LJMU Research Online

Pfeffer, J, Bastian, N, Kruijssen, JMD, Reina-Campos, M, Crain, RA and Usher, CG

 Young star cluster populations in the E-MOSAICS simulations

http://researchonline.ljmu.ac.uk/id/eprint/11736/

Article

For more information please contact researchonline@ljmu.ac.uk

Young star cluster populations in the E-MOSAICS simulations

Joel Pfeffer [1]★ Nate Bastian,[1] J. M. Diederik Kruijssen [2] Marta Reina-Campos [2]
Robert A. Crain [1] and Christopher Usher [1]

[1]*Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK*
[2]*Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Mönchhofstraße 12-14, D-69120 Heidelberg, Germany*

ABSTRACT
We present an analysis of young star clusters (YSCs) that form in the E-MOSAICS cosmological, hydrodynamical simulations of galaxies and their star cluster populations. Through comparisons with observed YSC populations, this work aims to test models for YSC formation and obtain an insight into the formation processes at work in part of the local galaxy population. We find that the models used in E-MOSAICS for the cluster formation efficiency and high-mass truncation of the initial cluster mass function ($M_{c,*}$) both quantitatively reproduce the observed values of cluster populations in nearby galaxies. At higher redshifts ($z \geq 2$, near the peak of globular cluster formation) we find that, at a constant star formation rate (SFR) surface density, $M_{c,*}$ is larger than at $z = 0$ by a factor of four due to the higher gas fractions in the simulated high-redshift galaxies. Similar processes should be at work in local galaxies, offering a new way to test the models. We find that cluster age distributions may be sensitive to variations in the cluster formation rate (but not SFR) with time, which may significantly affect their use in tests of cluster mass-loss. By comparing simulations with different implementations of cluster formation physics, we find that (even partially) environmentally independent cluster formation is inconsistent with the brightest cluster-SFR and specific luminosity-Σ_{SFR} relations, whereas these observables are reproduced by the fiducial, environmentally varying model. This shows that models in which a constant fraction of stars form in clusters are inconsistent with observations.

Key words: methods: numerical – stars: formation – globular clusters: general – galaxies: evolution – galaxies: formation – galaxies: star clusters: general.

1 INTRODUCTION

Star clusters are a natural by-product of the star formation process (for recent reviews, see Kruijssen 2014; Longmore et al. 2014; Adamo & Bastian 2018; Krumholz, McKee & Bland-Hawthorn 2019). Young star clusters (YSCs) are observed in all star-forming galaxies for which they can be resolved (e.g. Larsen & Richtler 1999); with the resolving power of the *Hubble Space Telescope* they can be detected out to distances of ∼100 Mpc (e.g. Adamo et al. 2010; Fedotov et al. 2011). This observability makes YSCs important tracers of the star formation process in galaxies. The most massive YSCs are also thought to be young analogues of globular clusters (GCs) (Portegies Zwart, McMillan & Gieles 2010; Kruijssen 2015; Forbes et al. 2018), therefore understanding the formation of YSCs may reveal important clues about the formation of GCs.

In recent years, observational studies have established the close connection between the properties of YSC populations and the intensity of star formation in their host galaxies (for a recent review,

see Adamo & Bastian 2018). The fraction of stars formed in bound clusters, i.e. the cluster formation efficiency (CFE or Γ, Bastian 2008), correlates with the star formation rate (SFR) surface density Σ_{SFR} (Goddard, Bastian & Kennicutt 2010; Adamo, Östlin & Zackrisson 2011; Silva-Villa, Adamo & Bastian 2013; Adamo et al. 2015; Johnson et al. 2016). At the low-mass end of their mass range, YSCs are observed to have a power-law mass function with an exponent $\beta \approx -2$ (Zhang & Fall 1999; Bik et al. 2003; Gieles et al. 2006b; McCrady & Graham 2007; Dowell, Buckalew & Tan 2008; Fall & Chandar 2012; Baumgardt et al. 2013). Both of these observations are consistent with being a natural outcome of star formation in a hierarchical gas distribution, with clusters forming in the densest regions of the gas (Elmegreen & Efremov 1997; Efremov & Elmegreen 1998; Elmegreen & Elmegreen 2001; Kruijssen 2012). However, at the high-mass end, evidence suggests that clusters form with a high-mass exponential truncation to the power-law mass function ($M_{c,*}$) that scales with Σ_{SFR} (Gieles et al. 2006a; Gieles 2009; Larsen 2009; Portegies Zwart et al. 2010; Johnson et al. 2017). The observed relation between the magnitude of the brightest cluster in a population ($M_V^{brightest}$) and the SFR of the galaxy (Billett, Hunter & Elmegreen 2002; Larsen

★ E-mail: j.l.pfeffer@ljmu.ac.uk

2002) also implies an upper truncation mass, as it cannot be simply explained by a statistical (size-of-sample) effect with a power-law mass function (Bastian 2008). Instead, the high-mass end of the initial cluster mass function is likely set by a combination of galactic dynamics and stellar feedback (Reina-Campos & Kruijssen 2017).

The dependence of star cluster formation on galactic scale properties means that modelling the formation of realistic star cluster populations also requires modelling the formation and evolution of galaxies and their environment. In part for this reason, simulations of YSC populations have lagged behind the progress of observations. For computational reasons, most works modelling YSC populations focus on isolated or merging galaxies in idealized, non-cosmological simulations (e.g. Li, Mac Low & Klessen 2004; Bournaud, Duc & Emsellem 2008; Kruijssen et al. 2011, 2012; Renaud, Bournaud & Duc 2015; Maji et al. 2017). For the same reason, simulations in a cosmological context also largely focus on high-redshift conditions (e.g. Li et al. 2017; Kim et al. 2018), which does not allow for direct comparisons to present-day galaxies. Moreover, most studies do not investigate *populations* of galaxies, meaning scaling relations between YSC and galaxy properties generally cannot be compared comprehensively to observations.

In this work we investigate the YSC populations in simulations from the MOdelling Star cluster population Assembly In Cosmological Simulations within EAGLE (E-MOSAICS) project (Pfeffer et al. 2018; Kruijssen et al. 2019a). E-MOSAICS couples the MOSAICS model for star cluster formation and evolution (Kruijssen et al. 2011; Pfeffer et al. 2018) to the Evolution and Assembly of GaLaxies and their Environments (EAGLE) galaxy formation model (Crain et al. 2015; Schaye et al. 2015), therefore capturing both the evolution of the galaxies and their environment, as well as the formation and evolution of their star cluster populations. The E-MOSAICS project aims to test the origin and evolution of GC populations within a YSC-based cluster formation scenario (Pfeffer et al. 2018, 2019; Reina-Campos et al. 2018, 2019; Usher et al. 2018) and the use of star clusters as tracers of galaxy formation and assembly (Hughes et al. 2019; Kruijssen et al. 2019a,b). In the fiducial cluster formation model, star cluster populations are fully described by the local, environmentally varying CFE (Kruijssen 2012) and cluster truncation mass (Reina-Campos & Kruijssen 2017). Though the analytical formulations of both models have previously been tested against observations (Kruijssen 2012; Silva-Villa et al. 2013; Adamo et al. 2015; Johnson et al. 2016; Reina-Campos & Kruijssen 2017; Messa et al. 2018b), the local formulation of the models and their coupling to hydrodynamical simulations through the MOSAICS model has not been systematically compared with observed YSC populations. The simulations allow for each component of the model to be switched off, such that their role in the formation of YSC populations, and the variance with galaxy properties, can be assessed. This work also serves as a means to validate the E-MOSAICS cluster formation model and thereby motivate its application to GC populations.

This paper is structured as follows. In Section 2, we briefly summarize the E-MOSAICS simulations and introduce a new set of 12.5 comoving Mpc (cMpc) periodic volumes, for which this paper presents the first results. In Section 3, we present the results from the simulations and compare them to observations, for the CFE-Σ_{SFR} relation (Section 3.1), $M_{\mathrm{c,*}}$-Σ_{SFR} relation (Section 3.2), power-law indices of the mass functions (Section 3.3), specific luminosities (Section 3.4), $M_{\mathrm{V}}^{\mathrm{brightest}}$-SFR relations (Section 3.5), and cluster age distributions (Section 3.6). Finally, we summarize and discuss our conclusions in Section 4.

2 METHODS

In this section we briefly describe the E-MOSAICS model and simulation suite, the selection of simulated galaxies and their star clusters and the methods for analysing the simulations. A full description of the MOSAICS model, the coupling of MOSAICS to the EAGLE model, along with extensive testing of the subgrid models, is given by Pfeffer et al. (2018), and the extension to the full suite of 25 zoom-in simulations is presented in Kruijssen et al. (2019a).

2.1 The E-MOSAICS simulations

The E-MOSAICS project (Pfeffer et al. 2018; Kruijssen et al. 2019a) is a suite of cosmological, hydrodynamical simulations of galaxy formation in the Λ cold dark matter cosmogony that couples the MOSAICS model for star cluster formation and evolution (Kruijssen et al. 2011; Pfeffer et al. 2018) to the EAGLE model of galaxy formation and evolution (Crain et al. 2015; Schaye et al. 2015). The simulations are run with a highly modified version of the *N*-body, smoothed particle hydrodynamics code GADGET3 (last described by Springel 2005). Bound galaxies (subhaloes) were identified within the simulations using the SUBFIND algorithm (Springel et al. 2001; Dolag et al. 2009), in the same manner as in the EAGLE simulations (for details see Schaye et al. 2015). EAGLE includes subgrid routines describing radiative cooling (Wiersma, Schaye & Smith 2009a), star formation (Schaye & Dalla Vecchia 2008), stellar evolution, and mass-loss (Wiersma et al. 2009b), the seeding and growth of black holes (BHs) via gas accretion and BH–BH mergers (Rosas-Guevara et al. 2015), and feedback associated with star formation and BH growth (Booth & Schaye 2009). As current cosmological simulations lack the resolution and physics necessary to compute the feedback efficiencies from first principles, the stellar and active galactic nuclei feedback parameters are calibrated such that the simulations of cosmologically representative volumes reproduce the galaxy stellar mass function, galaxy sizes, and BH masses at $z \approx 0$. The EAGLE simulations successfully reproduce a range of galaxy properties, including the stellar masses (Furlong et al. 2015) and sizes (Furlong et al. 2017) of galaxies, their luminosities and colours (Trayford et al. 2015), their cold gas properties (Lagos et al. 2015, 2016; Bahé et al. 2016; Marasco et al. 2016; Crain et al. 2017), and the properties of circumgalactic and intergalactic absorption systems (Rahmati et al. 2015, 2016; Oppenheimer et al. 2016; Turner et al. 2016, 2017). The simulations also largely reproduce the cosmic SFR density and relation between specific SFR and galaxy mass (Furlong et al. 2015). The simulations are therefore ideal for comparisons with observed galaxy populations.

