-II number density evolution tracks, solid lines indicate Illustris and dashed lines indicate Bolshoi. Agreement is reasonably good for the three low-mass bins but biased for the highest mass bin where the impact of quenching from black holes impacts Illustris differently than the halo growth rates in Millennium-II and Bolshoi and where small number statistics add significant variability to the number density evolution tracks.

The agreement for number density evolution for forward tracked galaxies in the mass range $10^8 < M_* < 10^{10}$ from $0 < z < 2$ (i.e. the top three bands in the left-hand and centre panels of Fig. 6) is very good between all three simulations (i.e. within a few per cent). This is notable since we are comparing number density evolution tracks derived from dark matter haloes (in Millennium and Bolshoi) against number density evolution tracks of galaxies derived from stellar masses (in Illustris). In this regime, we have strong confidence in the prescribed number density evolution fits because they are predominately driven by halo assembly – which is reasonably well understood in Λ cold dark matter. However, the agreement deteriorates elsewhere. Moving to higher redshift (i.e. galaxies selected at $z = 3$) results from the three simulations in the lowest two mass bins qualitatively agree, but with offsets of the order of 10–15 per cent between Illustris and the dark-matter-only simulations. More notably, the highest mass bins in each panel of Fig. 6 show qualitatively different behaviour when comparing Illustris against the dark-matter-only models. This distinct behaviour is driven by differences in the handling of stellar versus dark matter mass growth in the simulations. Whereas stellar-mass growth is strongly suppressed in Illustris for massive galaxies owing to strong stellar feedback, no similar feedback mechanism operates to slow halo-mass growth in the dark-matter-only simulations. Similar conclusions are drawn when galaxy populations are tracked backward in time, as demonstrated in Fig. 7. Specifically, the low galaxy mass bins agree reasonably well between the three simulations, while the most massive tracked bin shows distinct behaviour. This high-mass bin also suffers from low number statics, which could be improved with larger simulated box sizes.

Broadly speaking, offsets been Illustris and the number density evolution tracks of Millennium and Bolshoi arise where a decoupling of the stellar-mass growth from the halo-mass growth occurs. In Illustris, high-mass galaxies/haloes experience strong black hole feedback that regulates stellar-mass growth, which partially decouples stellar-mass growth from halo-mass growth. Stellar-mass halo-mass growth decoupling can also occur for lower mass galaxies that undergo environmental effects (e.g. strangulation/stripping) that are not similarly captured in the dark-matter-only models. Diagnosing the exact cause of the disagreement between Illustris and the dark-matter-only simulations is beyond the scope of this paper. Instead, we simply make the point the prescribed number density evolution fits should be treated as uncertain and beholden to the specific baryon physics model employed in Illustris in regions where the fits differ from those obtained with Millennium and Bolshoi. While it is unlikely that dark-matter-only simulations fully

capture the number density evolution of massive galaxies where the stellar-mass halo-mass relation is known to change slope, it is similarly unlikely that Illustris captures number density evolution in this regime without error. Number density evolution fits applied in this regime should therefore be treated with caution and consider the variation shown in Fig. 6 to be intrinsic to current modelling standards.

4.1.2 Comparison of Illustris cumulative stellar-mass functions with observations

A second source of uncertainty in the analysis presented in this paper is the imperfect match between the Illustris and observed cumulative stellar-mass functions. We would like to consider how uncertainties in the simulative cumulative stellar-mass function may propagate into the analysis presented in this paper. Fig. 8 shows a comparison of the cumulative stellar-mass functions derived from Illustris (solid bands) against the cumulative stellar-mass functions from Ilbert et al. (2013, top), Muzzin et al. (2013, middle) and Tomczak et al. (2014, bottom). Dashed lines indicate extrapolations of the best-fitting stellar-mass functions quoted in each paper beyond the completeness limits of the data. Some broad conclusions about tensions between the observed mass functions and Illustris mass functions can be drawn with consensus agreement based on these three data sets. Specifically, both the high- and low-mass ends of the low-redshift (i.e. $z < 0.5$) cumulative stellar-mass functions are overpopulated in Illustris compared to the observed cumulative mass functions (CMFs). These tensions represent inefficiencies with the Illustris feedback model, which has been addressed in much more detail in Torrey et al. (2014) and Genel et al. (2014). More detailed conclusions about the level of agreement between the observed and Illustris mass functions depend on the selected data set. For example, while the non-extrapolated Ilbert et al. (2013) and Muzzin et al. (2013) mass functions broadly agree with Illustris over the mass range $3 \times 10^9\,\mathrm{M}_\odot < M_* < 2 \times 10^{11}\,\mathrm{M}_\odot$, the Tomczak et al. (2014) mass functions are systematically biased towards lower number densities when compared with Illustris over this same range.

