

# Evolution of galaxy stellar masses and star formation rates in the EAGLE simulations

M. Furlong,<sup>1★</sup> R. G. Bower,<sup>1</sup> T. Theuns,<sup>1,2</sup> J. Schaye,<sup>3</sup> R. A. Crain,<sup>3</sup> M. Schaller,<sup>1</sup> C. Dalla Vecchia,<sup>4,5</sup> C. S. Frenk,<sup>1</sup> I. G. McCarthy,<sup>6</sup> J. Helly,<sup>1</sup> A. Jenkins<sup>1</sup> and Y. M. Rosas-Guevara<sup>7,8</sup>

<sup>1</sup>*Institute for Computational Cosmology, Durham University, South Road, Durham DH1 3LE, UK*

<sup>2</sup>*Department of Physics, University of Antwerp, Groenenborgerlaan 171, B-2020 Antwerp, Belgium*

<sup>3</sup>*Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands*

<sup>4</sup>*Instituto de Astrofísica de Canarias, C/ Vía Láctea s/n, E-38205 La Laguna, Tenerife, Spain*

<sup>5</sup>*Departamento de Astrofísica, Universidad de La Laguna, Av. del Astrofísicaco Francisco Sánchez s/n, E-38206 La Laguna, Tenerife, Spain*

<sup>6</sup>*Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK*

<sup>7</sup>*Université de Lyon, Lyon F-69003, France*

<sup>8</sup>*CNRS, UMR 5574, Centre de Recherche Astrophysique de Lyon, Ecole Normale Supérieure de Lyon, Lyon F-69007, France*

Accepted 2015 April 15. Received 2015 March 26; in original form 2014 October 1

## ABSTRACT

We investigate the evolution of galaxy masses and star formation rates in the Evolution and Assembly of Galaxies and their Environment (EAGLE) simulations. These comprise a suite of hydrodynamical simulations in a  $\Lambda$  cold dark matter cosmogony with subgrid models for radiative cooling, star formation, stellar mass-loss and feedback from stars and accreting black holes. The subgrid feedback was calibrated to reproduce the observed present-day galaxy stellar mass function and galaxy sizes. Here, we demonstrate that the simulations reproduce the observed growth of the stellar mass density to within 20 per cent. The simulations also track the observed evolution of the galaxy stellar mass function out to redshift  $z = 7$ , with differences comparable to the plausible uncertainties in the interpretation of the data. Just as with observed galaxies, the specific star formation rates of simulated galaxies are bimodal, with distinct star forming and passive sequences. The specific star formation rates of star-forming galaxies are typically 0.2 to 0.5 dex lower than observed, but the evolution of the rates track the observations closely. The unprecedented level of agreement between simulation and data across cosmic time makes EAGLE a powerful resource to understand the physical processes that govern galaxy formation.

**Key words:** galaxies: abundances – galaxies: evolution – galaxies: formation – galaxies: high-redshift – galaxies: star formation.

## 1 INTRODUCTION

Although the basic model for how galaxies form within the framework of a cold dark matter cosmogony has been established for many years (e.g. White & Rees 1978; White & Frenk 1991), many crucial aspects are still poorly understood. For example, what physical processes determine galaxy stellar masses and galaxy sizes? How do these properties evolve throughout cosmic history? How do stars and AGN regulate the evolution of galaxy properties? Numerical simulations and theoretical models are a valuable tool for

exploring these questions, but the huge dynamic range involved, and the complexity of the plausible underlying physics, limits the ab initio predictive power of such calculations (e.g. Schaye et al. 2010; Scannapieco et al. 2012).

We recently presented the EAGLE simulation project (Schaye et al. 2015, hereafter S15), a suite of cosmological hydrodynamical simulations in which subgrid models parametrize our inability to faithfully compute the physics of galaxy formation below the resolution of the calculations. Calibrating the parameters entering the subgrid model for feedback by observations of the present-day galaxy stellar mass function (GSMF) and galaxy sizes, we showed that EAGLE also reproduces many other properties of observed galaxies at  $z \sim 0$  to unprecedented levels. The focus of this paper is to explore whether

★ E-mail: [michelle.furlong@dur.ac.uk](mailto:michelle.furlong@dur.ac.uk)

the good agreement, specifically that between the simulated and observed stellar masses and star formation rates (SFRs), extends to higher redshifts.

Compared with semi-analytic models, hydrodynamical simulations such as *EAGLE* have fewer degrees of freedom and have to make fewer simplifying assumptions to model gas accretion and the crucial aspects of the feedback from star formation and accreting black holes that is thought to regulate galaxy formation. They also allow the study of properties of the circumgalactic and intergalactic media, providing important complementary tests of the realism of the simulation. Such a holistic approach is necessary to uncover possible degeneracies and inconsistencies in the model. Having a calibrated and well-tested subgrid model is of crucial importance, since it remains the dominant uncertainty in current simulations (Scannapieco et al. 2012).

**S15** present and motivate the subgrid physics implemented in *EAGLE*. An overriding consideration of the parametrization is that subgrid physics should only depend on local properties of the gas (e.g. density, metallicity), in contrast to other implementations used in the literature which for example depend explicitly on redshift, or on properties of the dark matter. Nevertheless, a physically reasonable set of parameters of the subgrid model for feedback exists for which the redshift  $z \sim 0$  GSMF and galaxy sizes agree to within 0.2 dex with the observations. This level of agreement is unprecedented, and similar to the systematic uncertainty in deriving galaxy stellar masses from broad-band observations. Other observations of the local Universe, such as the Tully–Fisher relation, the mass–metallicity relation and the column density distribution functions of intergalactic C IV and O VI are also reproduced, even though they were not used in calibrating the model and hence could be considered ‘predictions’.

In this paper, we focus on the build-up of the stellar mass density, and the evolution of galaxy stellar masses and SFRs, expanding the analysis of **S15** beyond  $z \sim 0$ . A similar analysis was presented by Genel et al. (2014), for the *ILLUSTRIS* simulation (Vogelsberger et al. 2014). They conclude that *ILLUSTRIS* reproduces the observed evolution of the GSMF from redshifts 0 to 7 well, but we note that they used the star formation history in their calibration process. Another difference with respect to Genel et al. (2014) is that we compare with recent galaxy surveys, which have dramatically tightened observational constraints on these measures of galaxy evolution. For example, *PRIMUS* (Moustakas et al. 2013), *ULTRAVISTA* (Ilbert et al. 2013; Muzzin et al. 2013) and *ZFOURGE* (Tomczak et al. 2014) provide improved constraints out to redshift 4. UV observations extend the comparison to even higher redshift, with inferred GSMFs available up to redshift 7 (González et al. 2011; Duncan et al. 2014). Observations of star formation rates also span the redshift range 0–7, with many different tracers of star formation (e.g. IR, radio, UV) providing consistency checks between data sets.

This paper is organized as follows: In Section 2, we provide a brief summary of *EAGLE* in particular the subgrid physics used. In Section 3, we compare the evolution of the stellar mass growth in the simulation to data out to redshift 7. We follow this with an analysis of the SFR density and specific star formation rates (SSFRs) in Section 4. In Section 5, we discuss the results and we summarize in Section 6. We generally find that the properties of the simulated galaxies agree with the observations to the level of the observational systematic uncertainties across all redshifts.

The *EAGLE* simulation suite adopts a flat  $\Lambda$  cold dark matter cosmogony with parameters from Planck (Planck Collaboration XVI 2014);  $\Omega_\Lambda = 0.693$ ,  $\Omega_m = 0.307$ ,  $\Omega_b = 0.048$ ,  $\sigma_8 = 0.8288$ ,  $n_s = 0.9611$  and  $H_0 = 67.77 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . The Chabrier (2003)

stellar initial mass function (IMF) is assumed in the simulations. Where necessary observational stellar masses and SFR densities have been renormalized to the Chabrier IMF<sup>1</sup> and volumes have been rescaled to the Planck cosmology. Galaxy stellar masses are computed within a spherical aperture of 30 proper kiloparsecs (pkpc) from the centre of potential of the galaxy. This definition mimics a 2D Petrosian mass often used in observations, as shown in **S15**. SFRs are computed within the same aperture. Distances and volumes are quoted in comoving units (e.g. comoving megaparsecs, cMpc), unless stated otherwise. Note that, unless explicitly stated, values are not given in  $h^{-1}$  units.

## 2 SIMULATIONS

The *EAGLE* simulation suite consists of a large number of cosmological simulations, with variations that include parameter changes relative to those of the reference subgrid formulation, other subgrid implementations, different numerical resolutions and a range of box sizes up to 100 cMpc boxes (**S15**; Crain et al. 2015). Simulations are denoted as, for example, L0100N1504, which corresponds to a simulation volume of  $L = 100$  cMpc on a side, using  $1504^3$  particles of dark matter and an equal number of baryonic particles. A prefix distinguishes subgrid variations, for example Ref-L100N1504 is our reference model. These simulations use advanced smoothed particle hydrodynamics (SPH) and state-of-the-art subgrid models to capture the unresolved physics. Cooling, metal enrichment, energy input from stellar feedback, black hole growth and feedback from AGN are included. The free parameters for stellar and AGN feedback contain considerable uncertainty (see **S15**), and so are calibrated to the redshift 0.1 GSMF, with consideration given to galaxy sizes. A complete description of the code, subgrid physics and parameters can be found in **S15**, while the motivation is given in **S15** and Crain et al. (2015). Here, we present a brief overview.

*CAMB* (Lewis, Challinor & Lasenby 2000, version Jan\_12) was used to generate the transfer function for the linear matter power spectrum with a Planck 1 (Planck Collaboration XVI 2014) cosmology. The Gaussian initial conditions were generated using the linear matter power spectrum and the random phases were taken from the public multiscale white noise *PANPHASIA* field (Jenkins 2013). Particle displacements and velocities are produced at redshift 127 using second-order Lagrangian perturbation theory (Jenkins 2010). See appendix B of **S15** for more detail.

The initial density field is evolved in time using an extensively modified version of the parallel *N*-body SPH code *GADGET-3* (Springel et al. 2008), which is essentially a more computationally efficient version of the public code *GADGET-2* described in detail by Springel (2005). In this Lagrangian code, a fluid is represented by a discrete set of particles, from which the gravitational and hydrodynamic forces are calculated. SPH properties, such as the density and pressure gradients, are computed by interpolating across neighbouring particles.

The code is modified to include updates to the hydrodynamics, as described in Dalla Vecchia et al. (in preparation, see also **S15** appendix A), collectively referred to as *ANARCHY*. The impact of these changes on cosmological simulations are discussed in Schaller et al. (in preparation). *ANARCHY* includes:

- (i) The pressure-entropy formulation of SPH described in Hopkins (2013).

<sup>1</sup> SSFRs are not renormalized as the correction for SFRs and stellar masses are similar and cancel each other.

**Table 1.** Box size, particle number, baryonic and dark matter particle mass, comoving and maximum proper gravitational softening for Ref-L100N1504 and Recal-L025N0752 simulations.

Simulation	L (cMpc)	N	$m_g$ ( $M_\odot$ )	$m_{dm}$ ( $M_\odot$ )	$\epsilon_{com}$ (ckpc)	$\epsilon_{prop}$ (pkpc)
Ref-L100N1504	100	$2 \times (1504)^3$	$1.81 \times 10^6$	$9.70 \times 10^6$	2.66	0.70
Recal-L025N0752	25	$2 \times (752)^3$	$2.26 \times 10^5$	$1.21 \times 10^6$	1.33	0.35

(ii) The artificial viscosity switch of Cullen & Dehnen (2010) and an artificial conduction switch described by Price (2008).

(iii) A C2 Wendland (1995) kernel with 58 neighbours to interpolate SPH properties across neighbouring particles.

(iv) The time step limiter from Durier & Dalla Vecchia (2012) that ensures feedback events are accurately modelled.

Two of the *EAGLE* simulations are analysed in this paper.<sup>2</sup> The first *EAGLE* simulation analysed in this paper is Ref-L100N1504, a  $(100 \text{ cMpc})^3$  periodic box with  $2 \times 1504^3$  particles. Initial masses for gas particles are  $1.81 \times 10^6 M_\odot$  and masses of dark matter particles are  $9.70 \times 10^6 M_\odot$ . Plummer equivalent comoving gravitational softenings are set to 1/25 of the initial mean interparticle spacing and are limited to a maximum physical size of 0.70 pkpc.

We also use simulation Recal-L025N0752 which has eight times better mass resolution and two times better spatial resolution in a  $(25 \text{ cMpc})^3$  box. The box sizes, particle numbers and resolutions are summarized in Table 1. Note that subgrid stellar feedback parameters and black hole growth and feedback parameters are recalibrated in the Recal-L025N0752 simulation, as explained in Section 2.2.

## 2.1 Subgrid physics

The baryonic subgrid physics included in these simulations is broadly based on that used for the *OWLS* (Schaye et al. 2010) and *GIMIC* (Crain et al. 2009) projects, although many improvements, in particular to the stellar feedback scheme and black hole growth, have been implemented. We emphasize that all subgrid physics models depend solely on local interstellar medium (ISM) properties.

(i) Radiative cooling and photoheating in the simulation are included as in Wiersma, Schaye & Smith (2009a). The element-by-element radiative rates are computed in the presence of the cosmic microwave background (CMB) and the Haardt & Madau (2001, hereafter *HM01*) model for UV and X-ray background radiation from quasars and galaxies. The 11 elements that dominate radiative cooling are tracked, namely H, He, C, N, O, Ne, Mg, Si, Fe, Ca and S. The cooling tables, as a function of density, temperature and redshift are produced using *CLOUDY*, version 07.02 (Ferland et al. 1998), assuming the gas is optically thin and in photoionization equilibrium.

Above the redshift of reionization the CMB and a Haardt & Madau (2001) UV-background up to 1 Ryd, to account for photodissociation of  $H_2$ , are applied. Hydrogen reionization is implemented by switching on the full Haardt & Madau (2001) background at redshift 11.5.

(ii) Star formation is implemented following Schaye & Dalla Vecchia (2008). Gas particles above a metallicity-dependent density threshold,  $n_H^*(Z)$ , have a probability of forming stars, determined by their pressure. The Kennicutt–Schmidt star formation law

(Kennicutt 1998), under the assumption of discs in vertical hydrostatic equilibrium, can be written as

$$\dot{m}_* = m_g A (1 \text{ Mpc}^{-2})^{-n} \left( \frac{\gamma}{G} f_g P \right)^{(n-1)/2}, \quad (1)$$

where  $m_g$  is the gas particle mass,  $A$  and  $n$  are the normalization and power index of the Kennicutt–Schmidt star formation law,  $\gamma = 5/3$  is the ratio of specific heats,  $G$  is the gravitational constant,  $f_g = 1$  is the gas fraction of the particle and  $P$  is its pressure. As a result the imposed star formation law is specified by the observational values of  $A = 1.515 \times 10^{-4} M_\odot \text{ yr}^{-1} \text{ kpc}^{-2}$  and  $n = 1.4$ , where we have decreased the amplitude by a factor of 1.65 relative to the value of Kennicutt (1998) to account for the use of a Chabrier, instead of Salpeter, IMF.

As we do not resolve the cold gas phase, a star formation threshold above which cold gas is expected to form is imposed. The star formation threshold is metallicity dependent and given by

$$n_H^*(Z) = 0.1 \text{ cm}^{-3} \left( \frac{Z}{0.002} \right)^{-0.64}, \quad (2)$$

where  $Z$  is the metallicity (from Schaye 2004, equations 19 and 24, also used in *SFTHRESHZ* model of the *OWLS* project).