In the MOSAICS model, star clusters are treated as a subgrid component of star particles in the simulation (Kruijssen et al. 2011). Star clusters are therefore 'attached' to star particles, such that they adopt the properties of the host particle (i.e. positions, velocities, ages, abundances). In a newly formed star particle, a population of star clusters may be formed with properties that depend on the cluster formation model. The model describes cluster formation in terms of two parameters: the CFE (Γ) and the high-mass exponential truncation of the Schechter (1976) cluster mass function $M_{\mathrm{c,*}}$ (with a power-law index of -2 at lower masses). Clusters are drawn from the mass function between masses of 10^2 and $10^8 \, \mathrm{M_\odot}$, while only clusters with masses $> 5 \times 10^3 \, \mathrm{M_\odot}$ are evolved to reduce the memory requirements of the simulations. Each stellar particle forms (statistically) a fraction of its mass in bound clusters (Γ times the particle mass). Thus particles may host clusters more massive

than the stellar mass of the particle, and the CFE and cluster mass function are only well sampled for an ensemble of star particles. However, the total cluster and field star mass is conserved on galactic scales (see Pfeffer et al. 2018 for details).

In the E-MOSAICS suite, we consider four variations of the cluster formation model to assess the importance of each component. The *fiducial* cluster formation model is fully environmentally dependent. The CFE is determined by the local formulation of the Kruijssen (2012) model, which varies as a function of the local natal gas pressure. The mass function truncation mass is determined by the local formulation of the Reina-Campos & Kruijssen (2017) model, where $M_{c,*}$ is related to the local Toomre (1964) mass. The model assumes that $M_{c,*}$ is proportional to the mass of the molecular cloud from which clusters form (Kruijssen 2014). As the simulations do not have the necessary physics and resolution to model molecular clouds, their (sub-grid) masses are calculated by assuming the local Toomre mass sets the maximum mass of molecular clouds, which may further decrease due to the effects of stellar feedback. In the model, the truncation mass generally increases with the natal gas pressure, but decreases in regions with high Coriolis or centrifugal forces (i.e. near the centres of galaxies).

The three other cluster formation model variations then consider environmentally independent versions of the CFE and $M_{c,*}$ (see also Reina-Campos et al. 2019): (i) a constant CFE of $\Gamma = 0.1$ with a pure power-law mass function (i.e. $M_{c,*} = \infty$; *no formation physics* model); (ii) an environmentally varying CFE with $M_{c,*} = \infty$ (*CFE only* model); (iii) an environmentally dependent $M_{c,*}$ with $\Gamma = 0.1$ (*$M_{c,*}$ only* model).

The simulations model several channels of mass-loss for star clusters, namely stellar evolution, two-body relaxation, tidal shock driven mass-loss, and complete disruption by dynamical friction (for details, see Kruijssen et al. 2011; Pfeffer et al. 2018). Stellar evolutionary mass-loss for clusters is proportional to that of the host stellar particle, calculated in the EAGLE model (Wiersma et al. 2009b). The mass-loss rate from two-body relaxation is determined by the strength of the local tidal field, which is calculated via the eigenvalues of the tidal tensor at the location of the star particle. The tidal shock mass-loss is also calculated via the tidal tensors, based on the derivations of Gnedin, Hernquist & Ostriker (1999) and Prieto & Gnedin (2008). Star clusters that reach a mass below $100\,\mathrm{M_\odot}$ are assumed to be fully disrupted. Additionally, the removal of star clusters due to dynamical friction is treated in post-processing and applied at every snapshot (though this mechanism is mainly important for massive, old clusters and has little effect on young cluster populations).

In this work, we use the 25 zoom-in simulations focused on Milky Way-mass haloes (Pfeffer et al. 2018; Kruijssen et al. 2019a) and a new set of E-MOSAICS simulations of a 12.5 cMpc periodic volume (L012N0376). All simulations were performed using a Planck Collaboration XVI (2014) cosmology, the 'recalibrated' EAGLE model (see Schaye et al. 2015) and initial baryonic particle masses of $\approx 2.25 \times 10^5\,\mathrm{M_\odot}$. The 25 zoom-in simulations of Milky Way-mass haloes ($M_{\rm vir} \approx 10^{12}\,\mathrm{M_\odot}$) were drawn from the high-resolution EAGLE simulation of a 25 cMpc volume (Recal-L025N0752) and resimulated in a zoom-in fashion with the E-MOSAICS model (see Pfeffer et al. 2018; Kruijssen et al. 2019a). Ten of these zoom-in simulations were run with all four MOSAICS model variations, while the other 15 were run only with the fiducial model. To increase the range of galaxy types and environments (mainly for galaxies with stellar masses $< 10^{10}\,\mathrm{M_\odot}$), we also performed (for all four model variations) new simulations of a periodic volume with side length $L = 12.5$ cMpc, using 2×376^3

particles (i.e. the EAGLE Recal-L012N0376 volume). The E-MOSAICS L012N0376 volume, and six example galaxies from the fiducial cluster formation model, are visualized in Fig. 1. In general, cluster formation is biased towards the centres of the galaxies ($\lesssim 10$ kpc for Milky Way-mass galaxies), in regions where the natal gas densities of star formation are highest.

2.2 Galaxy and star cluster selection

We select galaxies from bound subhaloes, including both central and satellite galaxies in a halo, from both the periodic volume and zoom-in simulations at redshifts $z = 0, 0.5, 1$, and 2. For the zoom-in simulations, we only consider galaxies that fall within the high-resolution region of the simulations (those that reside in haloes with <0.1 per cent contamination by low-resolution particles at any snapshot).

Galaxies (and their bound particles) are then selected from the simulations for analysis in the following way.

(i) First, we select star particles within a 30 kpc 3D radius (as for Schaye et al. 2015) from the centre of potential of the galaxy (i.e. the position of the most bound particle in the subhalo). This focuses the particle selection on the main galaxy and helps to exclude particles being stripped from merging satellites. Galaxies must have at least 20 star particles within this region, giving a minimum resolved stellar mass of $\approx 4 \times 10^6\,\mathrm{M_\odot}$.

(ii) Next, galaxies are limited to having at least 10 young (< 300 Myr) star particles within a projected radius $R_{\rm lim}$, where the galaxy is projected such that the disc is face on (using the angular momentum of the star particles). This selection imposes a minimum total stellar mass of $> 10^7\,\mathrm{M_\odot}$ at $z = 0$ and $> 4 \times 10^6\,\mathrm{M_\odot}$ at $z = 2$. We calculate $R_{\rm lim}$ as the minimum of $1.5R_{1/2}$ (the projected half-mass radius) and the radius containing 68 per cent of the recent (< 300 Myr) star formation in the galaxy. These selections are made in order to approximate the typical footprints for observations of nearby star-forming galaxies (e.g. Adamo et al. 2015; Messa et al. 2018a) and to limit the projected region such that area-averaged quantities (e.g. $\Sigma_{\rm SFR}$) are focused on the main star-forming component of each galaxy, respectively. The latter selection is important in galaxies with very centrally concentrated star formation. Note that, because of the scale-free nature of the interstellar medium (ISM) and star formation (e.g. Elmegreen & Falgarone 1996; Elmegreen 2002), there is no standard definition for the star-forming area of galaxies (see also Kruijssen & Bastian 2016, for a discussion on appropriate areas). Due to the heterogeneous coverage of observed galaxies, it is not possible to match the observational footprints directly (e.g. see fig. 1 of Larsen 2002).

(iii) Finally, we select star-forming galaxies based on their specific SFR (sSFR) measured within $1.5R_{1/2}$. Following Bourne et al. (2017), we use their equation (6) to define the star-forming galaxy sequence as a function of redshift, but set a constant sSFR for galaxies with $M_* \leq 10^{10}\,\mathrm{M_\odot}$ and apply a vertical shift to lower sSFRs (by setting $b_0 = -10.2$ and $b_1 = 2.3$ in their equation 6) to match the EAGLE main sequence (which predicts slightly lower sSFRs than observed, see Furlong et al. 2015). We then select galaxies with sSFRs that do not fall more than 0.5 dex below the sequence. At $z = 0$ and stellar masses of $M_* \leq 10^{10}\,\mathrm{M_\odot}$, this selects galaxies with $\mathrm{SFR}/M_* > 4 \times 10^{-10}\,\mathrm{yr^{-1}}$.

Star clusters are selected in galaxies following the same criteria as for the star particles to which they are attached. With the exception of the CFE (which is calculated from the total initial mass in clusters) and when fitting initial cluster mass functions,
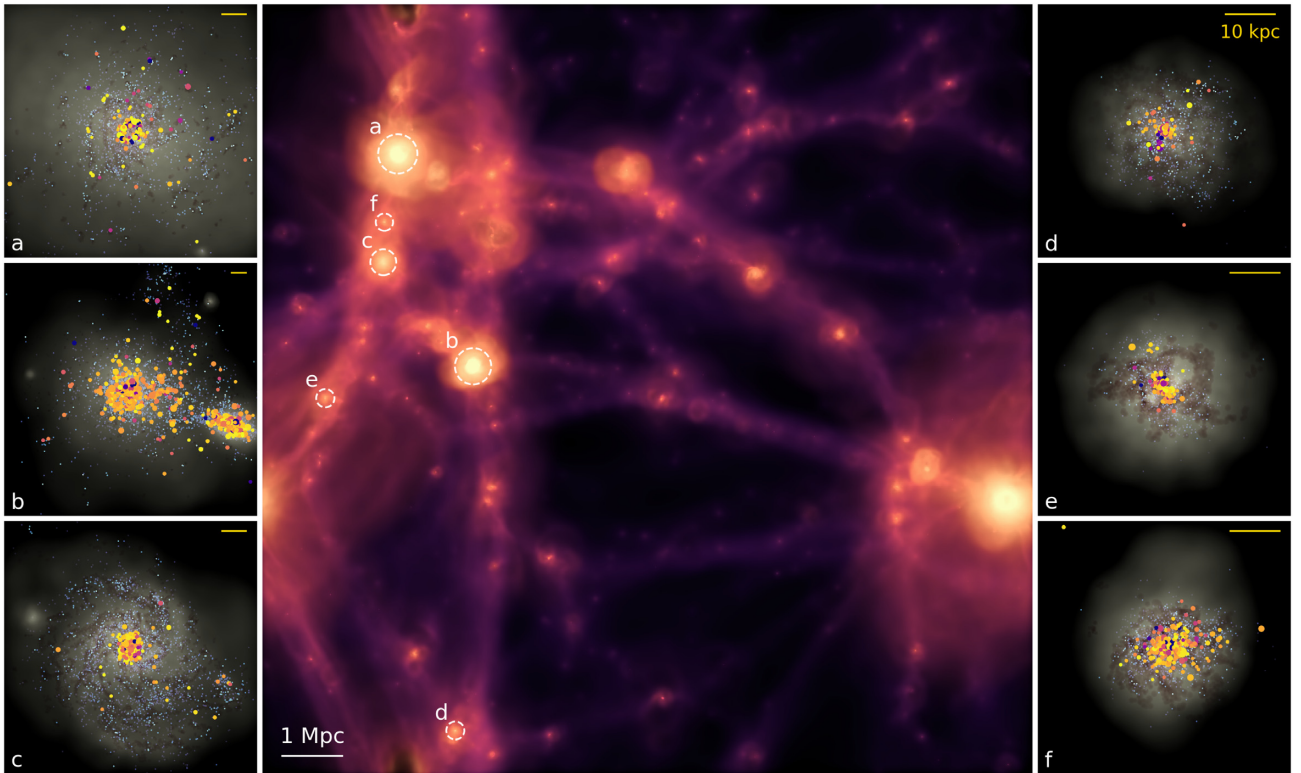
Figure 1. Visualization of the E-MOSAICS L012N0376 simulation at $z = 0$. The main panel shows the gas surface density, coloured by temperature, for the entire volume. The side panels show three approximately Milky Way-mass ($M_* \approx 1.5\text{-}6 \times 10^{10}\,M_\odot$; left-hand side) and Large Magellanic Cloud-mass galaxies ($M_* \approx 5 \times 10^9\,M_\odot$; right-hand side). The side panels show mock optical images of the galaxies (grey-scale shows stellar density, small light blue points show young star particles, brown shows dense star-forming gas; rotated such that the discs are face on) and the location of young clusters (age < 300 Myr, initial masses > $5 \times 10^3\,M_\odot$), where symbol colours show cluster age (with dark blue to yellow colours spanning the age range 10^7–$10^{8.5}$ yr) and symbol areas are proportional to cluster mass. The locations of the galaxies shown in the side panels are indicated in the main panel with dashed circles, where the radii of the circles show the virial radii (r_{200}) of the galaxies. The side panels show regions with side lengths of $L = 100$ kpc (left-hand panels) and 50 kpc (right-hand panels), with the exception of panel (b), which shows $L = 160$ kpc, as the galaxy is undergoing a major merger (stellar masses of 2.8×10^{10} and $1.3 \times 10^{10}\,M_\odot$). Scale bars in the upper right corner of the side panels indicate a length of 10 kpc.