Tension/errors between the Illustris CMFs and the observed CMFs do not necessarily impact the accuracy of the number density evolution fits (as discussed in the previous subsection) but will propagate into errors in the inferred stellar-mass evolution tracks. This point is demonstrated in Fig. 9, where the solid lines indicate the inferred median stellar-mass evolution for redshift $z = 0$ MW-mass galaxies using the number density evolution tracks prescribed in this paper. Different median mass evolution tracks are implied for the Ilbert et al. (2013), Muzzin et al. (2013), Tomczak et al. (2014) and Illustris mass functions. The inferred median mass evolution tracks are plotted only when they are above the mass completeness limits of the survey. We additionally show the best-fitting MW-mass evolution track from van Dokkum et al. (2013) based on the data from Marchesini et al. (2009) that assumed constant comoving number density evolution.

Mass evolution tracks derived from linking galaxies in comoving number density space directly inherit any uncertainty and/or errors in the assumed CMFs. Adopting the Illustris cumulative stellar-mass function as a way to infer the mass evolution of any particular galaxy class should be done only with caution given the uncertainties in the underlying model. Fig. 9 demonstrates how the inferred median mass evolution tracks vary depending on the employed mass function. For example, the inferred median MW

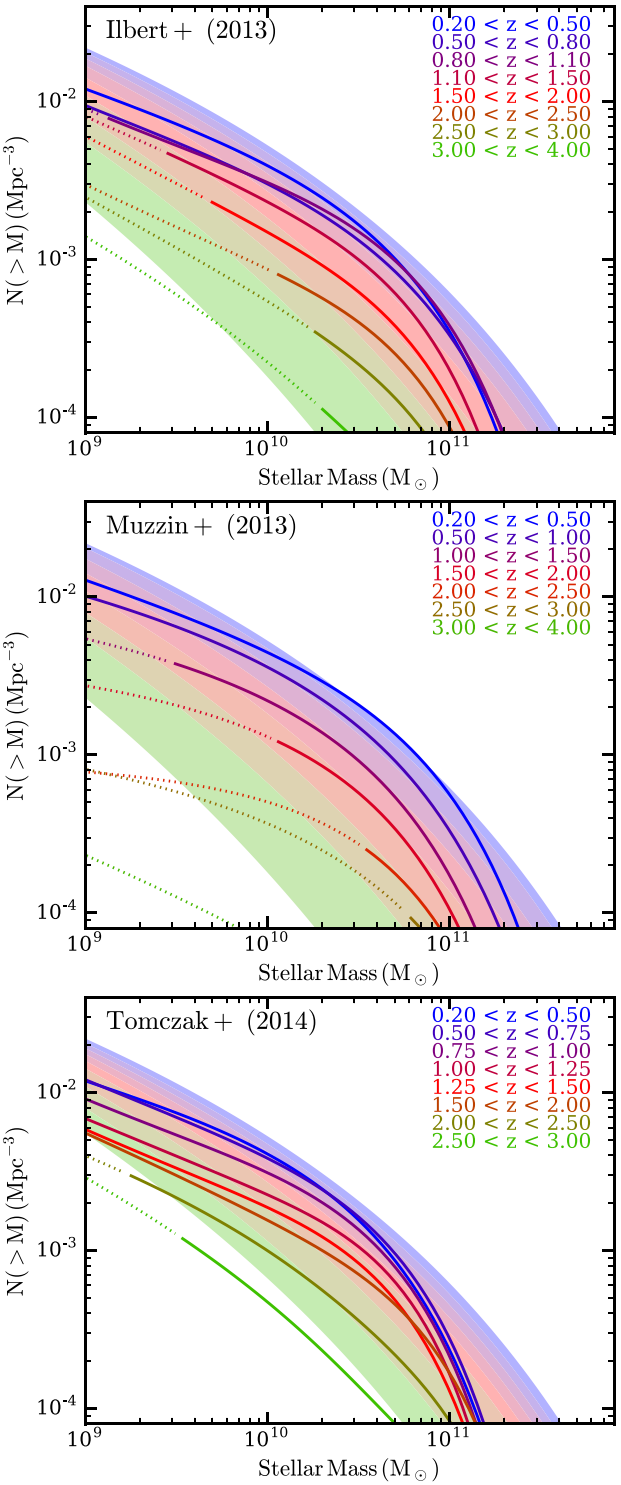

Figure 8. CMFs from Ilbert et al. (2013, top), Muzzin et al. (2013, middle) and Tomczak et al. (2014, bottom) are shown for comparison with Illustris (solid bands in each panel). The Illustris low-mass CMF slope is steeper than what is found in any of these data sets. Illustris broadly matches