A pressure floor as a function of density is imposed, of the form  $P \propto \rho^{\gamma_{\text{eff}}}$ , for gas with density above  $n_H^*(Z)$  and  $\gamma_{\text{eff}} = 4/3$ . This models the unresolved multiphase ISM. Our choice for  $\gamma_{\text{eff}}$  ensures that the Jeans mass is independent of density and prevents spurious fragmentation provided the Jeans mass is resolved at  $n_H^*(Z)$  (see Schaye & Dalla Vecchia 2008). Gas particles selected for star formation are converted to collisionless star particles, which represent a simple stellar population with a Chabrier (2003) IMF.

(iii) Stellar evolution and enrichment is based on Wiersma et al. (2009b) and detailed in S15. Metal enrichment due to mass-loss from AGB stars, winds from massive stars, core collapse supernovae and Type Ia supernovae of the 11 elements that are important for radiative cooling are tracked, using the yield tables of Marigo (2001), Portinari, Chiosi & Bressan (1998) and Thielemann, Argast & Brachwitz (2003). The total and metal mass lost from stars are added to the gas particles that are within an SPH kernel of the star particle.

(iv) Stellar feedback is treated stochastically, using the thermal injection method described in Dalla Vecchia & Schaye (2012). The total available energy from core-collapse supernovae for a Chabrier IMF assumes all stars in the stellar mass range  $6\text{--}100 M_\odot$  release  $10^{51}$  erg of energy into the ISM and the energy is injected after a delay of 30 Myr from the time the star particle is formed. Rather than heating all gas particle neighbours within the SPH kernel, neighbours are selected stochastically based on the available energy, then heated by a fixed temperature difference of  $\Delta T = 10^{7.5} \text{ K}$ . The stochastic heating distributes the energy over less mass than heating all neighbours. This results in a longer cooling time relative

<sup>2</sup> Two further simulations are considered in Appendix B.

<sup>3</sup>  $6\text{--}8 M_\odot$  stars explode as electron capture supernovae in models with convective overshoot, e.g. Chiosi, Bertelli & Bressan (1992).

to the sound crossing time across a resolution element, allowing the thermal energy to be converted to kinetic energy, thereby limiting spurious losses (Dalla Vecchia & Schaye 2012).

In *EAGLE*, the fraction of this available energy injected into the ISM depends on the local gas metallicity and density. The stellar feedback fraction, in units of the available core collapse supernova energy, is specified by a sigmoid function,

$$f_{\text{th}} = f_{\text{th,min}} + \frac{f_{\text{th,max}} - f_{\text{th,min}}}{1 + \left(\frac{Z}{0.1Z_{\odot}}\right)^{n_Z} \left(\frac{n_{\text{H,birth}}}{n_{\text{H,0}}}\right)^{-n_n}}, \quad (3)$$

where  $Z$  is the metallicity of the star particle,  $n_{\text{H,birth}}$  is the density of the star particle's parent gas particle when the star was formed and  $Z_{\odot} = 0.0127$  is the solar metallicity.

The values for  $f_{\text{th,max}}$  and  $f_{\text{th,min}}$ , the parameters for the maximum and minimum energy fractions, are fixed at 3 and 0.3 for both simulations analysed here. At low  $Z$  and high  $n_{\text{H,birth}}$ ,  $f_{\text{th}}$  asymptotes towards  $f_{\text{th,max}}$  and at high  $Z$  and low  $n_{\text{H,birth}}$  asymptotes towards  $f_{\text{th,min}}$ . Applying up to three times the available energy can be justified by appealing to the different forms of stellar feedback, e.g. supernova, radiation pressure, stellar winds which are not treated separately here as we do not have the resolution to resolve these forms of stellar feedback. This also offsets the remaining numerical radiative losses (Crain et al. 2015).

The power-law indexes are  $n_Z = n_n = 2/\ln(10)$  for the Ref model, with  $n_n$  changed to  $1/\ln(10)$  for the Recal model, resulting in weaker dependence of  $f_{\text{th}}$  on the density in the high-resolution model. The normalization of the density term,  $n_{\text{H,0}}$ , is set to  $0.67 \text{ cm}^{-3}$  for the Ref model and to  $0.25 \text{ cm}^{-3}$  for the Recal model. The feedback dependence is motivated in Crain et al. (2015).

(v) Black hole seeding and growth is implemented as follows. Haloes with a mass greater than  $10^{10} h^{-1} M_{\odot}$  are seeded with a black hole of  $10^5 h^{-1} M_{\odot}$ , using the method of Springel, Di Matteo & Hernquist (2005). Black holes can grow through mergers and accretion. Accretion of ambient gas on to black holes follows a modified Bondi–Hoyle formula that accounts for the angular momentum of the accreting gas (Rosas-Guevara et al. 2013). Differing from, e.g. Springel et al. (2005), Booth & Schaye (2009), Rosas-Guevara et al. (2013), the black hole accretion rate is not increased relative to the standard Bondi accretion rate in high-density regions. For the black hole growth there is one free parameter,  $C_{\text{visc}}$ , which is used to determine the accretion rate from

$$\dot{m}_{\text{accr}} = \min \left( \dot{m}_{\text{bondi}} \left[ C_{\text{visc}}^{-1} (c_s/V_{\phi})^3 \right], \dot{m}_{\text{bondi}} \right), \quad (4)$$

where  $c_s$  is the sound speed and  $V_{\phi}$  is the rotation speed of the gas around the black hole. The Bondi rate is given by

$$\dot{m}_{\text{bondi}} = \frac{4\pi G^2 m_{\text{BH}}^2 \rho}{(c_s^2 + v^2)^{3/2}}, \quad (5)$$

where  $v$  is the relative velocity of the black hole and the gas. The accretion rate is not allowed to exceed the Eddington rate,  $\dot{m}_{\text{Edd}}$ , given by

$$\dot{m}_{\text{Edd}} = \frac{4\pi G m_{\text{BH}} m_{\text{p}}}{\epsilon_r \sigma_{\text{T}} C}, \quad (6)$$

where  $m_{\text{p}}$  is the proton mass,  $\sigma_{\text{T}}$  is the Thomson scattering cross-section and  $\epsilon_r$  is the radiative efficiency of the accretion disc. The free parameter  $C_{\text{visc}}$  relates to the viscosity of the (subgrid) accretion disc and  $(c_s/V_{\phi})^3/C_{\text{visc}}$  relates the Bondi and viscous time-scales (see Rosas-Guevara et al. 2013, for more detail).

(vi) AGN feedback follows the accretion of mass on to the black hole. A fraction of the accreted gas is released as thermal energy into

**Table 2.** Values of parameters that differ between Ref-L100N1504 and Recal-L025N0752.

Simulation prefix	$n_{\text{H,0}}$ ( $\text{cm}^{-3}$ )	$n_n$	$C_{\text{visc}}^a$	$\Delta T_{\text{AGN}}$ (K)
Ref	0.67	$2/\ln(10)$	$2\pi$	$10^{8.5}$
Recal	0.25	$1/\ln(10)$	$2\pi \times 10^3$	$10^9$

<sup>a</sup>Note that the subgrid scheme is not very sensitive to the changes in  $C_{\text{visc}}$ , as shown in appendix B of Rosas-Guevara et al. (2013).

the surrounding gas. Stochastic heating, similar to the supernova feedback scheme, is implemented with a fixed heating temperature  $\Delta T_{\text{AGN}}$ , where  $\Delta T_{\text{AGN}}$  is a free parameter. The method used is based on that of Booth & Schaye (2009) and Dalla Vecchia & Schaye (2008), see S15 for more motivation.

The effect of varying some of the subgrid parameters is explored in Crain et al. (2015). The values of the parameters that differ between the two simulations used in this paper, Ref-L100N1504 and Recal-L025N0752, are listed in Table 2.

## 2.2 Resolution tests

We distinguish between the *strong* and *weak* numerical convergence of our simulations, as defined and motivated in S15. By strong convergence, we mean that simulations of different resolutions give numerically converged answer, *without* any change to the subgrid parameters. In S15, it is argued that strong convergence is not expected from current simulations, as higher resolution often implies changes in the subgrid models, for example energy injected by feedback events often scales directly with the mass of the star particle formed. In addition, with higher resolution, the physical conditions of the ISM and hence the computed radiative losses, will change. Without turning off radiative cooling or the hydrodynamics (which could be sensitive to the point at which they are turned back on), the changes to the ISM and radiative losses are expected to limit the strong convergence of the simulation.

The *EAGLE* project instead focuses on demonstrating that the simulations show good weak convergence (although S15 shows that the strong convergence of the simulation is on par with other hydrodynamical simulations). Weak convergence means that simulations of different resolutions give numerically converged results, *after* recalibrating one or more of the subgrid parameters. As it is argued in S15 that current simulations cannot make ab initio predictions for galaxy properties, due to the sensitivity of the results to the parameters of the subgrid models for feedback, and calibration is thus required, the high-resolution *EAGLE* simulation subgrid parameters are recalibrated to the same observables (the present-day GSMF, galaxy sizes and the stellar-mass black hole mass correlation) as the standard resolution simulations. This recalibrated high-resolution model, Recal-L025N0752, enables us to test the weak convergence behaviour of the simulation and to push our results for galaxy properties to eight times lower stellar mass. In Table 2, we highlight the parameters that are varied between the Ref and Recal models. In the main text of this paper, we consider weak convergence tests, strong convergence tests can be found in Appendix B.

As a simulation with a factor of 8 better mass resolution requires a minimum of eight times the CPU time (in practice the increase in time is longer due to the higher density regions resulting in shorter time steps and difficulties in producing perfectly scalable



algorithms), we compare the  $(100 \text{ cMpc})^3$  intermediate-resolution simulation to a  $(25 \text{ cMpc})^3$  high-resolution simulation. Note that for volume averaged properties the  $(25 \text{ cMpc})^3$  box differs from the  $(100 \text{ cMpc})^3$  box not only due to the resolution but also due to the absence of larger objects and denser environments in the smaller volume. As a result, for volume averaged quantities we present only the Ref-L100N1504 simulation in the following sections and revisit the convergence of these quantities in Appendix B. For quantities as a function of stellar mass, we present both the Ref-L100N1504 and Recal-L025N0752 simulations, although the comparison at high redshifts is limited by the small number of objects in the high-resolution simulation, which has a volume that is 64 times smaller.

### 2.3 Halo and galaxy definition

Halo finding is carried out by applying the friends-of-friends (FOF) method (Davis et al. 1985) on the dark matter, with a linking length of 0.2 times the mean interparticle separation. Baryonic particles are assigned to the group of their nearest dark matter particle. Self-bound overdensities within the group are found using `SUBFIND` (Springel et al. 2001; Dolag et al. 2009); these substructures are the galaxies in our simulation. A ‘central’ galaxy is the substructure with the largest mass within a halo. All other galaxies within a halo are ‘satellites’. Note that any FOF particles not associated with satellites are assigned to the central object, thus the mass of a central galaxy may extend throughout its halo.

A galaxy’s stellar mass is defined as the stellar mass associated with the subhalo within a 3D 30 pkpc radius, centred on the minimum of the subhalo’s centre of gravitational potential. Only mass that is bound to the subhalo is considered, thereby excluding mass from other subhaloes. This definition is equivalent to the total subhalo mass for low-mass objects, but excludes diffuse mass around larger subhaloes, which would contribute to the intracluster light (ICL). S15 shows that this aperture yields results that are close to a 2D Petrosian aperture, often used in observations, e.g. Li & White (2009). The same 3D 30 pkpc aperture is applied when computing the SFRs in galaxies, again considering only particles belonging to the subhalo. The aperture constraint has only a minimal effect on the SFRs because the vast majority of star formation occurs in the central 30 pkpc, even for massive galaxies.

## 3 EVOLUTION OF GALAXY STELLAR MASSES

We will begin this section by comparing the growth in stellar mass density across cosmic time in the largest *EAGLE* simulation, Ref-L100N1504, to a number of observational data sets. This is followed with a comparison of the evolution of the GSMF from redshift 0 to 7 and a discussion on the impact of stellar mass errors in the observations. We also consider the convergence of the GSMF in the simulation at different redshifts.

### 3.1 The stellar mass density

We begin the study of the evolution in the primary *EAGLE* simulation, Ref-L100N1504, by considering the build up of stellar mass. We present the stellar mass density ( $\rho_*$ ) as a function of lookback time in Fig. 1, with redshift on the upper axis. Plotting the stellar mass density as a function of time (rather than redshift, say) gives a better

visual impression of how much different epochs contribute to the net stellar build-up.