we apply a mass limit for evolved clusters of $M > 5 \times 10^3\,M_\odot$. Though this limit is necessary due to instantaneous disruption of low-mass clusters in the simulations, it is consistent with those imposed in observations of YSCs in nearby galaxies (depending on distance to the galaxy and the upper age limit for clusters; e.g. Annibali et al. 2011; Adamo et al. 2015; Johnson et al. 2017; Messa et al. 2018a; Cook et al. 2019).

For the $z = 0$ snapshot, these selection criteria give us a sample of 153 galaxies with stellar masses between 2×10^7 and $4 \times 10^{10}\,M_\odot$ (median $3.7 \times 10^8\,M_\odot$), SFRs between 8×10^{-3} and $3\,M_\odot\,yr^{-1}$ (median $0.04\,M_\odot\,yr^{-1}$) and Σ_{SFR} between 10^{-4} and $0.3\,M_\odot\,yr^{-1}\,kpc^{-2}$ (median $2 \times 10^{-3}\,M_\odot\,yr^{-1}\,kpc^{-2}$). Of these, 39 galaxies have >50 YSCs (ages < 300 Myr and initial masses > $5 \times 10^3\,M_\odot$) within R_{lim}.

2.3 Analysis

All cluster and SFR-related quantities (SFR, Σ_{SFR}) are calculated for clusters and star particles with ages < 300 Myr at the time of the relevant snapshot, with the exception of Section 3.5 (which investigates the $M_V^{brightest}$–SFR relation) and Section 3.6 (which investigates cluster age distributions). For clusters, this age limit is similar to observational studies for which YSC populations are typically only complete (in mass) below ages of a few hundred megayears

(depending on the mass limit). SFRs for the simulated galaxies are calculated directly from the mass in star particles formed over this time period. Observational studies often use SFR tracers (e.g. $H\alpha$ or UV flux, stellar counts) which are sensitive to time-scales $\lesssim 100$ Myr (Kennicutt & Evans 2012; Haydon et al. 2018).

Projected galaxy quantities (Σ_{SFR}, Σ_{gas}, Σ_*) are calculated within the surface area given by the projected radius R_{lim} (i.e. area-weighted surface densities). This method follows most observational studies which use the same procedure, though it remains sensitive to the region over which the properties are measured (e.g. particularly if star formation is highly concentrated or substructured; see also Johnson et al. 2016, who apply a mass-weighted method, and Appendix A).

In Section 3.5 we compare YSC properties against the SFR and sSFR of the galaxy. For this comparison, we calculate all properties within $1.5R_{1/2}$ so as not to bias the sSFR measurement for cases with a very small R_{lim}.

In Sections 3.2 and 3.3 we fit Schechter (1976) mass functions to the simulated YSC populations. We follow a similar analysis to that used in observational studies (e.g. Johnson et al. 2017; Messa et al. 2018b). For each population, we use the Markov chain Monte Carlo code PYMC (Fonnesbeck et al. 2015) to sample the posterior probability distribution of the Schechter truncation mass. For each population we sample the truncation mass in log -space with a uniform prior between $5 \times 10^3\,M_\odot$ and $10^8\,M_\odot$ (the lowest

mass YSC we consider and ∼30 times the mass of the most massive YSC at $z = 0$ in our study, respectively). When fitting for the initial cluster masses, we fix the power-law index of the mass function to $\beta = -2$, the input index in the cluster models. When fitting for the final (evolved) masses of clusters, we allow the power-law index to vary, and sample the index with a uniform prior between -3 and -0.5. For each population, we perform 10 000 iterations with 1000 burn-in steps.

For Sections 3.4 and 3.5, Johnson U and V-band luminosities were calculated for clusters assuming simple stellar populations using the clusters' age, metallicity, and mass in combination with the Flexible Stellar Population Synthesis (FSPS) model (Conroy, Gunn & White 2009; Conroy & Gunn 2010), using the MILES spectral library (Sánchez-Blázquez et al. 2006), Padova isochrones (Girardi et al. 2000; Marigo & Girardi 2007; Marigo et al. 2008), a Chabrier (2003) initial stellar mass function and assuming the default FSPS parameters. Mass-to-light ratios for the clusters were calculated by linearly interpolating the luminosities and relative stellar masses from the grid in ages and total metallicities $\log(Z/Z_\odot)$. Note that we do not include extinction in these estimates, as most observational studies correct for this effect.

3 RESULTS

3.1 Cluster formation efficiency

In Fig. 2, we first test the CFE-Σ_{SFR} relation in the fiducial cluster formation model for stars and clusters younger than 300 Myr at $z = 0$. Note that this is not strictly a prediction of the simulations, since the Kruijssen (2012) model, for which the galaxy-scale version has previously been tested against observations, was adopted for the CFE model in the simulations. However, the section serves as a validation of the implementation in terms of local variables within the E-MOSAICS model and its extension to galaxy-wide scales within the simulations, and enables the testing of effects that may induce scatter in the relation. To calculate the CFE in the simulations, we sum the total mass of clusters formed[1] in the star particles in the region of interest (i.e. the CFE at formation which includes any stochasticity in sampling initial cluster masses from the mass function, not the values calculated at the particle level from the natal gas pressure). This value does not include the effect of cluster mass-loss (which will lower the measured value of the CFE) or any observational uncertainties associated with cluster detection and measurement of the SFR. For the simulated galaxies, Σ_{SFR} is also measured over the same 300 Myr time-scale as the CFE. Our results are not sensitive to the exact age limit used, as we find consistent results[2] when using ranges of 1–10 Myr, 1–100 Myr, and 10–100 Myr.

Fig. 2 shows the galaxy-averaged CFE for all galaxies in the simulations with the fiducial cluster formation model. For comparison, in the figure we also include observed CFEs from Goddard et al. (2010), Adamo et al. (2011), Annibali et al. (2011), Silva-Villa & Larsen (2011), Cook et al. (2012, using the results for ages < 100 Myr and excluding galaxies with only upper limits for the CFE), Ryon et al. (2014), Adamo et al. (2015), Hollyhead et al.
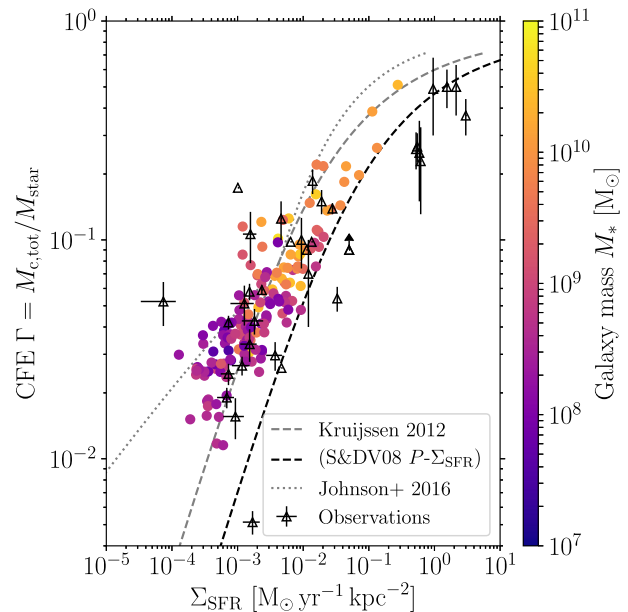

Figure 2. CFE as a function of the star formation rate surface density (Σ_{SFR}) of the galaxy. For the simulations, each point shows the result for stars and star clusters younger than 300 Myr at $z = 0$ for the fiducial cluster formation model. The points are coloured by the stellar mass of the galaxy. The open triangles show the CFEs for observed galaxies (see the text). The grey dashed line shows the fiducial prediction of the Kruijssen (2012) model (where Σ_{SFR} has been decreased by a factor of 1.65 to convert from a Salpeter 1955 to Chabrier 2003 initial stellar mass function for the Kennicutt–Schmidt relation), while the black dashed line shows the same relation shifted to match the pressure–Σ_{SFR} relation adopted in the EAGLE simulations (Schaye & Dalla Vecchia 2008; Schaye et al. 2015). The grey dotted line shows the same model but assuming a Σ_{gas}–Σ_{SFR} relation based on the Bigiel et al. (2008) observations (Johnson et al. 2016).

(2016), Johnson et al. (2016), Ginsburg & Kruijssen (2018), and Messa et al. (2018b). Note that the CFE for NGC 4449 is a lower limit (Annibali et al. 2011). The simulations show a similar level of scatter in the CFE at a given Σ_{SFR} (∼0.25 dex) to measurements of observed galaxies. In the figure there is a galaxy mass gradient along the CFE–Σ_{SFR} relation, such that the more massive galaxies generally have a higher Σ_{SFR} and CFE. This result is expected as, assuming pressure equilibrium in the galaxies, larger galaxy masses (therefore deeper potentials) result in higher characteristic ISM pressures, and thus higher Σ_{SFR} and CFE. However, it is important to note that the volumes of the simulations are not large enough to capture rare objects, such as rapidly star-forming dwarf galaxies with high star-formation densities and CFEs (e.g. blue compact dwarfs; Adamo et al. 2011). Additionally, the selection in sSFR for star-forming galaxies (Section 2.2) selects against high-mass galaxies with low star-formation densities and CFEs.