the knee of the Ilbert et al. (2013) and Muzzin et al. (2013) mass functions but is biased towards lower number densities when compared against Tomczak et al. (2014). We note that disagreements between the observed and Illustris mass functions propagate into uncertainties in the inferred mass evolution of galaxies, as demonstrated in Fig. 9, but do *not necessarily* imply uncertainties in the prescribed number density evolution tracks presented in this paper as discussed in Section 4.1.1.

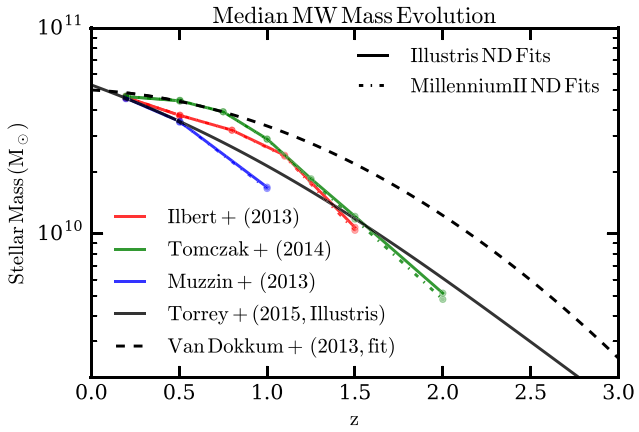

Figure 9. The inferred median MW progenitor mass evolution. Solid coloured lines indicate the mass tracks obtained using the Illustris number density evolution fits combined with the Ilbert et al. (2013), Tomczak et al. (2014) and Muzzin et al. (2013) cumulative stellar mass functions. Dot–dashed coloured lines indicate the mass tracks obtained using the Millennium-II number density evolution fits combined with the same cumulative stellar mass functions. While the difference between inferred mass tracks using different mass functions is large, the difference between the Millennium and Illustris number density evolution tracks is minimal. The van Dokkum et al. (2013) line, which has a significant offset from the other results, uses a constant comoving number density assumption with the Marchesini et al. (2009) stellar mass functions.

progenitor mass varies by nearly a factor of ∼1.5–2 between the Tomczak et al. (2014) and Muzzin et al. (2013) mass evolution tracks at redshift $z = 1$, with the Illustris results falling somewhere between. Uncertainty in the cumulative stellar-mass functions – Illustris or observed – intrinsically limits the accuracy with which conclusions about the (median) mass growth rates of galaxies can be assessed.