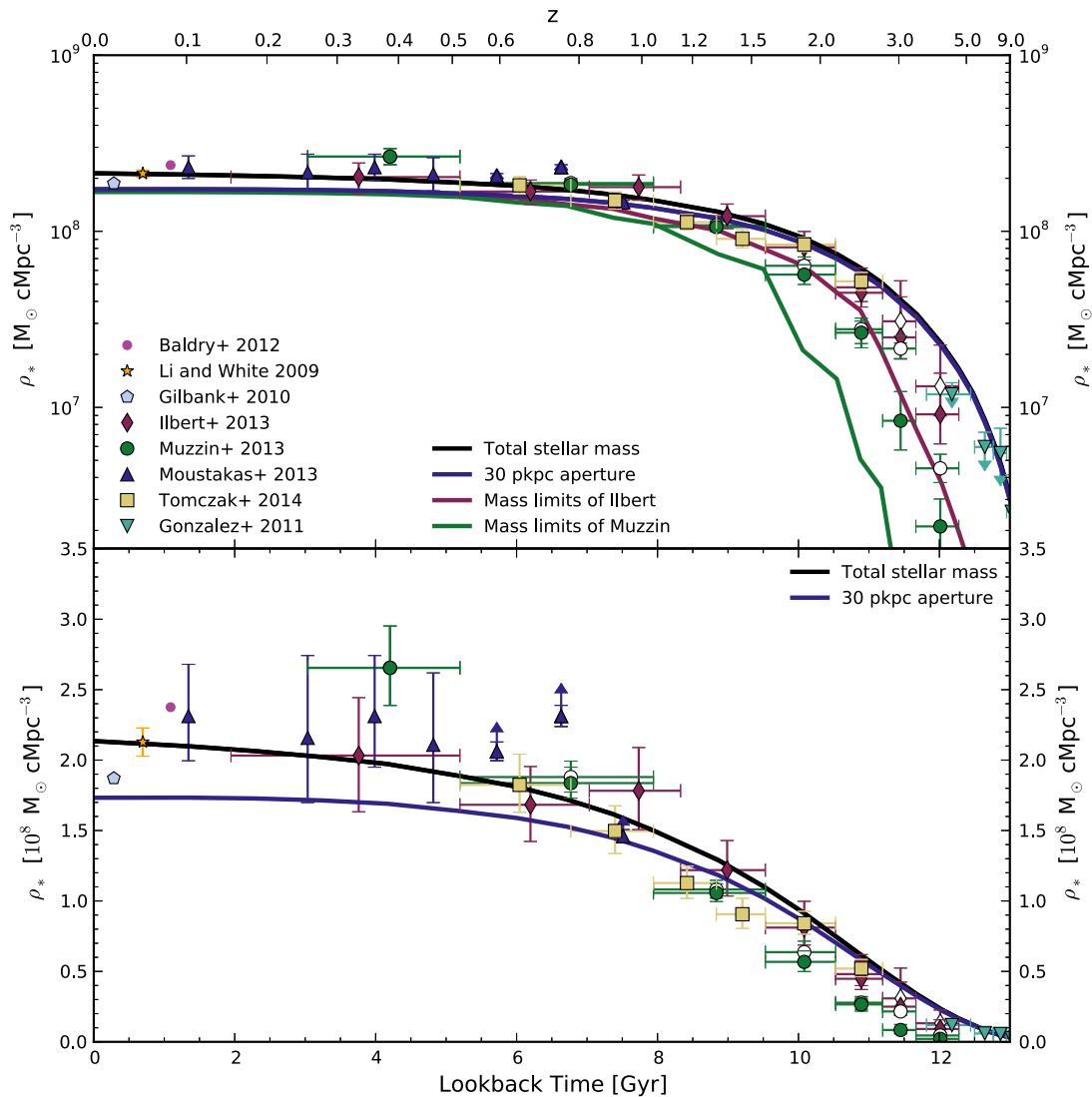
We added to this figure recent observational estimates of  $\rho_*$  from a number of galaxy surveys. Around redshift 0.1 we show data from Baldry et al. (2012) (GAMA survey), Li & White (2009) (SDSS), Gilbank et al. (2010b) (Stripe82 - SDSS) and Moustakas et al. (2013) (PRIMUS). The values agree to within  $0.55 \times 10^8 \text{ M}_\odot \text{ cMpc}^{-3}$ , which is better than 0.1 dex. The Moustakas et al. (2013) data set extends to redshift 1, providing an estimate for  $\rho_*$  for galaxies with masses greater than  $10^{9.5} \text{ M}_\odot$ . Note, however, that above redshift 0.725 the Moustakas et al. (2013) measurements of  $\rho_*$  are a lower limit as they only include galaxies with stellar masses of  $10^{10} \text{ M}_\odot$  or above. Ilbert et al. (2013) and Muzzin et al. (2013) estimate  $\rho_*$  from redshifts 0.2 to 4 from the *ULTRAVISTA* survey. These two data sets use the same observations but apply different signal-to-noise limits and analyses to infer stellar masses resulting in slightly different results. We include both studies in the figure to assess the intrinsic systematics in the interpretation of the data. Both data sets extrapolate the observations to  $10^8 \text{ M}_\odot$  to estimate a ‘total’ stellar mass density. The data sets are consistent within the estimated error bars up to redshift 3. Above redshift 3 they differ, primarily because of the strong dependence of  $\rho_*$  on how the extrapolation below the mass completeness limit of the survey is performed. The estimated  $\rho_*$  from observed galaxies can be compared to the extrapolated  $\rho_*$  for both data sets by comparing the filled and open symbols in Fig. 1. Tomczak et al. (2014) estimate stellar mass densities between redshifts 0.5 and 2.5 from the *ZFOURGE* survey. The mass completeness limits for this survey are below  $10^{9.5} \text{ M}_\odot$  at all redshifts, probing lower masses than other data sets at the same redshifts. For this data set no extrapolation is carried out in estimating  $\rho_*$ . In the simulations, galaxies with masses below  $10^9 \text{ M}_\odot$  contribute only 12 per cent to the stellar mass density at redshift 2 and their contribution decreases with decreasing redshift due to the flattening of the GSMF (see Section 3.2).

At redshifts below two the various observational measurements show agreement on the total stellar mass density to better than 0.1 dex. From redshift 2 to 4 the agreement is poorer, with differences up to 0.4 dex, primarily as a result of applying different extrapolations to correct for incompleteness. At redshifts above four only the UV observations of González et al. (2011) are shown. Note that these do not include corrections for nebular emission lines and hence may overestimate  $\rho_*$  (e.g. Smit et al. 2014). We therefore plot these values for  $\rho_*$  as upper limits.

The solid black line in each panel of Fig. 1 shows the build up of  $\rho_*$  in the simulation. The log scale used in the upper panel emphasizes the rapid fractional increase at high redshift. There is a rapid growth in  $\rho_*$  from the early Universe until 8 Gyr ago, around redshift 1, by which point 70 per cent of the present-day stellar mass has formed. The remaining 30 per cent forms in the 8 Gyr, from redshift 1 to 0. We find that 50 per cent of the present-day stellar mass was in place 9.75 Gyr ago, by redshift 1.6.

The simulation is in good agreement with the observed growth of stellar mass across the whole of cosmic time, falling within the error bars of the observational data sets. We find that 3.5 per cent of the baryons are in stars at redshift zero, which is close to the values of 3.5 and 4 per cent reported by Li & White (2009) and Baldry et al. (2012), respectively.

However, it should be noted that observed stellar mass densities are determined by integrating the GSMF, thereby excluding stellar mass associated with ICL. To carry out a fairer comparison, we apply a 3D 30 pkpc aperture to the simulated galaxies to mimic a 2D Petrosian aperture, as applied to many observations (see Section 2.3



**Figure 1.** The stellar mass density as a function of time on log and linear scales (top and bottom panels, respectively). The black solid curve is the total stellar mass density from the *EAGLE* simulation Ref-L0100N1504, and the blue curve is the stellar mass density in galaxies in that simulation (i.e. excluding ICL). Observational data are plotted as symbols, see the legend for the original source. Open symbols refer to observations that include extrapolations of the GSMF below the mass completeness of the survey, filled symbols are the raw data. Where necessary, data sets have been scaled to a Chabrier IMF and the Planck cosmology, as used in the simulation. The top panel shows  $\rho_*$  for all galaxies in the simulation in blue and  $\rho_*$  for galaxies above the completeness limit of observations by Ilbert et al. (2013) and Muzzin et al. (2013) in red and green, respectively. The corresponding data sets for Ilbert et al. (2013) and Muzzin et al. (2013) are coloured accordingly, and simulation lines should be compared to corresponding filled red and green symbols. From redshift 0 to 0.5,  $\rho_*$  in galaxies agrees with the observations at the 20 per cent level, with the simulated  $\rho_*$  lower by around 0.1 dex. At redshifts from 0.5 to 7, the model agrees well with the data, although the level of agreement above redshift 2 depends on the assumed incompleteness correction.

and S15). The aperture masses more accurately represent the stellar light that can be detected in observations. The result of the aperture correction is shown as a solid blue line in both panels.<sup>4</sup>

In this more realistic comparison of the model to observations, which excludes the ICL, we find that from high redshift to redshift 2 there is little difference between the total  $\rho_*$  and the aperture stellar mass density associated with galaxies. At these high redshifts, the simulation curve lies within the scatter of the total stellar mass density estimates from the observations of (González et al. 2011,

inverted triangles) and (Ilbert et al. 2013, open diamonds), although the simulation data are above the estimates of (Muzzin et al. 2013, open circles) above redshift 2. Between redshifts 2 and 0.1, the simulation data lie within the error bars from different observational estimates, although it is on the lower side of all observed values below redshift 0.9. At redshift 0.1, where  $\rho_*$  can be determined most accurately from observations, the simulation falls below the observations by a small amount, less than 0.1 dex, or 20 per cent. We will return to the source of this deficit in stellar mass at low redshift when studying the shape of the GSMF.

Returning to the agreement between redshifts 2 and 4, above redshift 2 the stellar mass density estimated from observations requires extrapolation below the mass completeness limit of the survey, as

<sup>4</sup> Note the mass in the simulation associated with the ICL resides in the largest haloes, as will be shown in a future paper.

**Table 3.** Mass completeness limit at redshifts 0.2 to 4 for GSMF observations of Ilbert et al. (2013) and Muzzin et al. (2013).

Redshift	Ilbert et al. (2013) $\log_{10}(M_*) (M_\odot)$	Muzzin et al. (2013) $\log_{10}(M_*) (M_\odot)$
0.2–0.5	7.93	8.37
0.5–0.8 <sup>a</sup>	8.70	8.92
0.8–1.1	9.13	–
1.1–1.5	9.42	9.48
1.5–2.0	9.67	10.03
2.0–2.5	10.04	10.54
2.5–3.0	10.24	10.76
3.0–4.0	10.27	10.94

<sup>a</sup>Muzzin et al. (2013) use redshift ranges 0.5–1.0 and 1.0–1.5.

discussed. To compare the simulation with the stellar mass density that is observed, without extrapolation, the red and green lines in the top panel show  $\rho_*$  from the simulation after applying the mass completeness limits of Ilbert et al. (2013) and Muzzin et al. (2013), respectively. The mass completeness limits applied are listed in Table 3. The red and green lines should be compared to the filled red diamonds and filled green circles, respectively, showing  $\rho_*$  from the observed galaxies without extrapolating below the mass completeness limit. Note that 30 pkpc apertures are still applied to the simulated galaxies for this comparison. When comparing with Ilbert et al. (2013), we find agreement at the level of the observational error bars from redshifts 0.2 to 4. However, Muzzin et al. (2013) find more stellar mass than the simulation after applying the mass completeness limits between redshifts 1.5 and 4. This can be understood by noting that the estimated mass completeness limit of Muzzin et al. (2013) is higher than that of Ilbert et al. (2013) (although both groups use the same survey data), resulting in only the most massive objects being detected at a given redshift. These objects are not sufficiently massive in the simulation when compared with the inferred GSMF from observations (without accounting for random or systematic mass errors), as will be shown next.

### 3.2 The evolution of the GSMF

The evolution of the stellar mass density of the Universe provides a good overview of the growth of stellar mass in the simulation. However, it does not test whether stars form in galaxies of the right mass. We now carry out a full comparison of the GSMFs in the simulation with those inferred from observations at different epochs.

The shape of the GSMF is often described by a Schechter (1976) function,

$$\Phi(M)dM = \Phi^* \left( \frac{M}{M_C} \right)^\alpha e^{-\frac{M}{M_C}} dM, \quad (7)$$

where  $M_C$  is the characteristic mass or ‘knee’,  $\Phi^*$  is the normalization and  $\alpha$  is the power-law slope for  $M \ll M_C$ . We will refer to the slope and knee throughout this comparison. In Appendix A, we fit the simulation GSMFs with Schechter functions to provide a simple way of characterizing the simulated GSMFs.

In Fig. 2, we compare the GSMF to the same observational data sets that were presented in Fig. 1 in terms of the total stellar mass density. The GSMFs from these different observations are consistent with each other within their estimated error bars up to redshift 2. Between redshifts 0 and 1, there is little evolution seen in the

observational data, all show a reasonably flat low-mass slope and a normalization that varies by less than 0.2 dex at  $10^{10} M_\odot$  over this redshift range. From redshift 1 to 2, there is a steepening of the slope at galaxy masses below  $10^{10} M_\odot$  and a drop in normalization of  $\sim 0.4$  dex. The drop in normalization appears to continue above redshift 2, although the observations do not probe below  $10^{10} M_\odot$  at redshifts 2–4.

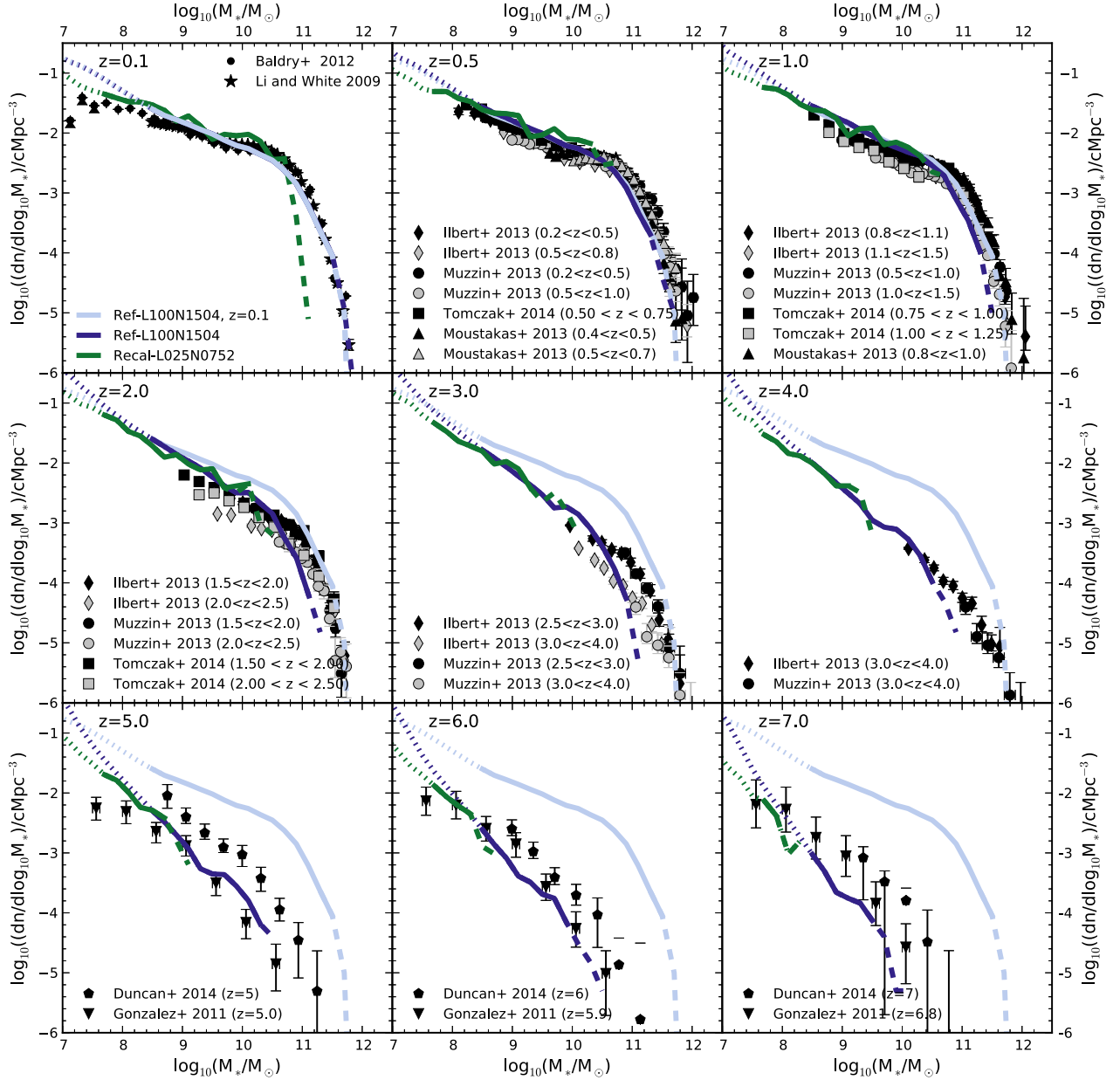
Observational data at redshifts 5, 6 and 7 from González et al. (2011) and Duncan et al. (2014), based on rest-frame UV observations, are shown in the bottom three panels of Fig. 2. There is no clear break in the GSMF at these high redshifts, so it is not clear that the distribution is described by a Schechter function in either data set. Both data sets show similar slopes above  $10^8 M_\odot$ . At low masses, below  $10^8 M_\odot$ , the data set of González et al. (2011) shows a flattening in the slope at all redshifts shown. These low masses are not probed by Duncan et al. (2014). At redshift 5, the data sets differ in amplitude by up to 0.8 dex. This offset reduces to  $\sim 0.2$  dex by redshift 7. A comparison of these data sets provides an impression of the systematic errors in determining the GSMF from observations at redshifts greater than 5.