For the projected version of the Kruijssen (2012) model (dashed lines in the figure), the gas surface density (Σ_{gas}, which is the fundamental quantity in the model setting the CFE, see e.g. Kruijssen & Bastian 2016 and Ginsburg & Kruijssen 2018) is converted to an SFR surface density assuming the Kennicutt–Schmidt star formation relation (Kennicutt 1998).[3] Additionally, star formation

---

[1] For a small number of particles where $M_{c,*} < 100 M_\odot$ (less than the lower mass for the cluster mass function) we assume that no clusters were formed.
[2] Though with significantly fewer galaxies in the 1–10 Myr age range, since this range is generally poorly resolved due to the resolution limits of the simulations.

[3] Note that for consistency with the simulations, we adopt $\Sigma_{SFR} = 1.515 \times 10^{-4} M_\odot \, yr^{-1} \, kpc^{-2} (\Sigma_{gas}/1 \, M_\odot \, kpc^{-2})^{1.4}$, consistent with a Chabrier (2003) initial stellar mass function (see Schaye et al. 2015).
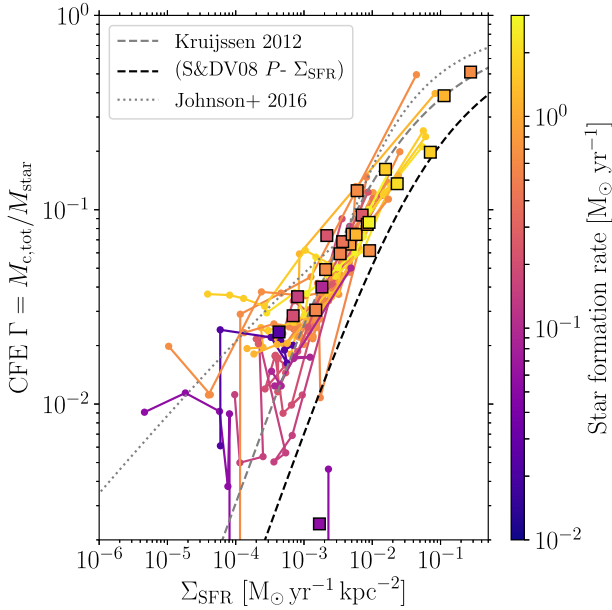
Figure 3. Radial CFE distributions (solid lines) in 2 kpc annuli, to a maximum of 16 kpc, for the target L^* galaxies in 23 of the 25 zoom-in simulations. Each galaxy in the figure is coloured by its star formation rate. The squares show the values of the same galaxies measured within $R_{\rm lim}$ (as in Fig. 2). The dashed and dotted lines show the same relations as in Fig. 2.

in EAGLE is implemented following the Kennicutt–Schmidt relation, rewritten as a pressure law (Schaye et al. 2015). Therefore, the naive expectation is that the simulations should broadly reproduce the (shifted) Kruijssen (2012) relation in Fig. 2 (black dashed line), where the relation has been shifted to higher $\Sigma_{\rm SFR}$ by ≈ 0.6 dex to account for the change from the Krumholz & McKee (2005) P–$\Sigma_{\rm SFR}$ relation to the (Schaye & Dalla Vecchia 2008) relation used in EAGLE. At $\Sigma_{\rm SFR} \gtrsim 5 \times 10^{-3}\,{\rm M_\odot\,yr^{-1}\,kpc^{-2}}$, the simulated galaxies broadly follow the expected relation (black dashed line), but are generally shifted to slightly lower $\Sigma_{\rm SFR}$, with most galaxies falling between the fiducial (grey dashed line) and shifted (black dashed line) Kruijssen (2012) relations. For the modelling in this work, the fundamental relation is between the CFE and natal gas pressure, and therefore some amount of uncertainty in the CFE–$\Sigma_{\rm SFR}$ relation arises simply due to the uncertainty in the P–$\Sigma_{\rm SFR}$ relation. We discuss this point in further detail in Appendix A, where we show that the offset from the expected CFE–$\Sigma_{\rm SFR}$ relation is due to galaxies being offset from the expected P–$\Sigma_{\rm SFR}$ (see also the discussion below).

At lower surface densities ($\Sigma_{\rm SFR} \lesssim 5 \times 10^{-3}\,{\rm M_\odot\,yr^{-1}\,kpc^{-2}}$), the simulations show a higher CFE at a given $\Sigma_{\rm SFR}$ than the Kruijssen (2012) relation, for which the cause may be multifold. First, this can be caused by highly concentrated or substructured star formation, such that star and cluster formation largely occurs in a much smaller area compared to the area for which $\Sigma_{\rm SFR}$ is calculated, which will lower the measurement for $\Sigma_{\rm SFR}$ at a given CFE. Secondly, at low $\Sigma_{\rm SFR}$ there is a physical effect that increases the CFE at lower metallicities (i.e. in lower mass galaxies) due to the metallicity-dependent density threshold for star formation implemented in EAGLE.[4] This threshold is included to model the

[4]Note that this effect of increasing CFE with metallicity is not expected to occur at high $\Sigma_{\rm SFR}$, since the densities of star-forming gas in this regime are well above the metallicity-dependent threshold.

effect of the thermogravitational collapse of warm, photoionized interstellar gas into a cold, dense phase, which is expected to occur at lower densities and pressures in metal-rich gas (Schaye 2004). This higher density threshold at lower metallicities results (through the lower density limit for star formation imposed by the polytropic equation-of-state implemented at high gas densities) in higher pressures of star formation at a given $\Sigma_{\rm SFR}$, and therefore in higher CFEs (see fig. 3 in Pfeffer et al. 2018). Finally, variations in the natal pressure–$\Sigma_{\rm SFR}$ relation in the galaxies will, in turn, lead to variations in the CFE–$\Sigma_{\rm SFR}$ relation, through the dependence of the CFE on natal gas pressure. Such variations can be driven by random fluctuations within the galaxies (which may be most important at low $\Sigma_{\rm SFR}$), or physical variations due to differing contributions of the gravity of stars to the mid-plane gas pressure in galaxies (i.e. variations in ϕ_P, see also Appendix C). We test the importance of these effects in Appendix A, finding the dominant effect to be the use of too large an area in the calculation for $\Sigma_{\rm SFR}$ (i.e. $\Sigma_{\rm SFR}$ is systematically underestimated). This effect may be mitigated by calculating a mass-weighted surface density (see Johnson et al. 2016, and Appendix A), or by judicious aperture choice, focussing on the main region of star formation. Since most studies apply the standard area-weighted calculations, we focus on that method in this paper.

The variation of the CFE at a given $\Sigma_{\rm SFR}$ can be further investigated by comparing the radial CFE distributions within the galaxies. In Fig. 3 we show the radial CFE distributions in 2 kpc annuli in 23 of the 25 L^* galaxies (Milky Way-mass haloes; MW16 and MW22 are quenched and do not have young clusters, thus are excluded from the figure) from the zoom-in simulations (Kruijssen et al. 2019a). For this figure, we have not applied the limit on sSFR for the galaxies (Section 2.2) in order to sample a wide range of galactic environments. The majority of measurements in the radial distributions fall along the Kruijssen (2012) relation (as expected), and galaxies with higher SFRs generally show higher CFE and $\Sigma_{\rm SFR}$. However at low $\Sigma_{\rm SFR}$ ($\lesssim 10^{-3}\,{\rm M_\odot\,yr^{-1}\,kpc^{-2}}$), the simulations again show significant scatter from the Kruijssen (2012) relation, with most points falling between the fiducial relation and the reinterpreted relation from Johnson et al. (2016, which uses the Kruijssen model, but assumes a $\Sigma_{\rm gas}$–$\Sigma_{\rm SFR}$ relation based on the observations of Bigiel et al. 2008, rather than the Kennicutt–Schmidt relation). This deviation can be attributed to $\Sigma_{\rm SFR}$ being averaged over a larger area than for which star and cluster formation is occurring and variations in the natal pressure–$\Sigma_{\rm SFR}$ relation, since the natal pressure is approximately constant at a given CFE (see also Appendix A). Similarly, $\Sigma_{\rm SFR}$ for the innermost radial bin in MW13 (at $\Gamma \approx 0.5$) deviates significantly from both the 'global' value (square symbol) due to very central star formation that dominates the cluster formation in the galaxy (for this galaxy $R_{\rm lim} = 0.75$ kpc). One galaxy, MW05, has a CFE that is significantly below other galaxies at $\Sigma_{\rm SFR} \approx 2 \times 10^{-3}\,{\rm M_\odot\,yr^{-1}\,kpc^{-2}}$. The galaxy has a very low median cluster truncation mass at $z = 0$ of $M_{\rm c,*} \approx 100\,{\rm M_\odot}$, meaning many star particles with $M_{\rm c,*} < 100\,{\rm M_\odot}$ form no clusters. This is caused by the low natal gas pressure for star formation ($P/k < 10^4\,{\rm K\,cm^{-3}}$) and a high stellar density (high ϕ_P, see the discussion in Section 3.2) in the galaxy at that epoch.

3.2 Mass function truncation

In this section we test the model for the upper exponential truncation of the cluster mass function ($M_{\rm c,*}$) implemented in the MOSAICS models (Reina-Campos & Kruijssen 2017). As for the CFE, this is not strictly an independent prediction since the galaxy-scale

version of the model has previously been tested against observations (Reina-Campos & Kruijssen 2017; Messa et al. 2018b), but it serves as a test and validation of the implementation in terms of local variables within the E-MOSAICS model. However, we also provide predictions at high redshifts which may be tested with future observations.

Using the fitting procedure described in Section 2.3, for each galaxy with >50 clusters we fit a Schechter (1976) mass function with an upper exponential truncation mass $M_{c,*}$ to the initial masses of clusters younger than 300 Myr, using a fixed low-mass index of $\beta = -2$ (i.e. the input value used in the simulations). In Appendix B, we compare the resulting $M_{c,*}$ for fitting initial cluster mass functions (with a fixed power-law index) and final (evolved) cluster mass functions (with a variable power-law index). We find that both methods generally give consistent measurements for $M_{c,*}$, with potentially a small offset to higher initial $M_{c,*}$ due to stellar evolutionary mass-loss (a factor of ~0.1 dex).

Following the fit, we exclude galaxies for which $M_{c,*}$ is larger than the most massive cluster in the population. In such cases, $M_{c,*}$ is poorly constrained since cluster formation does not fully sample up to the truncation mass. For galaxies with $M_{c,*} < \max(M)$, fits typically have 1σ uncertainties of <0.5 dex. For $M_{c,*} > \max(M)$ uncertainties reach up to ~2 dex, even for populations with >100 clusters. Due to the initial cluster mass limit ($5 \times 10^3\,\mathrm{M_\odot}$), truncation masses are typically only able to be fit above masses of a few times $10^4\,\mathrm{M_\odot}$. This limit also biases the results to galaxies ($M_* \gtrsim 10^9\,\mathrm{M_\odot}$ at $z = 0$) which have a large enough population of YSCs above the mass limit to fit a mass function. At $z = \{0, 0.5, 1, 2\}$ this leaves us with a sample of $\{33, 51, 65, 60\}$ galaxies.