Finally, we consider how the MW median progenitor mass would change if we had instead adopted the number density evolution tracks prescribed by Millennium. Fig. 9 includes a series of dotted coloured lines that indicate the inferred median MW progenitor mass as determined from the Millennium number density evolution best fits. Comparing any of the dotted and solid lines of the same colour, we find that the Millennium and Illustris inferred MW median progenitor mass evolution tracks are very similar. The uncertainty in the inferred median MW progenitor mass evolution is dominated by uncertainty in the cumulative stellar-mass function, not uncertainty in the number density evolution fits. This conclusion holds for low-mass galaxies (i.e. $M_* \lesssim 10^{10}\,\mathrm{M}_\odot$) but breaks down for massive systems where the consistency between Illustris and Millennium breaks down.

4.2 Application to observational data sets

The evolution tracks prescribed in this paper can be used to link selected galaxy populations to progenitor or descendant galaxy *populations*. In contrast to linking at a constant comoving number density, or at a non-constant comoving number density (Torrey et al. 2015), the PDFs and DDFs will identify a wider range of possible galaxy progenitors/descendants. The spread in the progenitor/descendant population properties can be set based on results of numerical simulations.

A detailed exploration of how our prescriptions can be applied to observational data is left to a companion paper (Wellons & Torrey 2017). There, we describe in detail and validate a straightforward method to translate the DDFs tabulated in this paper into mass, size and star formation rate evolution tracks based directly on observational data.

However, we note that the overall predictive power of linking galaxies via the comoving number density method as discussed in this paper may be limited when stellar mass is the only variable considered. Additional information – such as environment, galaxy colour, star formation rate, bulge fraction, black hole mass, etc. – may improve the intrinsic ability of this method to link progenitor and descendant galaxy populations.

5 CONCLUSIONS

In this paper, we have critically examined the nature of galaxy evolution in comoving number density space. We have provided a detailed description of how the median number density and standard deviation for a tracked galaxy population evolves with time. The primary conclusions of this paper are as follows.

(1) We defined the DDF and the PDF to capture the median number density evolution, the spread among the tracked galaxy population, and the survival fraction of tracked galaxies. Practical fits to the DDF and the PDF were provided based on fits from the Illustris simulation that can be applied to observational data. Publicly available PYTHON scripts to evaluate the fitting functions have been provided.[4]

(2) The median number density evolution described in the DDF and the PDF are not symmetric. This asymmetry is driven by the convolution of the DDF with the relative abundance of progenitor galaxies that is required to recover the PDF. A quantitative description of this is given in equation (5) that is demonstrated in Fig. 2.

(3) The total DDF evolution rate can be broken down into a contribution from galaxy coagulation and galaxy scattered growth rates. We quantified the relative contribution of galaxy coagulation and scattered growth rates and found that galaxy coagulation is the dominant driver of the number density evolution rate for a range of mass bins and redshift ranges (described in Section 3.3 and Fig. 5).

(4) The evolution tracks for galaxy populations in number density space defined in this paper for redshift $z \leq 3$ in the mass ranges $10^8\,\mathrm{M}_\odot < M_*$ can be used to analyse observational data sets. A forthcoming paper (Wellons & Torrey 2017) will practically demonstrate how the diversity of evolution tracks in number density space impact conclusions about the mass, star formation rate and velocity dispersion evolution of galaxy populations.

ACKNOWLEDGEMENTS

We thank the referee, Bart Clauwens, for the many thoughtful comments that have strengthened this work. PT is supported through Hubble Fellowship grant #HST-HF2-51384.001-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555. Support for PFH was provided by an Alfred P. Sloan Research Fellowship, NASA ATP Grant

[4] https://github.com/ptorrey/torrey_cmf

NNX14AH35G, NSF Collaborative Research Grant #1411920 and CAREER Grant #1455342. MV acknowledges support through an MIT RSC award. The Illustris simulation was run on the CURIE supercomputer at CEA/France as part of PRACE project RA0844, and the SuperMUC computer at the Leibniz Computing Centre, Germany, as part of project pr85je. Further simulations were run on the Harvard Odyssey and CfA/ITC clusters, the Ranger and Stampede supercomputers at the Texas Advanced Computing Center through XSEDE and the Kraken supercomputer at Oak Ridge National Laboratory through XSEDE. The analysis reported in this paper was performed on the Caltech compute cluster 'Zwicky' (NSF MRI award #PHY-0960291), the joint partition of the MIT-Harvard computing cluster 'Odyssey' supported by MKI and FAS, and allocation TG-AST150059 granted by the Extreme Science and Engineering Discovery Environment (XSEDE) supported by the NSF.