We compare these observations to the evolution of the GSMFs predicted by Ref-L100N1504 between redshift 0.1 and 7, spanning 13 Gyr. The GSMF for Ref-L100N1504 is shown as a blue curve in Fig. 2, and to guide the eye, we repeat the redshift 0.1 GSMF in all panels in light blue. To facilitate a direct comparison with observational data, the GSMF from Ref-L100N1504 is convolved with an estimate of the likely uncertainty in observed stellar masses. Random errors in observed masses will skew the shape of the stellar mass function because more low-mass galaxies are scattered to higher masses than vice versa. We use the uncertainty quoted by Behroozi, Wechsler & Conroy (2013),  $\sigma(z) = \sigma_0 + \sigma_z z$  dex, where  $\sigma_0 = 0.07$  and  $\sigma_z = 0.04$ . This gives a fractional error in the galaxy stellar mass of 18 per cent at redshift 0.1 and 40 per cent at redshift 2. Note that this error does not account for any systematic uncertainties that arise when inferring the stellar mass from observations, which could range from 0.1 to 0.6 dex depending on redshift (see Section 3.2.1).

Recall that the observed GSMF at redshift 0.1 was used to calibrate the free parameters of the simulation. At this redshift, the simulation reproduces the reasonably flat slope of the observed GSMF below  $10^{10.5} M_\odot$ , with an exponential turnover at higher masses, between  $10^{10.5} M_\odot$  and  $10^{11} M_\odot$ . Overall, we find agreement within 0.2 dex over the mass range from  $2 \times 10^8 M_\odot$  to over  $10^{11} M_\odot$  and a very similar shape for the simulated and observed GSMF. In our implementation, the interplay between the subgrid stellar and AGN feedback models at the knee of the GSMF, at galaxy masses of around  $10^{10.5} M_\odot$ , results in a slight underabundance of galaxies relative to observations. As the stellar mass contained in this mass range dominates the stellar mass density of the Universe, this small offset accounts for the shortfall of stellar mass at the 20 per cent level seen at redshift 0 in  $\rho_*$  in Fig. 1 (blue curve).

In the simulation, there is almost no evolution in the GSMF from redshift 0 to 1, apart from a small decrease of 0.2 dex in galaxy masses at the very high-mass end. This can be seen by comparing the blue and light blue lines in the top panels, where the light blue line repeats the redshift 0.1 GSMF. A similar minimal evolution was reported based on the observational data of (Moustakas et al. 2013, triangles) from redshift 0 to 1, and is also seen in the other data sets shown.

From redshift 1–2, the simulation predicts strong evolution in the GSMF, in terms of its normalization, low-mass slope and the location of the break. Between these redshifts, spanning just 2.6 Gyr in



**Figure 2.** The GSMF at the redshifts shown in the upper left of each panel for simulation Ref-L100N1504 and Recal-L025N0752, in blue and green, respectively. When the stellar mass falls below the mass of 100 baryonic particles curves are dotted, when there are fewer than 10 galaxies in a stellar mass bin curves are dashed. The redshift 0.1 GSMF is reproduced in each panel as a light blue curve, to highlight the evolution. Comparing Ref-L100N1504 to Recal-L025N0752, the simulations show good convergence over the redshift range shown, where there are more than 10 galaxies per bin. The data points show observations as indicated in the legends. Where necessary, observational data have been converted to a Chabrier IMF and Planck cosmology. The black points represent the observational redshift bin below the simulation redshift, while the grey curves are from the redshift bin above the simulation snapshot. Within the expected mass errors, we find good agreement with observations of the GSMF from redshift 0 to 7. Between redshifts two and four the model tends to underestimate the masses of the brightest galaxies by around 0.2 dex, but these are very sensitive to the stellar mass errors in the observations, see text for discussion.

time, the stellar mass density almost doubles, from  $0.75$  to  $1.4 \times 10^8 \text{ M}_\odot \text{ cMpc}^{-3}$ , and the GSMF evolves significantly. From redshift 2–4, the normalization continues to drop and the mass corresponding to the break in the GSMF continues to decrease.

Although the trend of a decrease in normalization of the GSMF between redshift 1 and 2 is qualitatively consistent with what is seen in the observations, the normalization at redshift 2 at  $10^{9.5} \text{ M}_\odot$  is too high in the simulation by around 0.2 dex. There is also a suggestion

that the normalization of the GSMF in the simulation is too high at redshift 3, although observations do not probe below  $10^{10} \text{ M}_\odot$  at this redshift. It is therefore difficult to draw a strong conclusion from a comparison above redshift 2 without extrapolating the observational data. At redshift 2, there is also an offset at the massive end of the GSMF. The exponential break occurs at a mass that is around 0.2 dex lower than observed. However, the number of objects per bin in the simulation at redshift 2 above  $10^{11} \text{ M}_\odot$  falls below 10 providing a



poor statistical sample of the massive galaxy population. Increasing the box size may systematically boost the abundance of rare objects, such as that of galaxies above  $10^{11} M_{\odot}$  at redshift 2 and above. The break is also particularly sensitive to any errors in the stellar mass estimates, a point we will return to below.

Comparing the simulated GSMF to observations at redshifts 5–7, we find a similar shape to the observational data. The simulation has a similar trend with mass to González et al. (2011), however, it is offset in stellar mass from Duncan et al. (2014). No break in the GSMF is visible, neither in the simulation nor in the observations, at these high redshifts over the mass ranges considered here. Hence, for redshifts above 5, a Schechter fit may not be an appropriate description of the data.

### 3.2.1 Galaxy stellar mass errors

When comparing the simulation to observations, it is important to consider the role of stellar mass errors, both random and systematic. We begin by considering the random errors. In Fig. 3, the GSMF from Ref-L100N1504 is plotted at redshift 2 assuming no stellar mass error (red), a random mass error of  $0.07 + 0.04z$  (Behroozi et al. 2013) as in Fig. 2 (blue), resulting in an error of 40 per cent in galaxy stellar mass at redshift 2, and a mass error of a factor of 2 (green), i.e. 100 per cent. Where the GSMF is reasonably flat, i.e. at masses below  $10^{10.5} M_{\odot}$ , the impact of random uncertainty is minimal. However, above this mass the shape of the GSMF depends strongly on the random stellar mass errors in the observations, because more low-mass galaxies are scattered to high masses than

vice versa. If we increase the random errors, the exponential break becomes less sharp and the simulation agrees better with the observations.

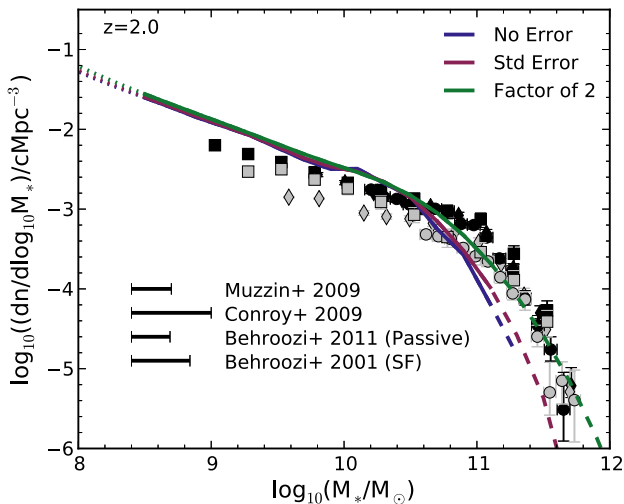
There are also systematic errors to consider in the determination of stellar masses from observed flux or spectra. Fitting the spectral energy distribution (SED) of a galaxy is sensitive to the choice of stellar population synthesis (SPS) model, e.g. due to the uncertainty in how to treat TP-AGB stars, the choice of dust model and the modelling of the star formation histories (e.g. Mitchell et al. 2013). Systematic variations in the stellar IMF would result in additional uncertainties, which are not considered here. The systematic uncertainties from SED modelling increase with redshift. At redshift 0, Taylor et al. (2011) quote  $\sim 0.1$  dex ( $1\sigma$ ) errors for GAMA data. At redshift 2, the estimated systematic error on stellar masses ranges from 0.3 dex (Muzzin et al. 2009) to 0.6 dex (Conroy et al. 2009), based on uncertainties in SPS models, dust and metallicities. Fig. 3 gives an impression of the size of these systematic errors by plotting values from Muzzin et al. (2009), Conroy et al. (2009) and Behroozi et al. (2013) in the bottom-left corner. The Behroozi et al. (2013) estimate is divided into star forming and passive galaxies due to the reduced sensitivity of passive galaxies to the assumed form of the star formation history. The systematic stellar mass errors are expected to shift the GSMF along the stellar mass axis. Considering the extent of the systematic uncertainties, we find the GSMF from EAGLE to be consistent with the observational data, although the low-mass slope may be slightly too steep. The observed evolutionary trends in the normalization and break are reproduced by the simulation, suggesting that the simulation is reasonably representative of the observed Universe.

### 3.2.2 Numerical convergence

Having found reasonable agreement between the evolution in the Ref-L100N1504 simulation and the observations, it is important to ask if the results are sensitive to numerical resolution. We consider only weak convergence tests here, i.e. we only examine the ability of the simulation to reproduce the observed evolution after recalibrating the high-resolution simulation to the same conditions (namely the redshift 0.1 GSMF) as used for the standard resolution simulation. In Fig. 2, the high-resolution model, Recal-L025N0752, is shown in green.

The 25 cMpc box is too small to sample the break in the GSMF accurately. To avoid box size issues, we do not consider the GSMF when there are fewer than 10 galaxies per bin, i.e. where the green curve is dashed. The 25 cMpc box also shows more fluctuations, due to poorer sampling of the large-scale modes in a smaller computational volume. At masses below  $10^8 M_{\odot}$ , when there are fewer than 100 star particles per galaxies in the Ref-L100N1504 simulation (blue dotted curve), the slope of the high-resolution simulation is flatter than that of Ref-L100N1504. Where the solid part of the blue and green curves overlap, there is excellent agreement, to better than 0.1 dex, between both resolutions across all redshifts. Overall, this amounts to good (weak) numerical convergence in the simulation across all redshifts that can be probed, given the limitations imposed on the test due to the small volume of the high-resolution run.

In summary, we have found the stellar mass density in the simulation to be close to the values estimated from observations, with a maximum offset of  $\sim 20$  per cent due to the slight undershooting of the EAGLE GSMF around the knee of the mass function. The observed evolutionary trends, in terms of changes in the shape and



**Figure 3.** The simulated GSMF at redshift 2 from EAGLE without random mass errors (red), convolved with the stellar mass error of Behroozi et al. (2013), used in Fig. 2, (blue) and with random errors of a factor of 2 (green). The random errors have a significant effect on the shape of the massive end of the GSMF, transforming the simulation from mildly discrepant with the observational data to being in excellent agreement with data. The Gaussian convolution with a stellar mass error is motivated by the random errors associated with the Malmquist bias. The horizontal black lines in the lower left of the figure indicate the estimated magnitudes of systematic errors in stellar masses according to Muzzin et al. (2009), Conroy, Gunn & White (2009) and Behroozi et al. (2013) at redshift 2. Systematic errors are expected to maintain the shape of the GSMF but would shift it horizontally. Within the estimated level of uncertainty in observations, the simulation shows agreement with observations of the GSMF, including the location of the break, although the low-mass slope may be slightly too steep.

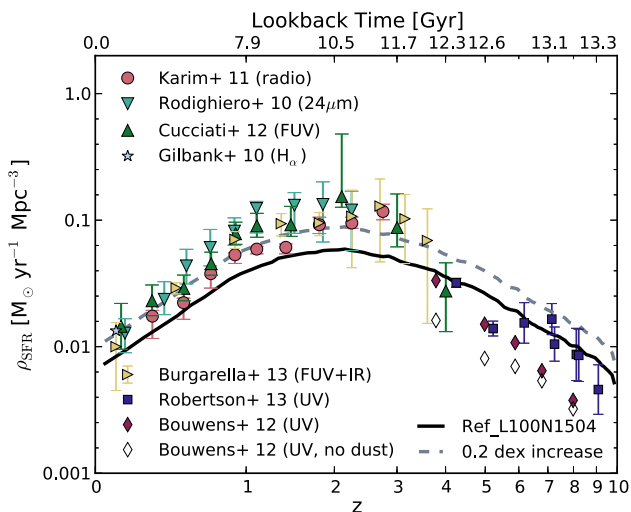
normalization of the GSMF between redshift 0.1 and 7 are reproduced, although the evolution in the normalization is not sufficiently strong in the simulation from redshift 1 to 2, with an offset in normalization at redshift 2 of  $\sim 0.2$  dex. The break in the GSMF occurs at too low a mass in the simulation compared to the observations at redshifts 2–4. However, the box size limits the number of objects produced in the simulation and we have shown that stellar mass errors play a significant role in defining the observed break of the GSMF. As a result of these uncertainties affecting the comparison, the remaining differences between the simulation and observations do not suggest significant discrepancies in the model.

## 4 EVOLUTION OF STAR FORMATION RATES

### 4.1 The cosmic star formation rate density

The SFR density ( $\rho_{\text{SFR}}$ ) as a function of redshift is plotted for simulation Ref-L100N1504 in Fig. 4. For comparison, observations from Gilbank et al. (2010b) [H $\alpha$ ], Rodighiero et al. (2010) [24 $\mu$ m], Karim et al. (2011) [Radio], Cucciati et al. (2012) [FUV], Bouwens et al. (2012) [UV], Robertson et al. (2013) [UV] and Burgarella et al. (2013) [FUV + FIR] are shown as well. This compilation of data covers a number of SFR tracers, providing an overview of  $\rho_{\text{SFR}}$  estimates from the literature, as well as an indication of the range of scatter and uncertainty arising from different methods of inferring  $\rho_{\text{SFR}}$ . There is a spread in the measured  $\rho_{\text{SFR}}$  of around 0.2 dex at redshifts less than two, while the estimated  $\rho_{\text{SFR}}$  include error bars of about  $\pm 0.15$  dex, with larger error bars above redshift two.

At high redshift, the simulated  $\rho_{\text{SFR}}$  (solid black curve) increases with time, peaks around redshift 2, followed by a decline of almost an order of magnitude to redshift 0. The simulation reproduces the shape of the observed  $\rho_{\text{SFR}}$  as a function of time very well, but falls below the measurements by an almost constant and small offset of 0.2 dex at  $z \leq 3$ . (The grey dashed line in Fig. 4 shows  $\rho_{\text{SFR}}$  increased by 0.2 dex.) While the simulation agrees reasonably well with the observational data at redshifts above 3, we caution



**Figure 4.** Evolution of the cosmic SFR density. The *EAGLE* simulation Ref-L100N1504 is plotted as a solid black curve, observational data are plotted as symbols. Open symbols from Bouwens et al. (2012) exclude a dust correction to the SFRs, giving an impression of the uncertainty in the measurement. The simulation tracks the evolution of the observed  $\rho_{\text{SFR}}$  very well, albeit with an almost constant 0.2 dex offset (grey dashed line) below redshift  $z \sim 3$ .

that these measurements are somewhat uncertain. For example, the difference between open and filled symbols for Bouwens et al. (2012) data shows the estimated dust correction that is applied to the observations.