In Fig. 4 we compare the fitted $M_{c,*}$ for the simulated galaxies at $z = 0$ with results from observed nearby galaxies (described in caption). The predicted $M_{c,*}$ for the simulated galaxies are in good agreement with the observed galaxies, falling about the relation described by the observations (Johnson et al. 2017) over the same range in Σ_{SFR}. More observations are needed to test whether the scatter found in $M_{c,*}$ for the simulated YSCs is consistent with observed YSC populations, which is possible with (e.g.) the LEGUS survey (Calzetti et al. 2015) and the PHANGS-*HST* survey (Lee et al., in preparation).

Overall, the Reina-Campos & Kruijssen (2017) model for $M_{c,*}$, and its implementation in terms of local gas and dynamical properties in the E-MOSAICS cluster formation model, performs well in reproducing YSC populations in realistic galaxy formation simulations at $z = 0$. Therefore we can be confident in extending the formation model to more extreme environments, such as to the epochs of GC formation. In Fig. 5 we perform the same comparison of $M_{c,*}$ against Σ_{SFR} for simulated galaxies at redshifts of $z = \{0, 0.5, 1, 2\}$. For reference, we also show the measurements from observed galaxies at $z = 0$ in each panel. At a given Σ_{SFR} or $M_{c,*}$, the typical stellar mass of galaxies forming clusters decreases with increasing redshift, implying that galaxies of a given stellar mass (at that epoch) can form higher mass clusters in the early universe compared to $z = 0$.

At $z = 0.5$, the $M_{c,*}$–Σ_{SFR} distribution is similar to $z = 0$. Due to the larger sample of galaxies at this snapshot, the distribution extends to higher $M_{c,*}$ and Σ_{SFR}, comparable with that found for the Antennae galaxies ($M_{c,*} \sim 2 \times 10^6\,\mathrm{M_\odot}$; Zhang & Fall 1999; Jordán et al. 2007). In fact, the simulated galaxy at $z = 0.5$ closest to the Antennae measurement is one of a pair of galaxies about to undergo a major merger (with stellar masses $\sim 10^{10}\,\mathrm{M_\odot}$), which are separated by $\approx 10\,\mathrm{kpc}$ at the time of the snapshot. Due to the infrequency of the snapshots (we output 29 between
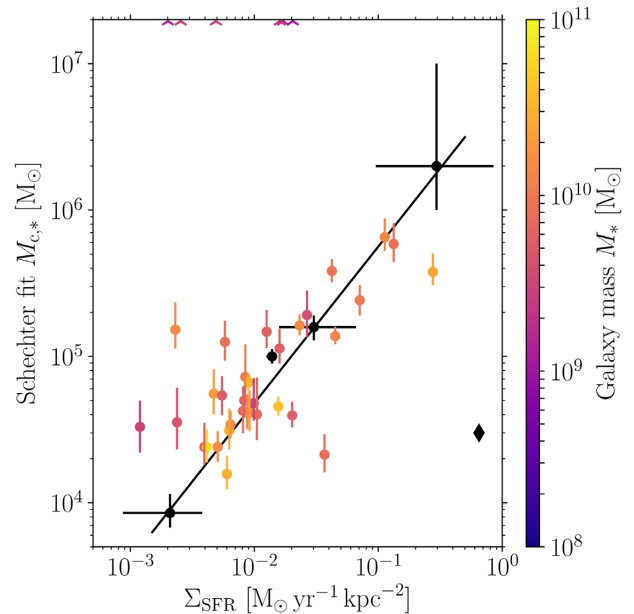

Figure 4. Mass function truncation $M_{c,*}$ as a function of Σ_{SFR} for the simulated galaxies with >50 clusters younger than 300 Myr at $z = 0$. The points show the median of the posterior distribution from the MCMC fit, coloured by the galaxy stellar mass, while errorbars show the 16 and 84 per cent confidence intervals. The arrows at the top of the figure indicate galaxies for which $M_{c,*}$ was unable to be constrained. The black points show the fits to observed cluster populations in the Antennae system (Zhang & Fall 1999; Jordán et al. 2007), M31 (Johnson et al. 2017), M51 (Gieles 2009; Messa et al. 2018b), and M83 (Adamo et al. 2015). The solid line shows the relation fit to the observations by Johnson et al. (2017). The black diamond shows the best-fitting truncation mass of YSCs in the Central Molecular Zone (CMZ) of the Milky Way (Trujillo-Gomez, Reina-Campos & Kruijssen 2019) versus its Σ_{SFR} (Barnes et al. 2017), demonstrating that the empirical relation from Johnson et al. (2017) is not fundamental, but must have an additional dependence, most likely on the epicyclic frequency as in Reina-Campos & Kruijssen (2017). This decreases $M_{c,*}$ towards galactic centres.

$z = 20$ and $z = 0$), catching a galaxy merger during its peak is extremely unlikely. At higher redshifts of $z = \{1, 2\}$, near the peak of GC formation for L^* galaxies in the E-MOSAICS model ($z \sim 1$–4; Reina-Campos et al. 2019), we find in the simulated galaxies that $M_{c,*}$ is higher at a given Σ_{SFR} than the relation at $z = 0$. Therefore, to reach a given $M_{c,*}$, galaxies require a lower Σ_{SFR} in the early universe compared to today (by ~0.5 dex at $z = 2$).

This increase in $M_{c,*}$ is caused by two effects. At late times, a higher contribution of the gravity of stars to the mid-plane gas pressure (i.e. ϕ_P, Elmegreen 1989; Krumholz & McKee 2005; equations 7 and 8 in Pfeffer et al. 2018) results in a lower gas surface density (and therefore Toomre mass) at a given pressure. Additionally, the density threshold for star formation increases with decreasing metallicity in the EAGLE model (which mainly has an effect at $\Sigma_{\mathrm{SFR}} < 10^{-2}\,\mathrm{M_\odot\,yr^{-1}\,kpc^{-2}}$, see the discussion in Section 3.1), thus resulting in a higher $M_{c,*}$. We further discuss the effect of ϕ_P in Appendix C, finding that the ϕ_P increases from $\phi_P \approx 1$ at $z = 2$ to $\phi_P \approx 2.5$ at $z = 0$. In Fig. 5, we show the effect of decreasing ϕ_P at higher redshifts by rescaling the fit to observed local galaxies at $z = 0$ assuming the median ϕ_P from the simulations at each redshift (dashed line). The simulated galaxies agree well with the rescaled relations at each redshift, demonstrating the effect of ϕ_P on $M_{c,*}$. Note that as $\phi_P \approx 1$ at $z = 2$ (right-hand panel in Fig. 5), galaxies at $z > 2$ should simply follow the
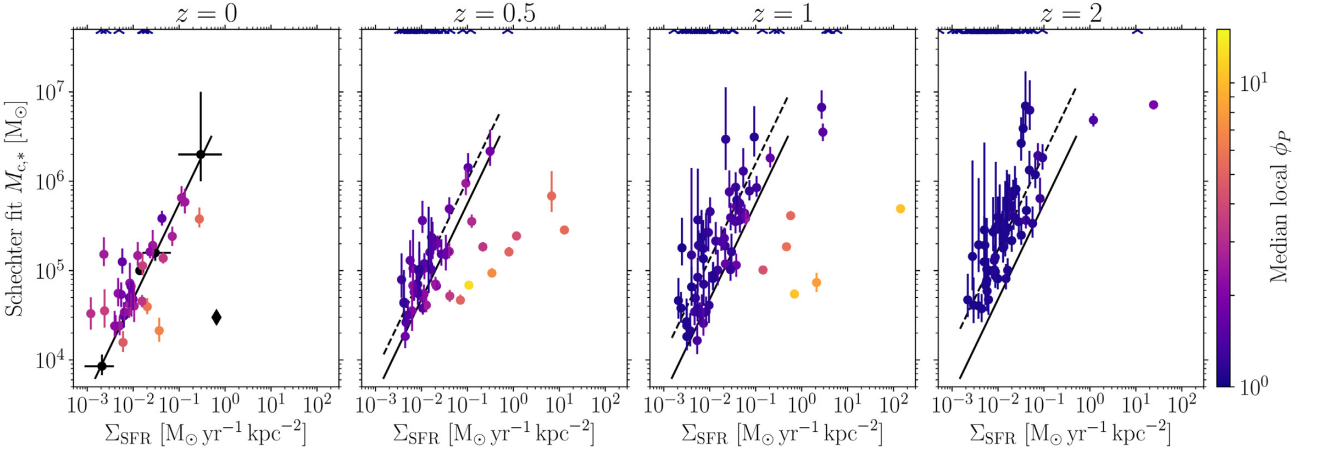
Figure 5. Mass function truncation $M_{c,*}$ as a function of Σ_{SFR} at redshifts $z = 0, 0.5, 1, 2$ (panels from left to right), coloured by the median ϕ_P of star particles younger than 300 Myr. The arrows at the top of the panels indicate galaxies for which $M_{c,*}$ was unable to be constrained. The black points show observed galaxies at $z = 0$ for reference (as in Fig. 4). The dashed lines show the fit to the observed galaxies rescaled assuming the median ϕ_P at each redshift found for the simulations (see Appendix C).

$z = 2$ relation, since ϕ_P cannot be less than unity. This result could be tested in galaxies from the local Universe by comparing cluster formation in regions of high stellar density (high ϕ_P) and low stellar density (low ϕ_P) at similar Σ_{SFR}.

In Fig. 5 (particularly evident at $z = \{0.5, 1\}$), a number of galaxies fall well below the present-day $M_{c,*}$–Σ_{SFR} relation, approaching the value found for the CMZ in the Milky Way (Trujillo-Gomez et al. 2019). This is driven by two effects, both due to star formation in regions of high stellar density (i.e. very central star formation in the galaxy). First, such galaxies have a decreased $M_{c,*}$ due to the higher contribution of the gravity of stars to the mid-plane gas pressure (high ϕ_P, as discussed above; see Fig. C1). However, a high stellar density (high ϕ_P) alone is not sufficient to account for the large decrease in $M_{c,*}$. For example, at $z = 0$ a number of galaxies are significantly elevated in ϕ_P ($\phi_P > 3$; Fig. C1), but fall along the present-day $M_{c,*}$–Σ_{SFR} relation (Fig. 4). The main contributing factor is due to the high Coriolis or centrifugal forces in the region of star formation,[5] resulting in decreased Toomre masses, and therefore $M_{c,*}$. This is captured by the Reina-Campos & Kruijssen (2017) model in terms of a dependence on the epicyclic frequency, which is not accounted for in the simple scaling with Σ_{SFR} from Johnson et al. (2017). Though this effect is most evident at $z = \{0.5, 1\}$ in the simulations, it may occur at any redshift and simply results from very central star formation. Such an effect has been observed at the centre of local Universe galaxies (Reina-Campos & Kruijssen 2017; Messa et al. 2018b; Trujillo-Gomez et al. 2019).