REFERENCES

Behroozi P. S., Marchesini D., Wechsler R. H., Muzzin A., Papovich C., Stefanon M., 2013, ApJ, 777, L10
Bezanson R. et al., 2011, ApJ, 737, L31
Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, MNRAS, 398, 1150
Brammer G. B. et al., 2011, ApJ, 739, 24
Clauwens B., Franx M., Schaye J., 2016, MNRAS, 463, L1
Conselice C. J., Mortlock A., Bluck A. F. L., Grützbauch R., Duncan K., 2013, MNRAS, 430, 1051
Fakhouri O., Ma C.-P., 2008, MNRAS, 386, 577
Genel S. et al., 2014, MNRAS, 445, 175
Ilbert O. et al., 2013, A&A, 556, A55
Jaacks J., Finkelstein S. L., Nagamine K., 2016, ApJ, 817, 174
Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ, 740, 102
Leja J., van Dokkum P., Franx M., 2013, ApJ, 766, 33
Lemson G., Springel V., 2006, in Gabriel C., Arviset C., Ponz D., Enrique S., eds, ASP Conf. Ser. Vol. 351, Astronomical Data Analysis Software and Systems XV. Astron. Soc. Pac., San Francisco, p. 212
Lidman C. et al., 2012, MNRAS, 427, 550
Lin Y.-T., Brodwin M., Gonzalez A. H., Bode P., Eisenhardt P. R. M., Stanford S. A., Vikhlinin A., 2013, ApJ, 771, 61
Marchesini D., van Dokkum P. G., Förster Schreiber N. M., Franx M., Labbé I., Wuyts S., 2009, ApJ, 701, 1765
Marchesini D. et al., 2014, ApJ, 794, 65
Mundy C. J., Conselice C. J., Ownsworth J. R., 2015, MNRAS, 450, 3696
Muzzin A. et al., 2013, ApJ, 777, 18
Ownsworth J. R., Conselice C. J., Mortlock A., Hartley W. G., Almaini O., Duncan K., Mundy C. J., 2014, MNRAS, 445, 2198
Papovich C., Finkelstein S. L., Ferguson H. C., Lotz J. M., Giavalisco M., 2011, MNRAS, 412, 1123
Papovich C. et al., 2015, ApJ, 803, 26
Patel S. G. et al., 2013, ApJ, 766, 15
Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, ApJ, 771, 30
Rodriguez-Gomez V. et al., 2015, MNRAS, 449, 49
Saglia R. P. et al., 2010, A&A, 524, A6
Salmon B. et al., 2015, ApJ, 799, 183
Shankar F. et al., 2015, ApJ, 802, 73
Springel V. et al., 2005, Nature, 435, 629
Tomczak A. R. et al., 2014, ApJ, 783, 85
Torrey P., Vogelsberger M., Genel S., Sijacki D., Springel V., Hernquist L., 2014, MNRAS, 438, 1985
Torrey P. et al., 2015, MNRAS, 454, 2770
van Dokkum P. G., Franx M., 1996, MNRAS, 281, 985
van Dokkum P. G., Franx M., 2001, ApJ, 553, 90
van Dokkum P. G. et al., 2010, ApJ, 709, 1018
van Dokkum P. G. et al., 2013, ApJ, 771, L35
Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, MNRAS, 436, 3031
Vogelsberger M. et al., 2014a, MNRAS, 444, 1518
Vogelsberger M. et al., 2014b, Nature, 509, 177
Wake D. A. et al., 2006, MNRAS, 372, 537
Wellons S., Torrey P., 2017, MNRAS, preprint (arXiv:e-prints)