### 4.2 Specific star formation rates

Observationally, a well-defined star forming sequence as a function of stellar mass has been found in the local Universe, which appears to hold up to a redshift of 3 (e.g. Noeske et al. 2007; Karim et al. 2011). It is described by a relation of the form

$$\frac{\dot{M}_*}{M_*} = \beta \left( \frac{M_*}{10^{10} M_\odot} \right)^\gamma, \quad (8)$$

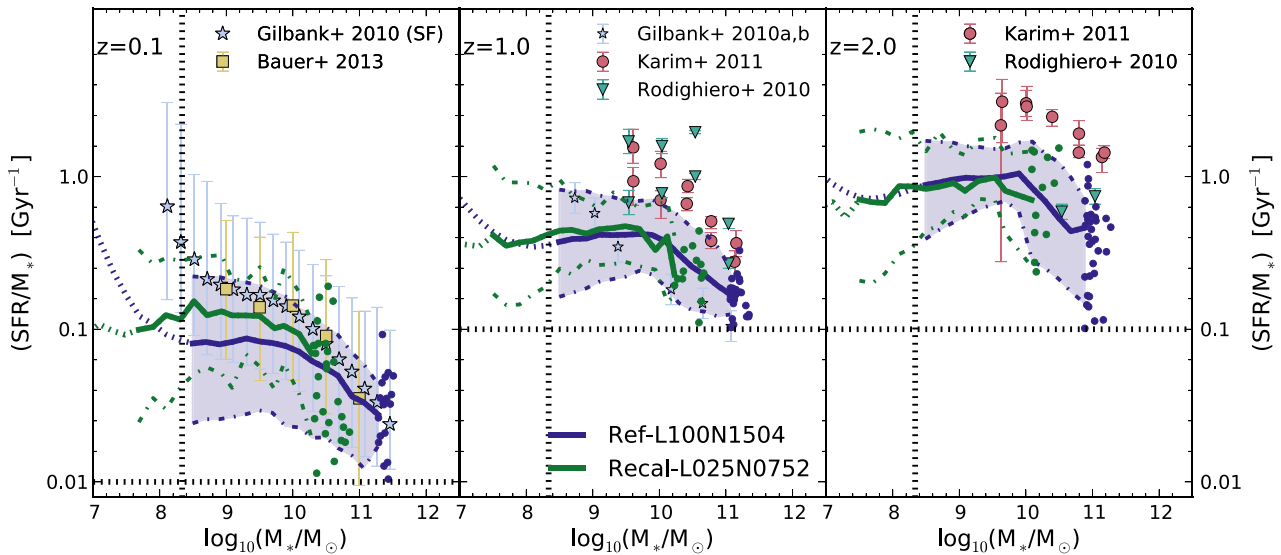
where  $\gamma$  is the logarithmic slope,  $\beta$  is the normalization and  $\dot{M}_*/M_*$  is the SSFR. Observations indicate that  $\gamma$  is negative but close to zero, and it is often assumed to be constant with stellar mass.

Fig. 5 shows the SSFR for star-forming galaxies as a function of galaxy stellar mass at redshifts 0.1, 1 and 2. The observational data sets for the SSFRs, we compare to at redshift 0.1 are from Gilbank et al. (2010b, stars) and Bauer et al. (2013, squares). These data sets show similar values for the normalization and slope and a similar scatter above  $10^9 M_\odot$ . Below  $10^9 M_\odot$  only Gilbank et al. (2010b) data are available. This data show an increase in the SSFR with decreasing stellar mass below  $10^{8.5} M_\odot$ . Rodighiero et al. (2010, inverted triangles), Karim et al. (2011, circles) and Gilbank et al. (2010a, stars) are shown at higher redshifts. Comparing these data sets, Rodighiero et al. (2010) and Karim et al. (2011) have similar slopes and normalization at redshifts 1 and 2. However, the Gilbank et al. (2010a) data are substantially (0.8 dex) lower in normalization over the mass ranges where it overlaps with Rodighiero et al. (2010) and Karim et al. (2011). The ROLES data used by Gilbank et al. (2010a) probes faint galaxies down to masses below  $10^9 M_\odot$ , but this deep survey covers only a small area of sky. The resulting small number statistics of massive galaxies may be driving this offset in SSFR from the other observational data sets.

The median SSFRs for star-forming galaxies from Ref-L100N1504 and Recal-L025N0752 are shown as blue and green curves, respectively. The horizontal dotted lines correspond to the SSFR cut ( $\sim 1$  dex below the observational data) used to separate star forming from passive galaxies.

At redshift 0.1, the SSFR in the simulations is reasonably independent of stellar mass (where well resolved) up to masses of  $10^{10} M_\odot$ . Above this mass, the SSFR decreases slowly with stellar mass. The simulations show a scatter of around 0.6 dex across the stellar mass range resolved by Ref-L100N1504. The normalization of the Recal-L025N0752 simulation lies 0.2 dex above that of Ref-L100N1504, as was already shown in S15. At low masses, when there are fewer than 100 star-forming particles per galaxy, there is an increase in SSFR with stellar mass in Ref-L100N1504. However, by comparing with Recal-L025N0752 we see that this is resolution driven.

The trend with stellar mass above  $10^9 M_\odot$  is similar in the simulations and the observations. However, there is an offset in the normalization from observations, where Recal-L025N0752 and Ref-L100N1504 are low by  $\sim 0.1$  and 0.3 dex, respectively. The increase in SSFR at a stellar mass of  $10^{8.5} M_\odot$  reported by Gilbank et al. (2010b) is not seen in the Recal-L025N0752 simulation, which has sufficient numerical resolution to compare to observations at these low masses. This could indicate that stellar feedback is too strong in low-mass galaxies, or perhaps that the observational data are not volume complete due to the difficulty in detecting low-mass



**Figure 5.** The SSFR,  $\dot{M}_*/M_*$ , as a function of galaxy stellar mass for Ref-L100N1504 and Recal-L025N0752 from left to right at redshifts 0.1, 1 and 2. The solid curves show the median relation for star forming galaxies, defined as those with an SSFR above the limit specified by the horizontal dotted line in each panel. The shaded region (dot-dashed curves) encloses the 10th to 90th percentiles for Ref-L100N1504 (Recal-L025N0752). Where there are fewer than 10 galaxies per bin, individual data points are shown. Lines are dotted when the stellar mass falls below that corresponding to 100 star-forming particles for the median SSFR and the mass of 100 baryonic particles, to indicate that resolution effects may be important. At redshift 0.1, the observational data of Gilbank et al. (2010b) and Bauer et al. (2013) are shown as light blue stars and yellow squares, respectively. Error bars enclose the 10th to 90th percentiles. At higher redshift, data from Gilbank et al. (2010a), Karim et al. (2011) and Rodighiero et al. (2010) are shown as light blue stars, pink circles and turquoise inverted triangles, respectively. The observed flat slope with stellar mass and the increase in normalization with redshift are reproduced by the simulations, but the simulations are lower in normalization by 0.2 to 0.4 dex, depending on redshift and the observational data set.

galaxies with low SFRs owing to their low surface brightness (see S15 for more discussion of the redshift 0.1 properties).

At higher redshifts, the simulation SSFRs increase in normalization, maintaining a flat slope below  $10^{10} M_\odot$ , with a shallow negative slope above this stellar mass. At redshifts between 1 and 2 the Recal-L025N0752 and Ref-L100N1504 SSFRs lie within 0.1 dex of each other across the stellar mass ranges for which both are resolved. The increase in normalization seen in the simulations reproduces the observed trend, although the offset in normalization increases to up to 0.5 dex when comparing to the data sets of Rodighiero et al. (2010) and Karim et al. (2011). Relative to the Gilbank et al. (2010a) data at redshift 1, the median SSFR from the simulation agrees to within around 0.2 dex. Comparing the slope of the SSFR– $M_*$  relation of Gilbank et al. (2010a) to the simulations, the simulation is flatter below  $10^{10} M_\odot$ , but is in agreement with the slopes of Karim et al. (2011) and Rodighiero et al. (2010).

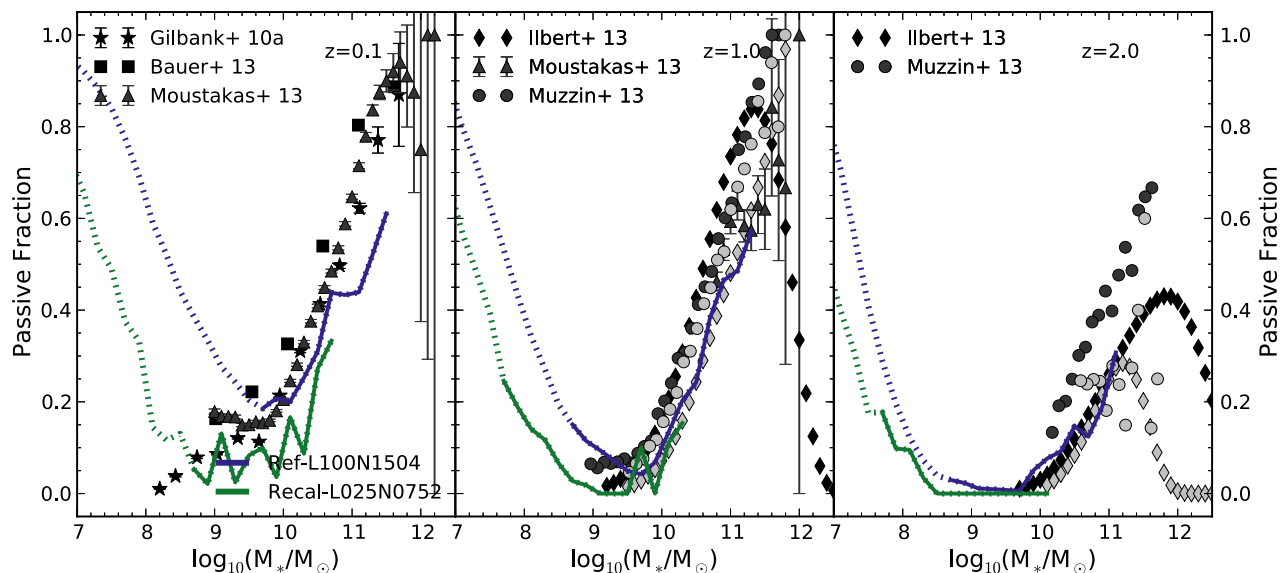
Observationally the galaxy population exhibits a bimodal colour distribution, which may imply a bimodality in the SSFR. To study this bimodality in the simulation, we show in Fig. 6 the passive fraction of galaxies as a function of mass at redshifts 0.1, 1 and 2. At higher redshifts, the simulation volume does not provide sufficiently massive galaxies to overlap with those detectable in observations. In the simulation, we define passive galaxies by a cut in SSFR that is an order of magnitude below the median observed SSFR (dotted horizontal line in Fig. 5). Varying this limit, while keeping it below the main star-forming sequence has negligible impact on the recovered median SSFR, although it can increase or decrease the passive fractions by around 10 per cent.

For comparison, passive fractions from Gilbank et al. (2010b), Bauer et al. (2013) and Moustakas et al. (2013) are shown at redshift 0.1 and from Moustakas et al. (2013), Muzzin et al. (2013) and Ilbert et al. (2013) at higher redshifts. For most observational data sets

shown, the passive fraction is determined based on a colour or SSFR cut as applied in the published data sets. Gilbank et al. (2010b) provide tabulated stellar masses and SFRs for each galaxy and we therefore apply the same SSFR cut as we use for the simulation data. At redshift 0.1, the dependence of passive fraction on stellar mass is similar for all observational data sets. At redshift 1, each observational data set shows the same trend, but there is a difference of up to 0.15 in the passive fraction for  $M_* \lesssim 10^{11} M_\odot$  for different data sets, and a larger difference above this mass. At redshift 2, agreement between data sets is poor.

The passive fraction from Ref-L100N1504 and Recal-L025N0752 are shown in blue and green, respectively. As a resolution guide, where the stellar mass is less than the maximum of 100 baryonic particles and 30 gas particles for the mass that corresponds to the SSFR cut, lines are dotted. As the SSFR cut evolves with redshift, this resolution guide evolves with redshift. The guide was chosen based on a comparison of the passive fractions for central galaxies in Ref-L100N1504 and Recal-L025N0752 (not shown). Both feedback and environment can quench star formation in galaxies. As different environments are probed in simulations of different box size, the passive fractions are expected to differ between Ref-L100N1504 and Recal-L025N0752, not only because of the resolution but also due to the box size. To overcome this, a comparison is carried out for central galaxies in the two simulations, which probe similar environments. This yields a difference in the passive fractions when a galaxy’s stellar mass is resolved by a minimum of 100 particles and the SSFR for the passive threshold is resolved by a minimum of 30 gas particles.

Over the resolved mass range, the passive fraction at redshift 0.1 follows a similar trend to the observational data, although there are too few passive galaxies between  $10^{10.5}$  and  $10^{11.5} M_\odot$  by around 15 per cent. In the simulations, passive fractions are lower at



**Figure 6.** The passive fraction as a function of galaxy stellar mass for Ref-L100N1504 and Recal-L025N0752 in blue and green, respectively, where galaxies with an SSFR below the horizontal dotted lines in Fig. 5 are defined as passive. Lines are dotted when the stellar mass falls below that corresponding to 30 star-forming particles for the SSFR limit. Data points show observations as indicated in the legends. The black points represent the observational redshift bin below the simulation redshift, while the grey curves are from the redshift bin above the simulation snapshot. Above  $10^9 M_\odot$ , the simulated passive fractions show similar normalization and slope with stellar mass to observations at all redshifts, with a small deficit of passive galaxies of around 15 per cent in the mass range  $10^{10.5}$  to  $10^{11.5} M_\odot$ . The upturn at low masses, below  $10^9 M_\odot$  is a numerical artefact.

redshift 1 than at redshift 0.1. This is consistent with what is seen in observational studies, although, there are again fewer passive galaxies in the range of  $10^{10.5}$  to  $10^{11.5} M_\odot$  than observed. At redshift 2, there is a further drop in the passive fraction of galaxies, both in the simulation and the observations. Summarizing, the passive fractions show the same trend as observations when galaxy masses and SFRs are resolved, although there are too few passive galaxies by  $\sim 15$  per cent in the stellar mass range  $10^{10.5}$ – $10^{11.5} M_\odot$ .

To better study the evolution of the SSFR and to extend the comparison to higher redshifts, we show in Fig. 7 the SSFR as a function of lookback time in three different stellar mass bins, of 0.5 dex, centred on  $10^{9.25}$ ,  $10^{9.75}$  and  $10^{10.25} M_\odot$ . The median SSFR for star-forming galaxies from Ref-L100N1504 and Recal-L025N0752 are shown in blue and green, respectively. In all mass bins the SSFR increases with lookback time. Comparing the two simulation, above redshift 1 the SSFRs of the two simulations are converged to within 0.1 dex. At lower redshifts, for stellar masses below  $10^{9.5} M_\odot$  Recal-L025N0752 has a slightly higher SSFR, by up to 0.2 dex. Similar trends are found when considering other mass bins of 0.5 dex between  $10^{8.5}$  and  $10^{11.5} M_\odot$ .