3.3 Mass function index

Cook et al. (2016) found that galaxies with higher SFR surface densities tend to have flatter cluster luminosity function indices. They suggest that this might be a result of the cluster formation process, with higher star formation efficiencies resulting in proportionally more massive star-forming regions. In this section, we investigate how other effects, namely increased relative mass-loss towards low cluster masses or the degeneracy between the power-law index and

$M_{c,*}$, may instead cause this effect. Following the method described in Section 2.3, we fit Schechter and power-law mass functions to the final (evolved) cluster populations, using a variable mass function index with a uniform prior of $-3 < \beta < -0.5$ (similar to Johnson et al. 2017; Messa et al. 2018b).

In Fig. 6 we compare the cluster mass function index with the SFR density of the galaxy, for galaxies in the $z = \{0, 0.5, 1, 2\}$ snapshots (in order to increase the galaxy sample and extend the range in Σ_{SFR}). The upper panel shows the results for Schechter function fits, while the lower panel shows the results for pure power-law fits. The figure includes all galaxies with >50 clusters with evolved masses $> 5 \times 10^3 \, M_{\odot}$ (regardless of how well $M_{c,*}$ is constrained in the case of Schechter fits), and therefore includes the effect of the degeneracy between β and $M_{c,*}$ in the fits. We find that the cluster mass function indices are flatter at higher SFR densities, similar to the effect found by Cook et al. (2016). This result is true for both Schechter and power-law fits, though the latter tend to find steeper mass function indices. In the simulations this is caused by two effects. At high SFR densities ($\Sigma_{SFR} \gtrsim 10^{-1.5} \, M_{\odot} \, yr^{-1} \, kpc^{-2}$), the mass function indices become flatter due to increased relative mass-loss towards low cluster masses, primarily due to tidal shocks from dense gas. At low SFR densities ($\Sigma_{SFR} \lesssim 10^{-1.5} \, M_{\odot} \, yr^{-1} \, kpc^{-2}$), mass function indices may appear steeper due to low truncation masses and the degeneracy between the index and the truncation mass (galaxies with $\beta < -2$ generally have poorly constrained $M_{c,*}$). This crossing point, where galaxies typically fall above or below an index of -2, depends on the lower cluster mass limit; higher or lower mass limits result in higher or lower crossing points in Σ_{SFR}, respectively. When using a lower cluster mass limit of $10^4 \, M_{\odot}$, rather than $5 \times 10^3 \, M_{\odot}$, the crossing point shifts to higher Σ_{SFR} by ≈ 0.3 dex. Observed local galaxies at low Σ_{SFR} ($\sim 10^{-3}$–$10^{-2} \, M_{\odot} \, yr^{-1} \, kpc^{-2}$) are also consistent with the power-law indices for the cluster mass function found in this work ($\beta \approx -2$, e.g. M31, Johnson et al. 2017; M51, Messa et al. 2018a).

The mass function indices at a given Σ_{SFR} for $z = 0$ and $z > 0$ galaxies are generally consistent. However, at $\Sigma_{SFR} \lesssim 10^{-2} \, M_{\odot} \, yr^{-1} \, kpc^{-2}$, galaxies at $z = 0$ tend to have steeper indices due to their smaller $M_{c,*}$ (Section 3.2).

Since their methods differ from ours (they fit luminosity functions and have a variable lower cluster luminosity limit between galaxies), we cannot make a direct quantitative comparison to

---

[5] A high ϕ_P does not imply high Coriolis/centrifugal forces, but high Coriolis/centrifugal forces generally occur in regions with high ϕ_P.
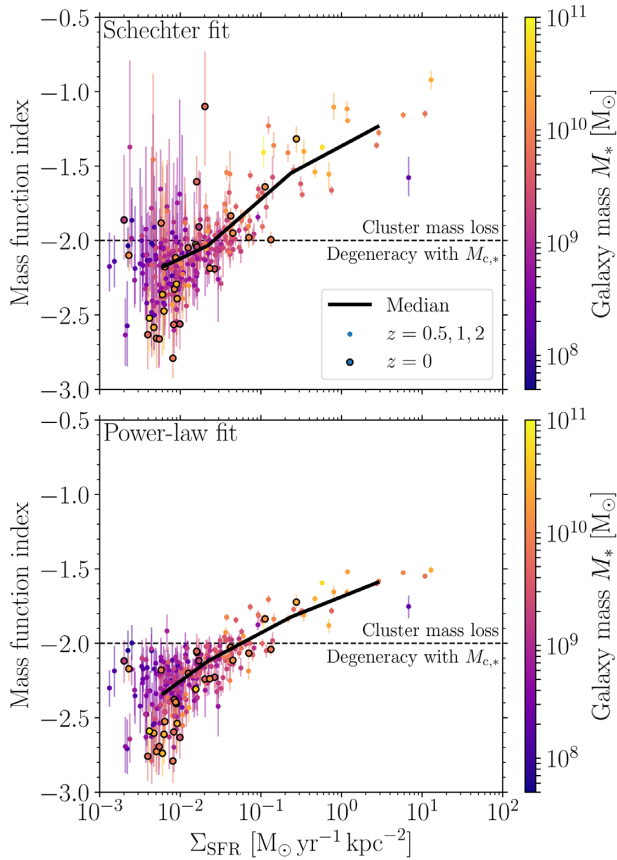
Figure 6. Power-law index of the final cluster mass function, fit with a Schechter (1976) function (upper panel) and a power-law function (lower panel), as a function of the star formation rate density of the galaxy, Σ_{SFR}. The points show the median of the posterior distribution from the MCMC fit, while error bars show the 16 and 84 per cent confidence intervals. Galaxies from the $z = \{0, 0.5, 1, 2\}$ snapshots are included in the figure, with galaxies at $z = 0$ highlighted with black circles. The black lines show the median mass function index and Σ_{SFR} in 1 dex intervals from 10^{-3} to $10^2\,\mathrm{M_\odot\,yr^{-1}\,kpc^{-2}}$. Initially, all clusters are drawn from a Schechter function with index $\beta = -2$ (dashed line in the figure). The flatter mass functions at higher star formation rate densities are caused by dynamical mass-loss of the clusters, while indices may be steeper than the initial value at low Σ_{SFR} due to the degeneracy between the index and mass function truncation.

the results from Cook et al. (2016). Additionally, the simulations and observations are largely biased to different SFR densities ($\Sigma_{\mathrm{SFR}} \gtrsim 10^{-2.5}\,\mathrm{M_\odot\,yr^{-1}\,kpc^{-2}}$ and $\Sigma_{\mathrm{SFR}} \lesssim 10^{-2}\,\mathrm{M_\odot\,yr^{-1}\,kpc^{-2}}$, respectively). However, the results from the simulations suggest that the relation between cluster luminosity function index and Σ_{SFR} found by Cook et al. (2016) can be explained by the degeneracy between the mass function index and truncation $M_{\mathrm{c,*}}$ (at low Σ_{SFR}). Similar measurements of the mass function index should be extended to higher SFR densities to assess and test the impact of cluster mass-loss.

3.4 Specific *U*-band cluster luminosity

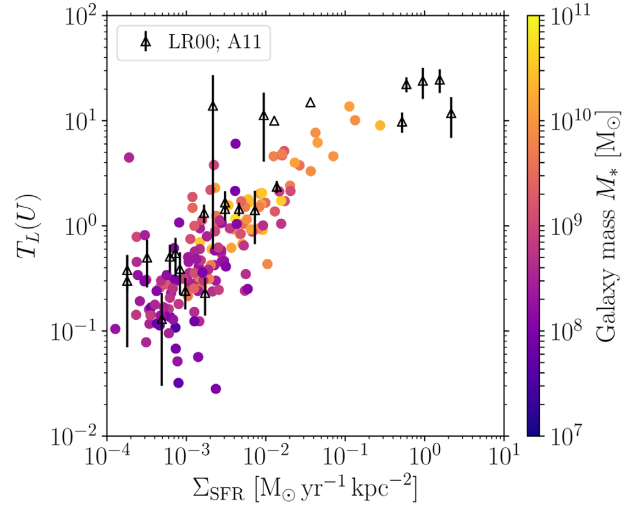An empirical precursor to the CFE, the specific *U*-band cluster luminosity, $T_L(U) = 100 L_{\mathrm{clusters}}/L_{\mathrm{galaxy}}$ (i.e. the percentage of *U*-band light of a galaxy contributed by star clusters), was introduced by Larsen & Richtler (2000), who found it correlated strongly with Σ_{SFR} for observed galaxies. While the CFE is the most relevant

Figure 7. Specific *U*-band cluster luminosity, $T_L(U) = 100 L_{\mathrm{clusters}}/L_{\mathrm{galaxy}}$ (i.e. the percentage of the *U*-band light of a galaxy contributed by star clusters), as a function of Σ_{SFR}, with points coloured by the stellar mass of the simulated galaxies. The open triangles show observed galaxies from Larsen & Richtler (2000) and Adamo et al. (2011).

quantity for simulations of cluster formation, it is not a direct observable and a number of caveats apply to its estimation (e.g. assumptions for and extrapolation of the cluster mass function, uncertainties in masses, and ages from stellar population modelling, SFR estimations, corrections for dust, etc.). On the other hand, $T_L(U)$ can be directly determined from observations of galaxies (though may depend on selection criteria for YSCs) and therefore presents a useful test for models of YSC populations.

In Fig. 7 we show $T_L(U)$ for the simulated galaxies at $z = 0$. We calculate the total luminosity of star clusters for all surviving clusters with initial masses $> 5 \times 10^3\,\mathrm{M_\odot}$. In order to calculate the total *U*-band luminosity of the galaxy, we assume simple stellar populations for each star particle and calculate their luminosities using the method described in Section 2.3. Both luminosities were calculated within R_{lim} and assuming no extinction.[6] We compare the simulated galaxies in Fig. 7 with observed galaxies from Larsen & Richtler (2000) and Adamo et al. (2011). We find good agreement in the trend of $T_L(U)$ with Σ_{SFR} between the simulated and observed galaxies. At low Σ_{SFR} ($\sim 10^{-3}\,\mathrm{M_\odot\,yr^{-1}\,kpc^{-2}}$), $T_L(U)$ may be slightly underestimated in the simulations due to the instantaneous disruption of clusters with initial masses $M < 5 \times 10^3\,\mathrm{M_\odot}$. However, similar limitations also apply to the observed galaxies, depending on the distance to the galaxy and detectability of clusters. In the simulations, $T_L(U)$ is largely determined by the CFE, such that those galaxies with a high CFE also have a high $T_L(U)$. The scatter in $T_L(U)$ at fixed Σ_{SFR} (or CFE) shows no clear trends with sSFR or metallicity, and arises from temporal variations in the CFE and Σ_{SFR}, as well as stochastic sampling at the high-mass end of the cluster mass function.