APPENDIX A: MEDIAN NUMBER DENSITY, SCATTER AND SURVIVAL FRACTION FITTING FUNCTIONS

In this appendix, we provide fits to the median number density, 1σ scatter in the number density, and survival fraction of a tracked galaxy population based on the Illustris simulation. The fits are not intended to carry physical meaning but instead to serve as tools that can be used to link galaxy populations based on observational data sets. A single fit for the forward evolution of galaxy populations is required for the evaluation of equation (5). The current fits have a large number of parameters because reduced parameters led to noticeable loss of fit accuracy.

In addition to publishing the fits here, PYTHON routines to evaluate these fits are provided through https://github.com/ptorrey/torrey_cmf.

A1 Forward median number density fit

The forward-tracked median number density evolution is well fitted with

$$\begin{aligned}\langle \mathcal{N}(z)\rangle = \mathcal{N}_0 &+\Delta z\left(A_0 + A_1\mathcal{N}_0 + A_2\mathcal{N}_0^2 + A_3\mathcal{N}_0^3 + A_4\mathcal{N}_0^4\right)\\ &+\Delta z^2\left(B_0 + B_1\mathcal{N}_0 + B_2\mathcal{N}_0^2 + B_3\mathcal{N}_0^3 + B_4\mathcal{N}_0^4\right)\\ &+\Delta z^3\left(C_0 + C_1\mathcal{N}_0 + C_2\mathcal{N}_0^2 + C_3\mathcal{N}_0^3 + C_4\mathcal{N}_0^4\right), \quad \text{(A1)}\end{aligned}$$

where $\mathcal{N}_0 = \log(N(z_0))$ and $\Delta z = |z_0 - z|$. The coefficients A_i, B_i and C_i are each functions of the initial selection redshift for the galaxy population according to $\alpha_i = \alpha_{i,0} + z_0\alpha_{i,1} + z_0^2\alpha_{i,2}$ (e.g. $A_1 = A_{1,0} + z_0A_{1,1} + z_0^2A_{1,2}$). This results in a fitting function that accounts for the dependence of initial selection redshift, initial selection number density and evolution time. The fit has 45 free variables that are determined via linear regression based on the number density evolution of all galaxies with stellar masses greater than $10^8\,\mathrm{M}_\odot$ tracked forward in time from any initial selection redshift below $z \leq 3$. This high number of coefficients is a result of making a single fit to simultaneous variations in $\langle N_f\rangle$, N_0, z_0 and Δz. The best-fitting coefficients are given in Table A1, the resulting fits are shown as dashed black lines in Fig. 1 and easily applied PYTHON functions to evaluate these fits are publicly available.[5] Equation (A1) describes the median comoving number density evolution for galaxy population selected at any redshift $z \leq 3$ in the mass ranges $10^8\,\mathrm{M}_\odot < M_*$ tracked forward in time. This single fit accurately captures the number density evolution for a wide range of galaxies owing to the high order of the polynomial expansion.

[5] https://github.com/ptorrey/torrey_cmf

Table A1. Best-fitting parameters to the forward median number density evolution (equation A1), scatter number density evolution (equation A2) and galaxy survival fraction (equation A3). The resulting best fits are demonstrated in Fig. 1.