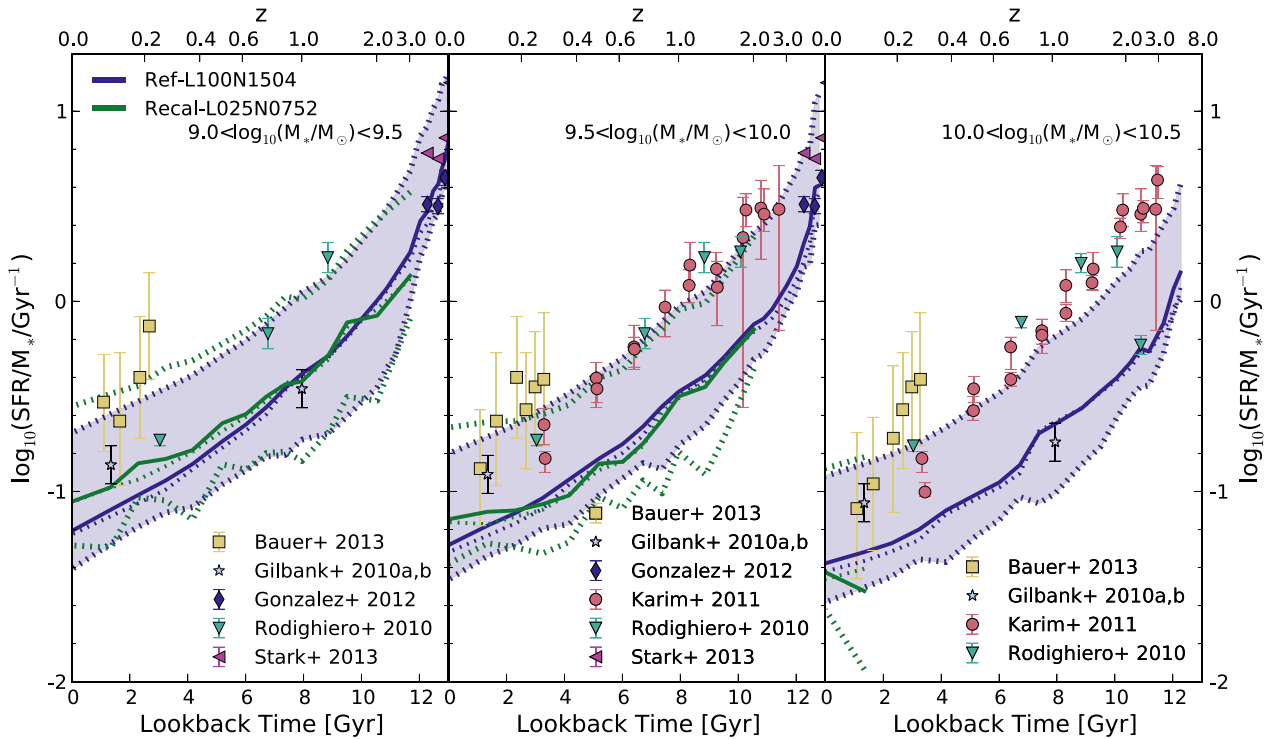
We compare the simulation data with the observations presented in Fig. 5, adding González et al. (2012) and Stark et al. (2013) at redshifts 4 and above. The observed trend with redshift is reproduced, there is, however, an offset in normalization of 0.2–0.5 dex at all times, across all mass ranges, as seen in Fig. 5. We found previously that the global SFR density was low by  $\sim 0.2$  dex across all redshifts relative to the values estimated from observations (Section 4.1). An offset in  $\rho_{\text{SFR}}$  does not convert directly into an offset in SSFR, due to the potential increase in stellar mass if SFRs were to increase. The offset in  $\rho_{\text{SFR}}$  thus cannot fully account for the offset in SSFR. If the SFRs were boosted by 0.3 dex across all mass ranges at all redshifts, as required to produce more consistent results relative to the observational data, the agreement for the stellar mass density from Section 3.1 would be broken. A possible solution to

the low SSFRs is that the star formation in the simulated galaxies is not sufficiently bursty. We will return to this possibility in the discussion.

As for the stellar mass, there are also uncertainties in the SFRs inferred from observations. Differences in the measured SFR density from different star formation tracers are of order 0.2 dex (as in Fig. 4), while Utomo et al. (2014) claim that SFRs inferred from UV and IR observations may be overestimated relative to those obtained by simultaneously modelling of stellar and dust emission. A recent study by Boquien, Buat & Perret (2014) also find SFRs to be overestimated, in FUV and *U* bands. Attempting to quantify the level of uncertainty in SFRs is difficult owing to the different sensitivity of each star formation tracer. UV observations require a large correction for the light that is absorbed. IR observations require information about the peak of the SED to constrain the total infrared luminosity and must assume all star formation is shrouded in dust if information from the UV is unavailable. Radio (and IR) observations can suffer from contamination by AGN and rely on an empirical calibration between the flux and SFR. At high redshift, where stacking is often necessary due to decreased ability to detect individual objects, there is a risk that the sample is incomplete, biasing results towards higher SFRs. Pérez-González et al. (2008) quote a factor of 2 (0.3 dex) in the uncertainty of IR SFRs due to dust, Muzzin et al. (2009) find a scatter of a factor of 2.8 (0.45 dex) depending on the bands available for fitting the SED. The SSFRs from the *EAGLE* Ref-L100N1504 model are only consistent with observations if the values inferred from the data are systematically high by about a factor of 2.

The systematic offset in SSFRs between models and observations has been noted before. Weinmann et al. (2012) and Genel et al. (2014) reported this issue for hydrodynamical simulations, while recent studies such as Mitchell et al. (2014) and White, Somerville & Ferguson (2015) revisited the issue with semi-analytic models. White et al. (2015) propose two plausible solutions to the issue





**Figure 7.** Evolution of the SSFR as a function of lookback time for stellar mass bins  $10^{9.0} < M_* < 10^{9.5} M_\odot$  (left),  $10^{9.5} < M_* < 10^{10.0} M_\odot$  (middle) and  $10^{10.0} < M_* < 10^{10.5} M_\odot$  (right) for Ref-L100N1504 and Recal-L025N0752, in blue and green, respectively. Solid curves show the median SSFR from the simulation for star-forming galaxies, the shaded region (dotted curves) enclose the 10th and 90th percentile values for Ref-L100N1504 (Recal-L025N0752). Medians are only shown when there are more than 10 galaxies per bin. Observational data from Gilbank et al. (2011), Bauer et al. (2013), Karim et al. (2011), Rodighiero et al. (2010), González et al. (2012) and Stark et al. (2013) are shown. The simulation shows good agreement with the observed shape of the SSFR evolution, but there is an offset in normalization of 0.2 to 0.4 dex, as seen in Fig. 5.

based on their semi-analytic modelling. In the first solution, star formation in low-mass galaxies forming at early times is preferentially suppressed, delaying star formation and providing further fuel for stars to form at later times. In the simulations presented here, Ref-L100N1504 and Recal-L025N0752, the dependence of the feedback on local gas metallicity and density does indeed result in preferential suppression of low-mass galaxies at early times and this does improve the behaviour of the SSFRs relative to *EAGLE* simulations with constant feedback or velocity dispersion dependent feedback (presented in Crain et al. 2015). However, to fully resolve the offset in SSFRs much stronger feedback is required in low-mass, high-redshift galaxies than the feedback that is implemented here. (Although the requirement for more efficient feedback may in part be a result of numerical radiative losses.) The second solution that White et al. (2015) appeal to, with a similar solution proposed by Mitchell et al. (2014), is limiting the cold gas available for star formation by reducing the accretion of gas from hot and ejected reservoirs on to haloes (see also Bower, Benson & Crain 2012). As our simulation follows the gravity and hydrodynamics of the gas, it is not a reasonable solution to apply to the accretion of gas in hydrodynamical simulation.

In summary, the simulation reproduces the shape of the evolution of  $\rho_{\text{SFR}}$  with redshift seen in observations with a 0.2 dex offset. The bimodality in SSFR, the slope with mass and the shape of the evolution of the SSFRs as a function of time are also reproduced by the simulation. However, the normalization is 0.2–0.5 dex too low at all redshifts and across all masses. This offset cannot be resolved by a simple systematic shift in SFRs in the simulation due

to the implications such a shift would have for  $\rho_*$ . However, the level of uncertainty in the data is such that the level of inconsistency in the *EAGLE* SSFRs may be smaller than suggested by current observations.

## 5 DISCUSSION

We have presented the evolution of the stellar masses and SFRs in two of the *EAGLE* cosmological hydrodynamical simulations. We have focused on Ref-L100N1504, a  $(100 \text{ cMpc})^3$  box with baryonic particle masses of  $1.81 \times 10^6 M_\odot$ , and Recal-L025N0752, a  $(25 \text{ cMpc})^3$  box with baryonic particle masses of  $2.26 \times 10^5 M_\odot$ . These simulations use advanced SPH techniques and state-of-the-art subgrid models, including cooling, metal enrichment, energy input from stellar feedback, black hole growth and feedback from AGN. The subgrid parameters depend only on local gas properties. The free parameters of the model have been calibrated to reproduce the observed local Universe GSMF, with consideration given to galaxy sizes [Crain et al. (2015)]. The resulting model has been shown to reproduce many observations around redshift 0, including the Tully–Fisher relation, SSFRs, the mass–metallicity relation, black hole masses and the column density distribution functions of intergalactic C IV and O VI (S15).

In this paper, we extend the comparison with observations of galaxy stellar masses and SFRs from redshift 0 to 7. This comparison with observations enables us to carry out a multi-epoch verification of the *EAGLE* galaxy formation model, where the galaxy properties in this comparison are predictions of the model, i.e.

evolution histories were not considered during the calibration of model parameters.

We began our comparison by finding a better than 20 per cent agreement with the evolution of the stellar mass density across all epochs (Fig. 1). For the GSMF, good agreement was typically found for the evolution of the normalization and break when comparing the simulation to observationally inferred data (Fig. 2). The normalization remains reasonably constant from redshift 0.1 to 1 and then decreases to redshift 2. The decrease continues at higher redshifts. Although this behaviour is qualitatively consistent with observations, at redshift 2 the normalization below  $10^{10.5} M_{\odot}$  is too high by  $\sim 0.2$  dex. Semi-analytical models have also reported normalizations that are too high relative to observations at  $z \sim 2$  (e.g. Weinmann et al. 2012, although see Henriques et al. (2014) for a possible solution in semi-analytics). In the current *EAGLE* implementation of stellar feedback, galaxies with low metallicity and high density, typical in the early Universe, experience strong feedback. The available feedback energy can be up to three times that available from core-collapse supernova, which compensates for numerical radiative losses. A comparison with the normalization of the observed GSMF at redshift 2 suggests that even more efficient stellar feedback is required in low-mass objects at redshifts above 2. More efficient feedback at high redshifts could provide surplus gas at later times, through recycling, helping to boost the SSFRs ( $= \dot{M}_*/M_*$ ), as is required based on the comparison with observational data in Fig. 7.

The break in the GSMF in the simulation evolves in a similar way to that observed, however, between redshifts 2 and 4 there is too little mass in simulated galaxies above  $10^{11} M_{\odot}$ , suggesting that less efficient AGN feedback (or stellar feedback in high-mass objects) at high redshifts is required to produce the observed evolution of the break in the GSMF. Less efficient AGN feedback at high redshifts would also result in more star formation around the peak epoch of star formation, at redshift 2, as favoured by current observational data for the SFR density. The requirement for weaker AGN feedback, however, is very sensitive to the stellar mass errors that arise from inferring the GSMF from observations. While recent observations of the GSMF are typically consistent with each other within their error bars, it is important to consider both random and systematic uncertainties in inferring stellar mass from observed flux, as shown in Fig. 3. As a result of the sensitivity of the exponential break in the GSMF to the stellar mass errors, it is difficult to determine if the AGN are indeed overly effective in the simulation.

The largest discrepancy we find with observational data is in the SSFRs of star-forming galaxies, which are 0.2 to 0.5 dex below the values inferred from observations across all of cosmic time (Fig. 7). This discrepancy cannot be explained as a simple systematic offset in the simulation, as we have shown the stellar mass density to be consistent with observations to within 0.1 dex. Applying a systematic boost to the SFRs of 0.3 dex would undo the agreement in the stellar mass density. It is puzzling that the SSFRs are systematically low, yet the stellar mass growth is consistent with the observational data. However, we have also found that the galaxy passive fractions appear too low by up to 15 per cent between  $10^{10.5}$  and  $10^{11.5} M_{\odot}$  (Fig. 6). Assuming that the observed SFRs are accurate, a potential solution to the low SSFRs is that the star formation is not sufficiently bursty. More bursty episodes of star formation could produce the same stellar mass with higher SFRs over shorter time periods than in the current simulation. This solution has the advantage that it would also increase the passive fractions, as galaxies would be star forming for a smaller fraction of the time.

Observed stellar masses and SFRs are uncertain at the 0.1 to 0.3 dex level across all observed redshifts. Until recently hydrodynamical simulations have struggled to reproduce redshift 0 galaxy populations within the observational uncertainties, not to mention the evolution of the galaxy population. The simultaneous comparison to stellar masses and SFRs across cosmic time thus provides a stringent test for the evolution of galaxy properties in our galaxy formation model. The *EAGLE* Ref-L100N1504 simulation performs relatively well in this test, verifying that the simulation produces galaxies with reasonable formation histories, for a redshift 0 galaxy population that is representative of the observed Universe. The agreement with observational data from redshifts 0 to 7 is at the level of the systematic uncertainties and follows the observed evolutionary trends. This gives us confidence that the model can be used as a reliable tool for interpreting observations and to explore the physics of galaxy formation. To give further confidence, our simulation shows weak numerical convergence, as defined in Section 2.2, of the GSMF to within 0.1 dex for galaxies of stellar masses greater than 100 baryonic particles<sup>5</sup> and of the SSFRs to within 0.1 dex when SFRs are resolved by a minimum of 100 star forming particles when going to a factor of 8 higher resolution. This level of convergence enables us to extend the galaxy population to lower stellar masses, by a factor of 8, using Recal-L025N0752, the higher-resolution simulation.

While there is scope to improve agreement with observational data, it is not clear that this should currently be a priority for a number of reasons. Given that the level of systematic uncertainty in the observations are similar to the level of agreement with the simulation, better agreement with observations would not automatically translate into more confidence in the model. Secondly, as hydrodynamical simulations are computationally expensive, full parameter space searches are unfeasible using current technology. Finally, it is likely that achieving better agreement with observations would require more complex parametrization of the subgrid models, which would be better motivated if changes were supported by small-scale simulations modelling ISM physics and smoothed to the resolution of current cosmological simulations. While many studies of this kind are underway (e.g. Creasey, Theuns & Bower 2013), they do not yet model all the relevant physics and currently require too much computational time to be incorporated into full cosmological simulations.

## 6 SUMMARY

We have compared the build-up of the stellar mass density, and the evolution of the GSMF and galaxy SFRs in the *EAGLE* cosmological simulations to recent observations. The *EAGLE* suite includes cosmologically representative volumes of up to  $(100 \text{ cMpc})^3$ , as well as smaller boxes run with higher numerical resolution to assess convergence and to extend the results to lower-mass galaxies. The simulations include physically motivated subgrid models for processes that cannot be resolved, with parameters calibrated to reproduce the observed redshift  $z \sim 0$  GSMF and galaxy sizes. *EAGLE* is described in detail and compared with a variety of observations of the present-day Universe in S15. In this paper, we investigated whether the good agreement between simulations and observations of galaxy masses and SFRs at  $z \sim 0$  extends to higher redshift,  $z = 0 \rightarrow 7$ . Our main findings are as follows.

- (i) The stellar mass density in the simulation tracks the observed value to within 20 per cent across cosmic time (Fig. 1). Observed

<sup>5</sup> Strong numerical convergence tests are presented in Appendix B.

trends in the evolution of the GSMF are reproduced to within plausible observational uncertainties, over the full redshift range  $z = 0 \rightarrow 7$  (Fig. 2).

(ii) The observed shape of the evolution of the SFR density (Fig. 4), and the trends of SSFR,  $\dot{M}_*/M_*$ , as a function of stellar mass and lookback time (Figs 5 and 7), are all reproduced accurately. The fraction of passive galaxies increases with stellar mass in the simulation, in agreement with the observed trend (Fig. 6).