In Fig. 8, we quantify the effect on $T_L(U)$ of varying the star cluster formation physics in the simulations and show the four cluster formation models described in Section 2.1. Each model

---

[6]Adopting a basic model for extinction where all stars and clusters are embedded within an optically thick cloud until a specific age (e.g. 10 Myr; c.f. Charlot & Fall 2000) has no effect on the results, because extinction has the same effect on stars and clusters.
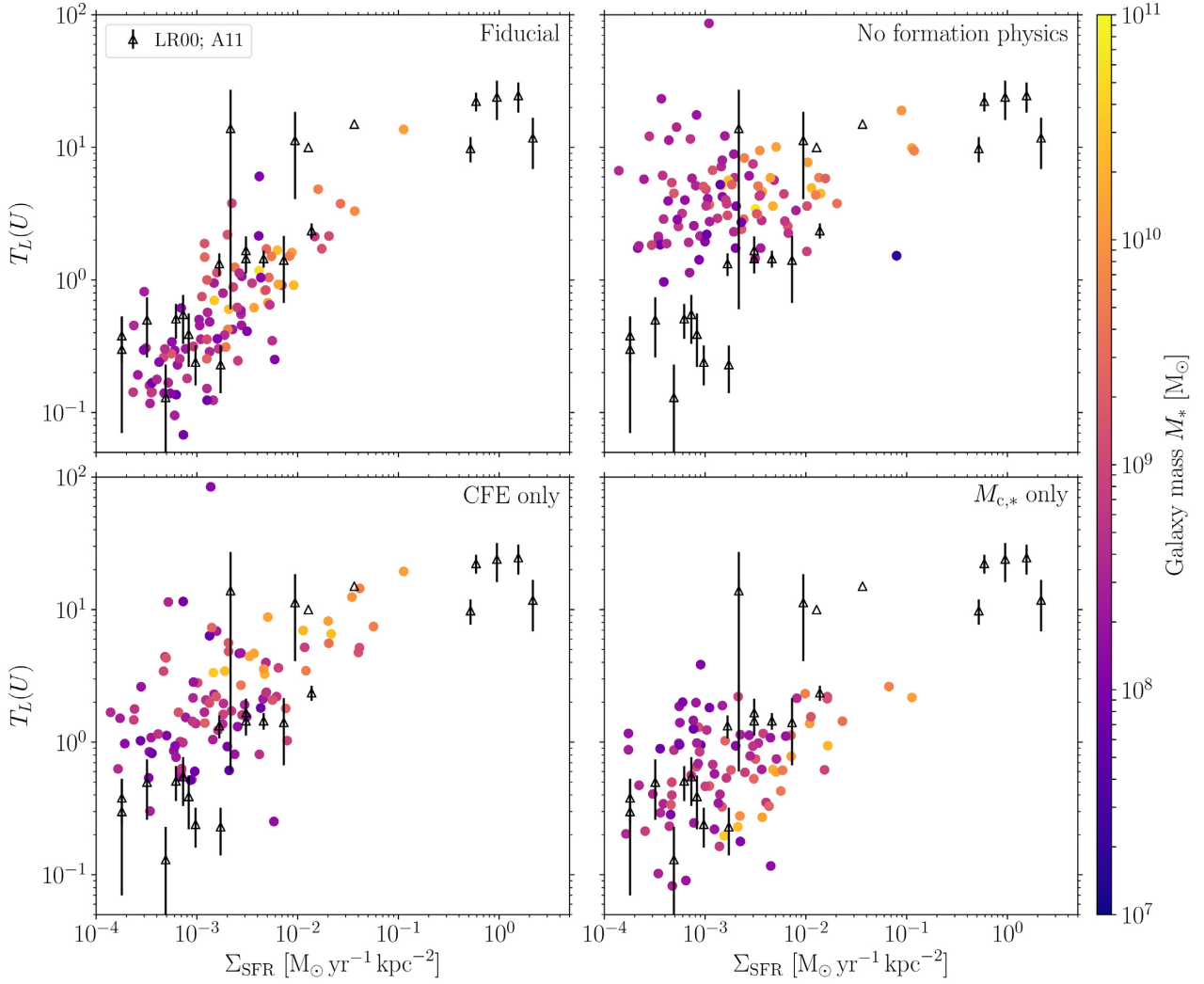
Figure 8. $T_L(U)$ for the simulated galaxies with different cluster formation physics. The open triangles show observed galaxies from Larsen & Richtler (2000) and Adamo et al. (2011). *Top left*: Fiducial E-MOSAICS model. *Top right*: Constant CFE (10 per cent), $M_{c,*} = \infty$. *Bottom left*: environmentally dependent CFE, $M_{c,*} = \infty$. *Bottom right:* environmentally dependent $M_{c,*}$, constant CFE (10 per cent).

is a variation on the fiducial, environmentally dependent model (top left-hand panel), with environmentally independent versions for either the truncation mass or CFE (*CFE only* and $M_{c,*}$ *only* models, bottom left-hand and right-hand panels, respectively), or both (*no formation physics* model, top right-hand panel). The figure only includes galaxies from the L012N0376 volume and the first ten zoom-in simulations (MW00–MW09), i.e. simulations with all four variations of the cluster formation physics (and thus the top left-hand panel shows fewer galaxies than in Fig. 7).

Fig. 8 clearly shows the critical role of both the CFE and $M_{c,*}$ models in reproducing the observations of $T_L(U)$. With a constant CFE ($\Gamma = 0.1$) and pure power-law mass function (upper right-hand panel), the *no formation physics* model implies a (roughly) constant $T_L(U)$, and therefore cannot simultaneously reproduce galaxies of high (>10) and low (<1) specific luminosities. The *CFE only* model (bottom left-hand panel) assumes an environmentally dependent CFE and a pure power-law mass function. Due to the variation of the CFE with Σ_{SFR}, the model agrees better with the observed galaxies than for the model with constant CFE. However, a variation in CFE alone (at least in the current formulation of the Kruijssen 2012

model) is also largely unable to account for galaxies with $T_L(U) < 1$. The *CFE only* model predicts higher $T_L(U)$ than observed at $\Sigma_{SFR} \sim 10^{-3}$ $M_\odot$ yr^{-1} kpc^{-2}, but agrees with the observed galaxies at higher Σ_{SFR}. The bottom right-hand panel of Fig. 8 shows the $M_{c,*}$ *only* model, which assumes a constant CFE = 0.1 and an environmentally dependent $M_{c,*}$. Though the model assumes the same constant CFE as for the *no formation physics* model, the $M_{c,*}$ *only* model shows good agreement for galaxies with $\Sigma_{SFR} \lesssim 10^{-2}$ $M_\odot$ yr^{-1} kpc^{-2}, but underpredicts $T_L(U)$ at higher Σ_{SFR}. In the $M_{c,*}$ *only* model, $T_L(U)$ is significantly lower than the assumed 10 per cent CFE due to the low $M_{c,*}$ (at low Σ_{SFR}) and the preferential formation of very low mass clusters, which are not detected.[7] Therefore, we conclude that environmental variations in both the CFE and $M_{c,*}$ are necessary for reproducing the observed $T_L(U)$ relation.

---

[7]Note that $T_L(U)$ will therefore depend upon the lower initial cluster mass limit in the simulations (we adopt 5×10^3 $M_\odot$). However, similar detection limits also apply for observed galaxies, depending on cluster age (e.g. Adamo et al. 2015; Johnson et al. 2015; Messa et al. 2018a)
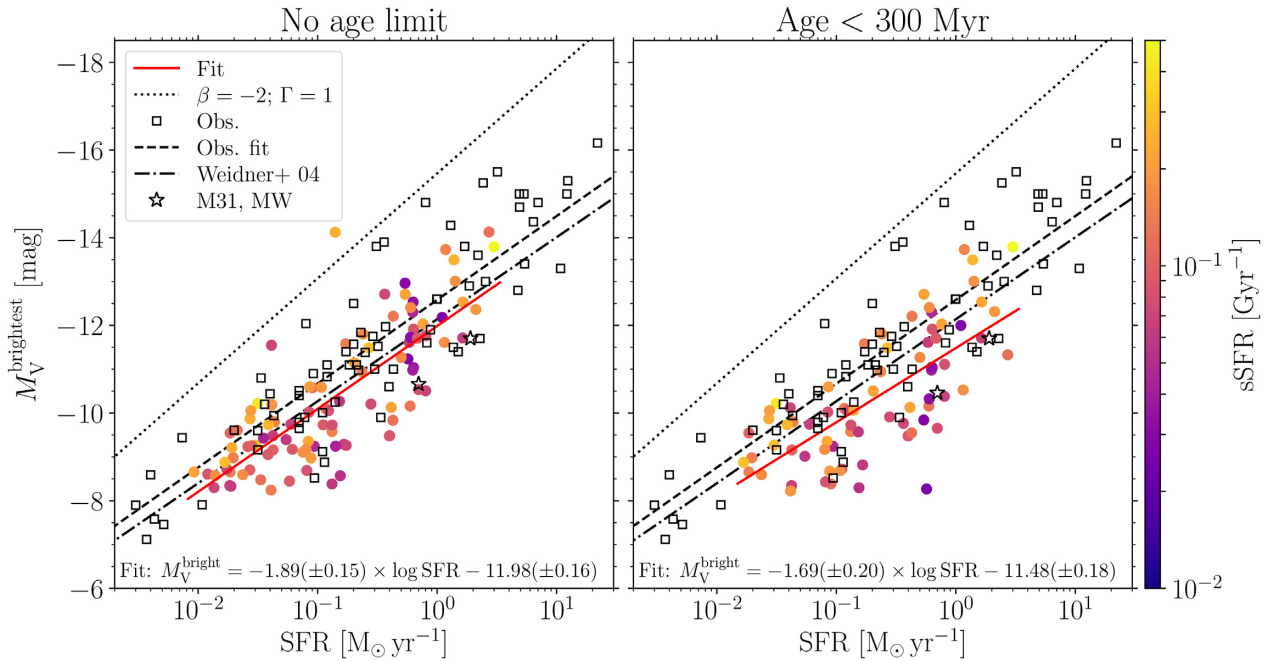
Figure 9. The $M_V^{\text{brightest}}$–SFR relation for the fiducial cluster formation model with no cluster age limit (left-hand panel) and an age limit of 300 Myr (right-hand panel). Each point represents an individual galaxy, with the symbols coloured by specific star formation rate of the galaxy. The solid red line is a linear regression fit to the simulations, with the equation shown in each panel. The black squares show the sample of observed galaxies compiled by Adamo et al. (2015) and the dashed line shows the best-fitting relation to this sample (equation 1). The dash-dotted line shows the best-fitting observed relation from Weidner, Kroupa & Larsen (2004, their equation 2). Stars show the Milky Way (SFR = $1.9\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$) and M31 (SFR = $0.7\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$) for comparison (see Section 3.5.1 for details). The dotted line is the expectation for a $\beta = -2$ power-law mass function and constant 100 per cent CFE (Bastian 2008), which is inconsistent with the observed relation. The slope of the best-fitting relation from the simulations is fully consistent with the slope of the relations from the observed galaxies.