	$j=0$	$j=1$	$j=2$
$A_{0,j}$	−6.616 92	4.772 10	−0.620 78
$A_{1,j}$	−12.553 86	10.062 38	−1.593 52
$A_{2,j}$	−8.490 63	7.305 74	−1.297 25
$A_{3,j}$	−2.393 11	2.170 88	−0.415 97
$A_{4,j}$	−0.237 25	0.224 54	−0.045 41
$B_{0,j}$	20.215 98	−19.237 83	3.786 53
$B_{1,j}$	35.822 74	−34.933 41	7.113 85
$B_{2,j}$	22.317 93	−22.302 32	4.677 40
$B_{3,j}$	5.845 23	−5.964 09	1.282 03
$B_{4,j}$	0.546 39	−0.566 17	0.124 11
$C_{0,j}$	−12.516 57	11.750 15	−2.437 51
$C_{1,j}$	−21.967 28	20.806 09	−4.367 87
$C_{2,j}$	−13.612 42	13.014 45	−2.763 50
$C_{3,j}$	−3.562 42	3.431 87	−0.735 92
$C_{4,j}$	−0.334 57	0.323 61	−0.069 90
$D_{0,j}$	0.307 54	−0.357 06	0.078 12
$D_{1,j}$	0.116 59	−0.363 10	0.087 81
$D_{2,j}$	0.054 32	−0.085 68	0.019 39
$E_{0,j}$	0.341 05	−0.265 66	0.058 67
$E_{1,j}$	0.404 05	−0.278 61	0.057 34
$E_{2,j}$	0.063 19	−0.046 06	0.010 22
$F_{0,j}$	−0.576 10	0.253 95	−0.069 09
$F_{1,j}$	−0.323 07	0.305 92	−0.084 16
$F_{2,j}$	−0.058 80	0.071 59	−0.019 36
$G_{0,j}$	0.344 38	−0.217 27	0.041 29
$G_{1,j}$	0.298 59	−0.211 86	0.042 28
$G_{2,j}$	0.038 89	−0.031 77	0.006 90

A2 Forward number density scatter fit

We quantify the scatter evolution in the DDF using the standard deviation of the number density distribution for the tracked galaxy population in log space. The fitting function

$$\sigma(z) = \sigma_0 + \Delta z \left(D_0 + D_1\mathcal{N}_0 + D_2\mathcal{N}_0^2\right) + \Delta z^2 \left(E_0 + E_1\mathcal{N}_0 + E_2\mathcal{N}_0^2\right) \quad \text{(A2)}$$

describes the forward-tracked lognormal standard deviation well. The coefficients D_i and E_i are functions of the selection redshift of the galaxy population, $\alpha_i = \alpha_{i,0} + z_0\alpha_{i,1} + z_0^2\alpha_{i,2}$, and are set via a linear regression from the same Illustris galaxies used in Appendix A1. The best-fitting values for the D_i and E_i coefficients are given in Table A1.

A3 Forward survival fraction fit

We quantify the survival fraction of tracked galaxy populations by calculating the fraction of systems that are consumed by a larger galaxy. The fitting function

$$f_s(z) = 1.0 + \Delta z \left(F_0 + F_1\mathcal{N}_0 + F_2\mathcal{N}_0^2\right) + \Delta z^2 \left(G_0 + G_1\mathcal{N}_0 + G_2\mathcal{N}_0^2\right) \quad \text{(A3)}$$

describes the forward tracked survival fraction well. The best-fitting values for the F_i and G_i coefficients are given in Table A1, with the results shown in Fig. A1. The fits broadly capture the survival fraction evolution over the mass range $10^8\,\mathrm{M}_\odot < M_{*,0} < 10^{11}\,\mathrm{M}_\odot$, with errors of the order of $\lesssim$10 per cent when compared against the merger-tree-defined survival fraction evolution from Illustris.

A4 Backward median number density fit

We showed in Section 2.2 that the PDF can be broadly recovered using the DDF and forward scatter fits. However, we also argued that the accuracy of this recovery procedure is limited by deviations of the DDF from a strictly lognormal distribution and by progenitors drawn from outside the valid fitting regions (as set by the resolution and box size of Illustris). For consistency with the fits presented in the previous subsection, we also fit to the median number density for the PDF using

$$\langle\mathcal{N}(z)\rangle' = \mathcal{N}_0 + \Delta z \left(A'_0 + A'_1\mathcal{N}_0 + A'_2\mathcal{N}_0^2\right) + \Delta z^2 \left(B'_0 + B'_1\mathcal{N}_0 + B'_2\mathcal{N}_0^2\right), \quad \text{(A4)}$$

where as before $\mathcal{N}_0 = \log(N(z_0))$ and $\Delta z = |z_0 - z|$. Primes in equation (A4) are used to denote fits to the progenitor distribution. We assume that galaxy populations are selected at redshift $z = 0$ and therefore we can directly specify with only six parameters. Best-fitting values for equation (A4) are given in Table A2 and