(iii) Below stellar masses of  $\sim 10^{10.5} M_\odot$ , the normalization of the galaxy stellar mass function is above the observations by  $\sim 0.2$  dex at redshift 2. There is a similar offset in the normalization of the SSFRs, which are low by 0.2–0.5 dex across all redshifts. The recent papers of Mitchell et al. (2014) and White et al. (2015) highlighted a similar discrepancy with the data, based on semi-analytical models. These apparent discrepancies may result from systematic uncertainties in the observations. However, if they are real, then this would imply that even stronger feedback is required at high redshift than what is currently implemented in EAGLE. Burstier star formation histories could possibly also resolve the apparent discrepancy.

(iv) GSMFs and SFRs are reasonably well converged across all redshifts at which the convergence can be tested (Figs 2 and 5).

## ACKNOWLEDGEMENTS

The authors are very grateful for the endless technical support provided by Dr Lydia Heck during the preparation of these simulations and during post processing. MF thanks Violeta Gonzalez-Perez and Peter Mitchell for providing observational data sets.

RAC is a Royal Society University Research Fellow. This work used the DiRAC Data Centric system at Durham University, operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility ([www.dirac.ac.uk](http://www.dirac.ac.uk)). This equipment was funded by BIS National E-infrastructure capital grant ST/K00042X/1, STFC capital grant ST/H008519/1, STFC DiRAC Operations grant ST/K003267/1 and Durham University. DiRAC is part of the National E-Infrastructure. We also gratefully acknowledge PRACE for awarding us access to the resource Curie based in France at Trés Grand Centre de Calcul. This work was sponsored by the Dutch National Computing Facilities Foundation (NCF) for the use of supercomputer facilities, with financial support from the Netherlands Organization for Scientific Research (NWO). The research was supported in part by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreements 278594-GasAroundGalaxies, GA 267291 Cosmiway, and 321334 dustygal, the Interuniversity Attraction Poles Programme initiated by the Belgian Science Policy OWNce ([AP P7/08 CHARM]), the National Science Foundation under Grant No. NSF PHY11-25915, the UK Science and Technology Facilities Council (grant numbers ST/F001166/1 and ST/I000976/1), Rolling and Consolidating Grants to the ICC, Marie Curie Reintegration Grant PERG06-GA-2009-256573, Marie Curie Initial Training Network Cosmocomp (PITN-GA-2009-238356).

## REFERENCES

- Baldry I. K. et al., 2012, MNRAS, 421, 621  
 Bauer A. E. et al., 2013, MNRAS, 434, 209  
 Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57  
 Booth C. M., Schaye J., 2009, MNRAS, 398, 53  
 Boquien M., Buat V., Perret V., 2014, A&A, 571, A72  
 Bouwens R. J. et al., 2012, ApJ, 754, 83  
 Bower R. G., Benson A. J., Crain R. A., 2012, MNRAS, 422, 2816  
 Burgarella D. et al., 2013, A&A, 554, A70  
 Chabrier G., 2003, PASP, 115, 763  
 Chiosi C., Bertelli G., Bressan A., 1992, ARA&A, 30, 235  
 Conroy C., Gunn J. E., White M., 2009, ApJ, 699, 486  
 Crain R. A. et al., 2009, MNRAS, 399, 1773  
 Crain R. A. et al., 2015, MNRAS, 450, 1937  
 Creasey P., Theuns T., Bower R. G., 2013, MNRAS, 429, 1922  
 Cucciati O. et al., 2012, A&A, 539, A31  
 Cullen L., Dehnen W., 2010, MNRAS, 408, 669  
 Dalla Vecchia C., Schaye J., 2008, MNRAS, 387, 1431  
 Dalla Vecchia C., Schaye J., 2012, MNRAS, 426, 140  
 Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371  
 Dolag K., Borgani S., Murante G., Springel V., 2009, MNRAS, 399, 497  
 Duncan K. et al., 2014, MNRAS, 444, 2960  
 Durier F., Dalla Vecchia C., 2012, MNRAS, 419, 465  
 Ferland G. J., Korista K. T., Verner D. A., Ferguson J. W., Kingdon J. B., Verner E. M., 1998, PASP, 110, 761  
 Genel S. et al., 2014, MNRAS, 445, 175  
 Gilbank D. G. et al., 2010a, MNRAS, 405, 2419  
 Gilbank D. G., Baldry I. K., Balogh M. L., Glazebrook K., Bower R. G., 2010b, MNRAS, 405, 2594  
 Gilbank D. G. et al., 2011, MNRAS, 414, 304  
 González V., Labbé I., Bouwens R. J., Illingworth G., Franx M., Kriek M., 2011, ApJ, 735, L34  
 González V., Bouwens R. J., Labbé I., Illingworth G., Oesch P., Franx M., Magee D., 2012, ApJ, 755, 148  
 Haardt F., Madau P., 2001, in Neumann D. M., Tran J. T. V., eds, Clusters of Galaxies and the High Redshift Universe Observed in X-rays CEA, Saclay, p. 64 (HM01)  
 Henriques B., White S., Thomas P., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2014, Few-Body Syst., 55, 1013  
 Hopkins P. F., 2013, MNRAS, 428, 2840  
 Ilbert O. et al., 2013, A&A, 556, A55  
 Jenkins A., 2010, MNRAS, 403, 1859  
 Jenkins A., 2013, MNRAS, 434, 2094  
 Karim A. et al., 2011, ApJ, 730, 61  
 Kennicutt R. C., Jr, 1998, ApJ, 498, 541  
 Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473  
 Li C., White S. D. M., 2009, MNRAS, 398, 2177  
 Marigo P., 2001, A&A, 370, 194  
 Mitchell P. D., Lacey C. G., Baugh C. M., Cole S., 2013, MNRAS, 435, 87  
 Mitchell P. D., Lacey C. G., Cole S., Baugh C. M., 2014, MNRAS, 444, 2637  
 Moustakas J. et al., 2013, ApJ, 767, 50  
 Muzzin A., Marchesini D., van Dokkum P. G., Labbé I., Kriek M., Franx M., 2009, ApJ, 701, 1839  
 Muzzin A. et al., 2013, ApJ, 777, 18  
 Noeske K. G. et al., 2007, ApJ, 660, L43  
 Pérez-González P. G. et al., 2008, ApJ, 675, 234  
 Planck Collaboration XVI, 2014, A&A, 571, A16  
 Portinari L., Chiosi C., Bressan A., 1998, A&A, 334, 505  
 Price D. J., 2008, J. Comput. Phys., 227, 10040  
 Robertson B. E. et al., 2013, ApJ, 768, 71  
 Rodighiero G. et al., 2010, A&A, 518, L25  
 Rosas-Guevara Y. M. et al., 2013, preprint ([arXiv:1312.0598](https://arxiv.org/abs/1312.0598))  
 Scannapieco C. et al., 2012, MNRAS, 423, 1726  
 Schaye J., 2004, ApJ, 609, 667  
 Schaye J., Dalla Vecchia C., 2008, MNRAS, 383, 1210  
 Schaye J. et al., 2010, MNRAS, 402, 1536  
 Schaye J. et al., 2015, MNRAS, 446, 521 (S15)  
 Schechter P., 1976, ApJ, 203, 297  
 Smit R. et al., 2014, ApJ, 784, 58  
 Springel V., 2005, MNRAS, 364, 1105  
 Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726  
 Springel V., Di Matteo T., Hernquist L., 2005, MNRAS, 361, 776

- Springel V. et al., 2008, MNRAS, 391, 1685  
 Stark D. P., Schenker M. A., Ellis R., Robertson B., McLure R., Dunlop J., 2013, ApJ, 763, 129  
 Taylor E. N. et al., 2011, MNRAS, 418, 1587  
 Thielemann F. K., Argast D., Brachwitz F., 2003, in Hillebrandt W., Leibundgut B., eds, *From Twilight to Highlight: The Physics of Supernovae*. Springer, Berlin, p. 331  
 Tomczak A. R. et al., 2014, ApJ, 783, 85  
 Utomo D., Kriek M., Labbé I., Conroy C., Fumagalli M., 2014, ApJ, 783, L30  
 Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, MNRAS, 436, 3031  
 Vogelsberger M. et al., 2014, MNRAS, 444, 1518  
 Weinmann S. M., Pasquali A., Oppenheimer B. D., Finlator K., Mendel J. T., Crain R. A., Macciò A. V., 2012, MNRAS, 426, 2797  
 Wendland H., 1995, Adv. Comput. Math., 4, 389  
 White S. D. M., Frenk C. S., 1991, ApJ, 379, 52  
 White S. D. M., Rees M. J., 1978, MNRAS, 183, 341  
 White C. E., Somerville R. S., Ferguson H. C., 2015, ApJ, 799, 201  
 Wiersma R. P. C., Schaye J., Smith B. D., 2009a, MNRAS, 393, 99  
 Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009b, MNRAS, 399, 574

## APPENDIX A: SCHECHTER FUNCTION FITS

To provide a simple way of reproducing the *EAGLE* GSMFs and to quantify the trends seen in the evolution of the normalization and the exponential break, we have fit the *EAGLE* GSMFs with Schechter functions. We fit the GSMFs of Ref-L100N1504 from redshifts 0.1 to 4 that were shown in Fig. 2 (blue curves) with single Schechter functions (equation 7) and double Schechter functions,

$$\Phi(M)dM = \left[ \Phi_1^* \left( \frac{M}{M_C} \right)^{\alpha_1} + \Phi_2^* \left( \frac{M}{M_C} \right)^{\alpha_2} \right] e^{-M/M_C} dM, \quad (A1)$$

which is the sum of two Schechter functions with the same characteristic mass,  $M_C$ , but different normalizations,  $\Phi_1^*$  and  $\Phi_2^*$  and different low-mass slopes,  $\alpha_1$  and  $\alpha_2$ . Double Schechter fits are increasingly used in observational studies fitting the GSMF. We use least-squares fitting with bins of width 0.2 dex in stellar mass. Bins are weighted by their Poisson error, thereby down weighting the poorly sampled galaxies in the most massive stellar mass bins. The

fits over the mass range  $10^8$ – $10^{12} M_\odot$  are presented in Table A1. These fits compared to the simulation data can be seen in Fig. A1.

To understand the dependence of the Schechter function parameters on the fitted mass range, we applied our fitting routine over three mass ranges, from  $10^8$ ,  $10^9$  and  $10^{10}$  to  $10^{12} M_\odot$ . Fig. A2 shows the evolution of the Schechter function parameters  $M_C$ ,  $\Phi^*$  and  $\alpha$  for the single Schechter function fits. For the single Schechter fit,  $M_C$  drops over the redshift range 0–4 for all mass ranges. However, the extent of the decrease depends on the fitting range. For example, there is a decrease of 0.5 dex when fitting above  $10^8 M_\odot$  compared to a 0.3 dex decrease for fits above  $10^9$  and  $10^{10} M_\odot$ .  $\Phi^*$  is reasonably flat until redshift one, with a decrease at redshifts above one for all fits. There is however an obvious difference in the value of  $\Phi^*$  recovered for different fitting ranges, and there is also a difference in their variation with redshift. The opposite changes in  $M_C$  and  $\Phi^*$  for the different mass ranges highlight the degeneracy between these two parameters.

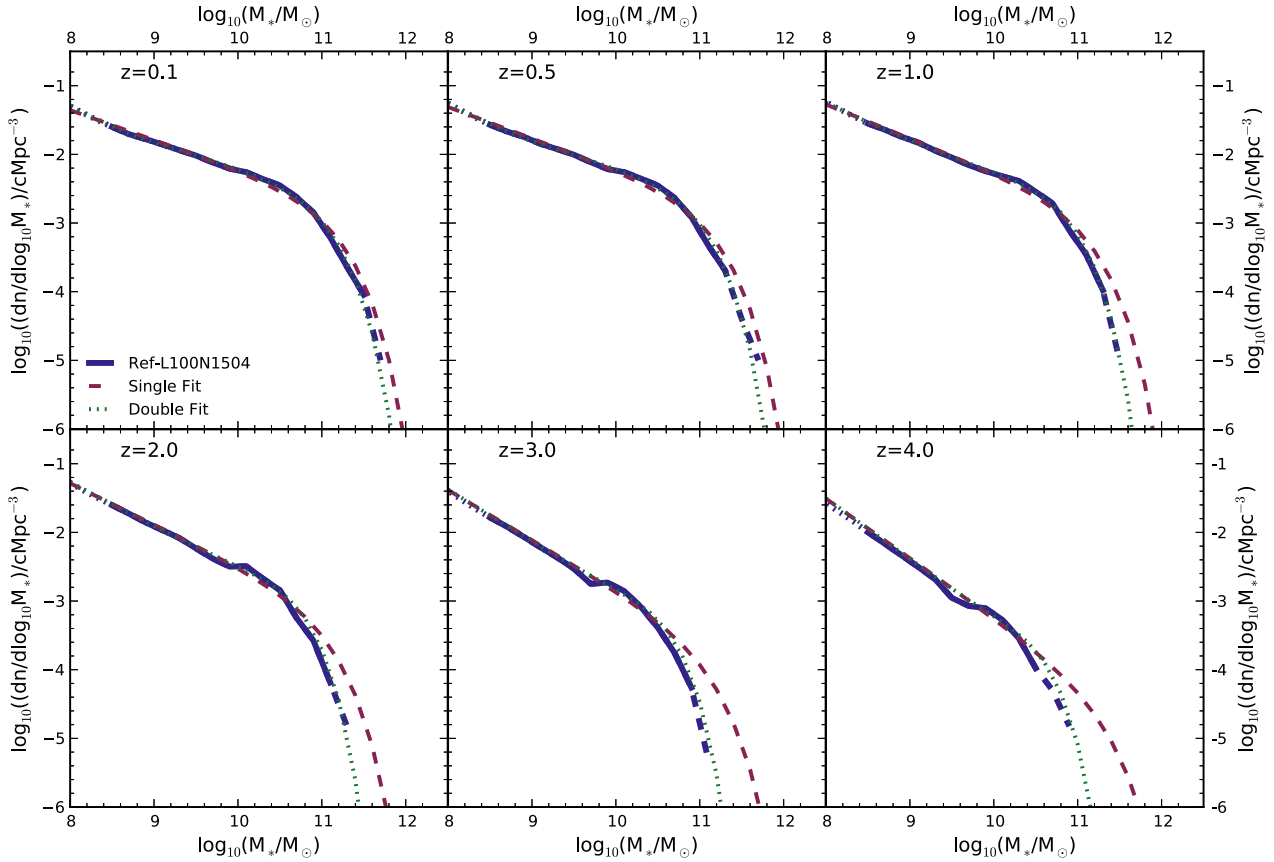
The  $\alpha$  parameter becomes more negative with increasing redshift for fits above  $10^8$  and  $10^9 M_\odot$ , showing that the low-mass slope steepens with redshift. However, different behaviour is seen for fits above  $10^{10} M_\odot$  where  $\alpha$  increases to redshift 1, then decreases. This is not unexpected given that fitting for stellar masses above  $10^{10} M_\odot$  does not provide enough information to constrain the slope for masses  $\ll M_C$ .

We find larger differences between different mass ranges, and in particular larger error bars, when fitting double Schechter functions than what is presented for single Schechter functions in Fig. A2. Due to the sensitivity of the Schechter fitting to the mass range over which it is done, it is very difficult to compare the fitting parameters directly to observations. This is especially true when we consider the evolving mass completeness limit for observations. Any trends with redshift could easily be a result of the changing mass range. The degeneracy between  $M_C$  and  $\Phi^*$  also makes a comparison of Schechter parameters difficult to interpret. The final issue with directly comparing Schechter parameters from observations and/or simulations is the sensitivity of the break in the Schechter function to stellar mass errors, as shown in Section 3.2.1. As a result of these issues, we choose not to compare the Schechter function parameters

**Table A1.** Single (equation 7) and double (equation A1) Schechter function parameters for the *EAGLE* Ref-L100N1504 GSMFs presented in Fig. 2, fitting over the mass range  $10^8$ – $10^{12} M_\odot$ .  $1\sigma$  errors, determined from the covariance matrix, are also listed. The Schechter function parameters provide a simple way of reproducing the GSMFs from the *EAGLE* simulation over the range where the fitting is carried out.

Redshift	$\log_{10}(M_C)$ ( $M_\odot$ )	$\Phi^*$ ( $10^{-3} \text{ cMpc}^{-3}$ )	$\alpha$		
0.1	$11.14 \pm 0.09$	$0.84 \pm 0.13$	$-1.43 \pm 0.01$	–	–
0.5	$11.11 \pm 0.08$	$0.84 \pm 0.12$	$-1.45 \pm 0.01$	–	–
1.0	$11.06 \pm 0.08$	$0.74 \pm 0.10$	$-1.48 \pm 0.01$	–	–
2.0	$10.91 \pm 0.08$	$0.45 \pm 0.07$	$-1.57 \pm 0.01$	–	–
3.0	$10.78 \pm 0.11$	$0.22 \pm 0.05$	$-1.66 \pm 0.01$	–	–
4.0	$10.60 \pm 0.15$	$0.12 \pm 0.04$	$-1.74 \pm 0.02$	–	–
Redshift	$\log_{10}(M_C)$ ( $M_\odot$ )	$\Phi_1^*$ ( $10^{-3} \text{ cMpc}^{-3}$ )	$\alpha_1$	$\Phi_2^*$ ( $10^{-3} \text{ cMpc}^{-3}$ )	$\alpha_2$
0.1	$10.95 \pm 0.03$	$1.45 \pm 0.12$	$-1.31 \pm 0.03$	$0.00 \pm 0.00$	$-2.22 \pm 0.22$
0.5	$10.88 \pm 0.04$	$1.61 \pm 0.13$	$-1.24 \pm 0.08$	$0.08 \pm 0.10$	$-1.79 \pm 0.15$
1.0	$10.74 \pm 0.05$	$1.51 \pm 0.18$	$-0.98 \pm 0.17$	$0.48 \pm 0.20$	$-1.62 \pm 0.05$
2.0	$10.44 \pm 0.08$	$1.06 \pm 0.24$	$-0.25 \pm 0.29$	$0.80 \pm 0.14$	$-1.58 \pm 0.02$
3.0	$10.19 \pm 0.09$	$0.63 \pm 0.13$	$0.23 \pm 0.37$	$0.61 \pm 0.13$	$-1.64 \pm 0.02$
4.0	$10.00 \pm 0.11$	$0.24 \pm 0.08$	$0.89 \pm 0.58$	$0.43 \pm 0.12$	$-1.69 \pm 0.03$





**Figure A1.** The GSMF from the Ref-L100N1504 simulation (blue), using single Schechter fits (red dashed) and double Schechter fits (green dotted) at six redshifts. The parameters for the fitting functions can be found in Table A1.

to those determined observationally and consider the comparison of the data presented in Fig. 2, from which Schechter parameters are derived, to be sufficient to determine the agreement between observations and simulations. However, the Schechter function parameters do provide a simple way of representing the GSMFs from the Eagle simulation over the range where the fitting is carried out.

## APPENDIX B: STRONG NUMERICAL CONVERGENCE

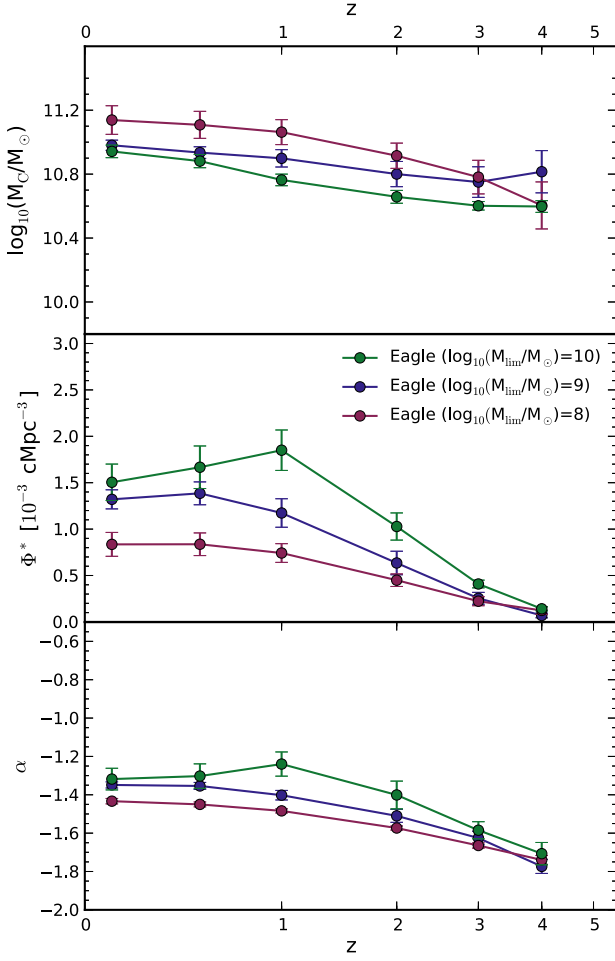
Here, we show strong and weak resolution tests for the evolution of the global stellar mass and SFR densities in Fig. B1. Three models are considered, Ref-L025N0376, equivalent in resolution and model parameters to Ref-L100N1504 except in a 25 cMpc box as opposed to 100 cMpc; Ref-L025N0752, with the same subgrid parameters as Ref-L100N1504 but with eight times higher mass resolution in a 25 cMpc box; and Recal-L025N0752, with recalibrated subgrid parameters and eight times higher mass resolution than Ref-L100N1504 in a 25 cMpc box. The 25 cMpc boxes for which we have higher-resolution simulations are too small to be representative. To ensure we do not obscure the effects of resolution with other effects such as box size, we compare the same box size for all models.

Fig. B1 shows  $\rho_{\text{SFR}}$  for all three 25 cMpc simulations in the top panel. Between redshifts 9 and 5 the Ref-L025N0376 simulation

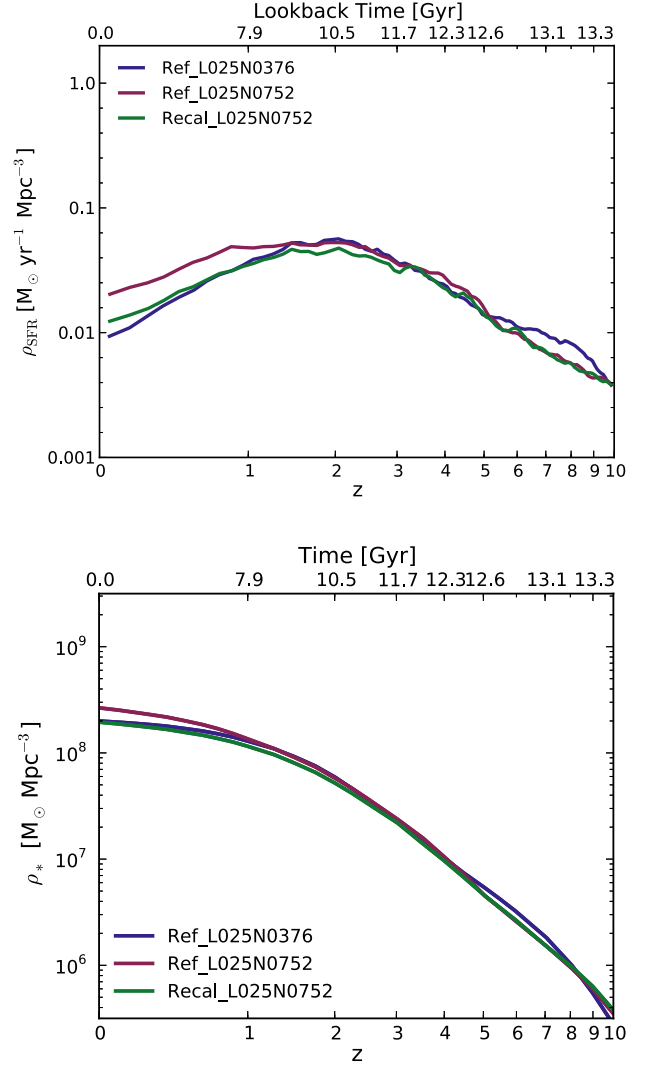
has an excess of star formation relative to both higher-resolution simulations, of less than 0.2 dex, which results from the coarser minimum SFR per particle at the standard resolution. The largest difference between the three simulations is at redshift 0.1, where the Ref-L025N0752 has a higher  $\rho_{\text{SFR}}$  by 0.3 dex. The  $\rho_*$  is shown in the bottom panel of Fig. B1. As  $\rho_*$  is the integral of  $\rho_{\text{SFR}}$  modulo stellar mass-loss, the differences seen here, at redshifts above 4 for Ref-L025N0376 and at redshift 0 for Ref-L025N0752 reflect those seen in  $\rho_{\text{SFR}}$ .

In Fig. B2 again three models are compared, in this case Ref-L100N1504, Recal-L025N0752 and Ref-L025N0752. The agreement between Ref-L100N1504 and Recal-L025N0752, testing weak convergence, is around 0.1 dex at redshift 0.1 over the range of stellar masses that can be probed, as reported in Section 3.2.2. The agreement is similar across all redshift ranges. Comparing Ref-L100N1504 and Ref-L025N0752, to test strong convergence, the stellar mass functions agree to within  $\sim 0.2$  dex at redshift 0.1 and the agreement improves with increasing redshift. At redshifts 4 and above the level of agreement is similar to Recal-L025N0752. In S15, the redshift 0.1 strong convergence was found to be similar to that obtained by simulations from other groups (e.g. Vogelsberger et al. 2013), while the agreement in *EAGLE* improves at higher redshifts.

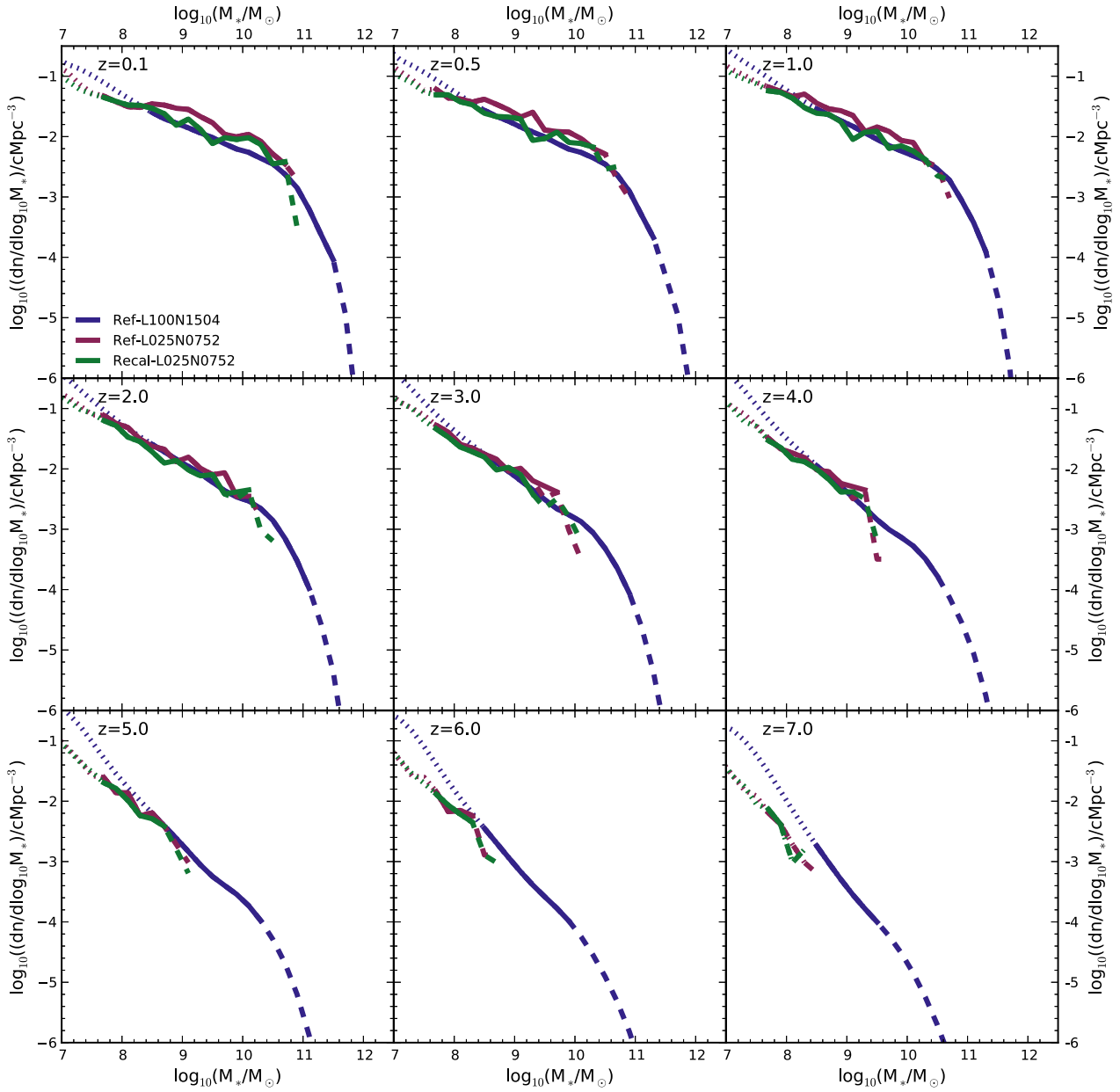
Overall the level of agreement shown for the strong, and particularly for the weak convergence, is good.



**Figure A2.** The Schechter function parameters,  $M_C$ ,  $\Phi^*$  and  $\alpha$  for the *EAGLE* GSMFs (as shown in Fig. 2) as a function of redshift. These panels show single Schechter function parameters fit from  $10^8$ ,  $10^9$  and  $10^{10} M_\odot$  to  $10^{12} M_\odot$  in red, blue and green, respectively, with  $1\sigma$  error bars from the fitting. The Schechter function fitting is sensitive to the mass range over which the fitting is done and the values for both  $M_C$  and  $\Phi^*$  are degenerate. For double Schechter function parameters the agreement between different stellar mass ranges is worse due to the increased freedom (not shown).



**Figure B1.** The SFR density and stellar mass density as a function of redshift in the top and bottom panels for Ref-L025N0376, using the same physics as for Ref-L100N1504 shown in all previous plots, Ref-L025N0752, with eight times the resolution and Recal-L025N0752, with eight times the resolution and recalibrated subgrid parameters in blue, red and green, respectively.



**Figure B2.** The GSMF from Ref-L100N1504 (blue), Ref-L025N0752 (red) and Recal-L025N0752 (green) at nine redshifts. At redshift 0.1, the Recal-L025N0752 simulation is within 0.1 dex of Ref-L100N1504, while the Ref-L025N0752 simulation is within 0.2 dex. The agreement between intermediate and high-resolution simulations improves with increasing redshift.

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