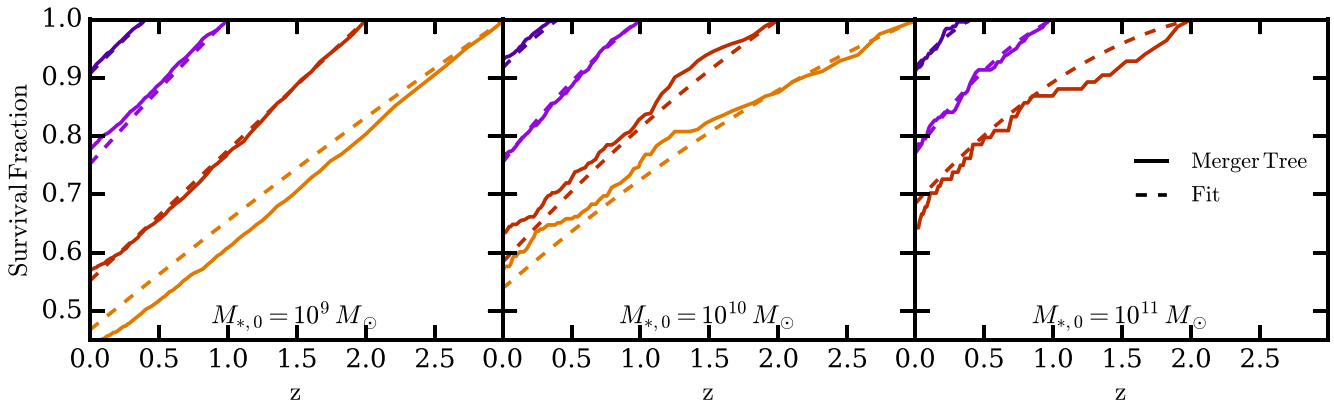

Figure A1. The survival fraction for galaxies selected with initial mass $M_{*,0} = \{10^9, 10^{10}, 10^{11}\}\,\mathrm{M}_\odot$ from left to right, respectively. Each panel shows galaxy populations selected at redshifts $z_0 = 0.5, 1, 2, 3$ in bins of width $\Delta M = 0.15$ dex tracked forward in time. Solid lines indicate results from the merger trees; dashed lines indicate the best-fitting results presented in equation (A3) using coefficients presented in Table A1. No line is shown for $M_{*,0} = 10^{11}\,\mathrm{M}_\odot$ with $z_0 = 3$ because we require at least 50 galaxies initially be in each tracked bin (and only 20 are found in that bin).

Table A2. Best-fitting parameters to the backward median number density evolution. The resulting best fits are demonstrated in Fig. 2.

	$j=0$	$j=1$	$j=2$
A'_j	0.125 773	−0.001 150	0.003 375
B'_j	0.045 737	0.035 451	0.006 921
D'_i	0.110 200	0.025 691	0.024 996
E'_i	−0.029 972	−0.021 199	−0.007 629

the resulting best-fitting tracks are demonstrated with black dashed lines in Fig. 2.

A5 Backward number density scatter fit

We fit to the 1σ standard deviation in the number density evolution of backward tracked galaxy populations using

$$\sigma(z) = \sigma_0 + \Delta z \left(D'_0 + D'_1\mathcal{N}_0 + D'_2\mathcal{N}_0^2\right) + \Delta z^2 \left(E'_0 + E'_1\mathcal{N}_0 + E'_2\mathcal{N}_0^2\right). \quad \text{(A5)}$$

Best-fitting values for equation (A4) are given in Table A2 and the resulting best-fitting tracks are demonstrated with black dashed lines in Fig. 2.

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