3.5 Brightest cluster–SFR relation

3.5.1 Fiducial E-MOSAICS model

The relation between the brightest cluster and the SFR of the galaxy is an empirical relation, of which the construction does not rely on modelled quantities such as clusters ages or masses. The relation is sensitive to cluster formation physics and therefore presents a useful test of cluster formation models. Moreover, neither the star cluster nor galaxy formation physics implemented in the simulations were calibrated to reproduce the relation, thus a comparison affords an independent test of the predictions of the simulations.

In this section we consider two cases. First, we consider the relation without applying an age limit to the simulated clusters, since in a number of cases observational measurements are obtained with single-band photometry, prohibiting age measurements (e.g. Larsen 2002). Secondly, we apply an upper age limit of 300 Myr, consistent with the calculation for the SFR. For the simulations we apply a cluster luminosity limit of $M_V < -8.2$ (assuming a metallicity of $\log(Z/Z_\odot) = -0.5$, similar to the metallicity of star-forming gas in an $M_* \sim 10^8\,\mathrm{M}_\odot$ galaxy in the EAGLE Recal model; Schaye et al. 2015) to reflect the $5 \times 10^3\,\mathrm{M}_\odot$ initial cluster mass limit, which is similar to the luminosity limit of observational programmes (e.g. Whitmore et al. 2014). V-band mass-to-light ratios for the simulated clusters are calculated using the FSPS model (see Section 2.3). Luminosities were then determined using the current cluster masses, which include cluster mass-loss.

Fig. 9 shows the predictions for the brightest cluster in the V-band as a function of galaxy SFR (both measured within $1.5R_{1/2}$) for the fiducial E-MOSAICS model, with no cluster age limit (left-hand

panel) and with an upper limit of 300 Myr (right-hand panel). The best-fitting relation for the simulated galaxies from a least-squares linear regression is given in each panel and shown as the solid red line. The brightest cluster is generally consistent for both with and without an age limit. However, for some galaxies, the brightest young cluster is significantly fainter than the brightest cluster of any age, which results in a slightly flatter slope of the best-fitting relation for the $< 300\,\mathrm{Myr}$ age limit. For comparison, the figure also shows the sample of observational measurements compiled by Adamo et al. (2015). Note that we have not attempted to match the sample selection for the observations, other than the selection in sSFR for the simulations, and therefore some bias between the simulated and observed galaxy populations may exist. The best-fitting relation for the observed sample of galaxies is given by

$$M_V^{\text{brightest}} = -1.91(\pm 0.09) \times \log \frac{\text{SFR}}{\mathrm{M}_\odot\,\mathrm{yr}^{-1}} - 12.58(\pm 0.13), \quad (1)$$

shown as a dashed line. To this compilation of galaxies we also add measurements of the Milky Way and M31. For the Milky Way we assume SFR = $1.9\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$ (Chomiuk & Povich 2011) and the brightest cluster to be Westerlund 1, with $M_V^{\text{brightest}} \approx -11.7$ (assuming a mass of $6 \times 10^4\,\mathrm{M}_\odot$ and age of 5 Myr, Mengel & Tacconi-Garman 2007, and a V-band mass-to-light ratio from FSPS assuming a Solar metallicity). For M31 we assume SFR = $0.7\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$ (Kang, Bianchi & Rey 2009; Lewis et al. 2015). We take the brightest cluster in M31 (of any age) to be the globular cluster G1, with $M_V^{\text{brightest}} = -10.66$ (Galleti et al. 2004). For the brightest young cluster we use the brightest cluster from the PHAT survey, $M_V^{\text{brightest}}(< 1\,\mathrm{Gyr}) \approx -10.46$ (Johnson et al. 2015), using
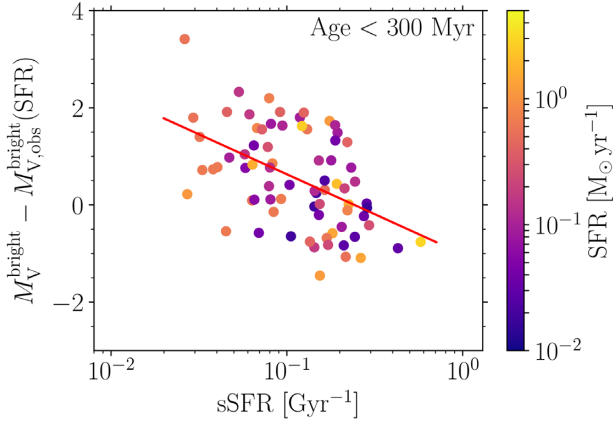
Figure 10. Difference in magnitude of the brightest cluster from the Weidner et al. (2004) $M_V^{\rm brightest}$–SFR relation (their equation 2) compared with the sSFR of the galaxy, for the simulated galaxies in the right-hand panel of Fig. 9. The solid red line shows the best-fitting relation $\Delta M_V = -1.64(\pm 0.35) \times \log({\rm sSFR/Gyr}^{-1}) - 1.01(\pm 0.34)$. Galaxies with lower sSFRs typically have fainter clusters due to having lower cluster mass function truncations ($M_{\rm c,*}$).

their equation (6) to calculate a V-band magnitude and assuming a distance modulus of 24.47 (McConnachie et al. 2005).

The slope of the best-fitting relation for the simulated galaxies (-1.89 ± 0.15 for no age limit, -1.69 ± 0.2 for clusters < 300 Myr) is fully consistent with that of the observed galaxies (-1.91 ± 0.09 for the Adamo et al. 2015 sample; -1.87 ± 0.06 for the relation from Weidner et al. 2004). The scatter around the observed relation is also very similar for the simulations and observations, with standard deviations in $M_V^{\rm brightest}$ from the predicted and observed relations of ≈ 0.95 and 1.01 mag, respectively. Therefore, the observed $M_V^{\rm brightest}$–SFR relation is reproduced by the fiducial E-MOSAICS model with an environmentally varying CFE and upper mass function truncation mass, such that star formation in environments with high natal gas pressures results in more star formation in bound clusters and up to higher cluster masses.

To investigate the origin of scatter away from the $M_V^{\rm brightest}$–SFR relation, in Fig. 9 we colour the simulated galaxies by their sSFR. In the right-hand panel (ages < 300 Myr), at a fixed SFR the simulations show a gradient in sSFR, such that the galaxies with the brightest clusters also typically have the highest sSFR (or lowest galaxy masses). This trend is weaker in the left-hand panel (no age limit) as cluster luminosities may be uncorrelated with the present SFR. We explore this further in Fig. 10, where we show the magnitude difference from the observed $M_V^{\rm brightest}$–SFR relation (Weidner et al. 2004) compared with the sSFR for the simulated galaxies for cluster ages < 300 Myr. Simulated galaxies with brightest clusters that are significantly fainter than the observed relation typically have the lowest sSFRs. The fitted relation crosses the zero-point (in $\Delta M_V^{\rm brightest}$) at sSFR ≈ 0.2 Gyr^{-1}, similar to the typical sSFR (which is a weak function of galaxy mass) for galaxies on the star-forming main sequence (Schiminovich et al. 2007). By selecting galaxy populations at different constant sSFRs, a prediction of our fiducial model is that we expect to find (age-limited) $M_V^{\rm brightest}$–SFR relations that are offset to fainter luminosities at lower sSFRs. Additionally, galaxies at higher redshifts, with higher sSFRs, should be offset to brighter $M_V^{\rm brightest}$. The physical cause of this effect in the simulations is lower cluster mass function truncations in higher mass galaxies (which at fixed SFR have lower sSFR). These galaxies

typically have lower gas mass fractions, which result in a higher ϕ_P, while larger galaxy masses result in an increased importance of Coriolis/centrifugal forces in setting the Toomre mass. Both factors result in a lower cluster mass function truncation. This result could be tested with future observations in galaxies of different masses at fixed sSFR.

3.5.2 Alternative cluster formation models

The result from Fig. 9 that environmentally dependent cluster formation reproduces the $M_V^{\rm brightest}$–SFR relation, can be further tested by considering the alternative cluster formation model variations in the E-MOSAICS suite. In Fig. 11, we compare the $M_V^{\rm brightest}$–SFR relation for the four cluster formation models described in Section 2.1. As in Fig. 8 (Section 3.4), each model is a variation on the fiducial, environmentally dependent model (top left-hand panel), with environmentally independent versions for either the truncation mass or CFE (*CFE only* and $M_{\rm c,*}$ *only* models, bottom left-hand and right-hand panels, respectively), or both (*no formation physics* model, top right-hand panel) and only includes galaxies from simulations with all four versions of the cluster formation physics (the L012N0376 volume and first ten zoom-in simulations, MW00-MW09). Fig. 11 highlights the importance of cluster formation physics in governing the $M_V^{\rm brightest}$–SFR relation.

In the *no formation physics* model (top right-hand panel), which assumes $\Gamma = 0.1$ and a $\beta = -2$ power-law mass function (i.e. with our standard cluster formation physics disabled), the simulations are inconsistent with the observed relation and recover the slope of the expected relation for a pure power-law mass function and constant CFE (dotted line in the figure; Bastian 2008). The relation is therefore determined by a size-of-sample effect, with larger cluster populations more likely to have brighter clusters. The stochasticity at the high-mass end of the mass function induces a large scatter between galaxies at a given SFR. Also for this reason, the absolute offset in the relation is determined by the choice of CFE. However, for any choice of constant CFE, the slope of the relation will remain inconsistent with the observations.

In the *CFE only* model (bottom left-hand panel), which assumes an environmentally dependent CFE and a $\beta = -2$ power-law mass function, the simulations are again inconsistent with the observed relation. Due to the correlation of the CFE with galaxy mass (and thus SFR; see also Fig. 2), the *CFE only* model yields a relation that is even steeper than the *no formation physics* model, since low-mass galaxies with low SFRs typically have CFEs below 10 per cent. Thus at low SFRs in the figure (SFR $< 0.1 \, {\rm M}_\odot \, {\rm yr}^{-1}$) many galaxies are consistent with the observed relation, while those at higher SFRs are inconsistent with observed counterparts due to the lack of a truncation in the cluster mass function. Again, there is large galaxy-to-galaxy scatter at a given SFR due to the stochasticity at the high-mass end of the mass function.

Finally, the $M_{\rm c,*}$ *only* model, with a constant $\Gamma = 0.1$ and an environmentally dependent exponential truncation mass to the cluster mass function ($M_{\rm c,*}$), is shown in the bottom right-hand panel of the figure. At low SFRs, the simulations are consistent with observed galaxies. However, since the CFE does not vary between galaxies, this model predicts a flatter slope than is observed, and thus at higher SFRs the brightest clusters are underluminous compared to observed galaxies.

These results therefore show that an environmental dependence of both the CFE and cluster truncation mass is necessary for reproducing the observed properties of YSC populations.
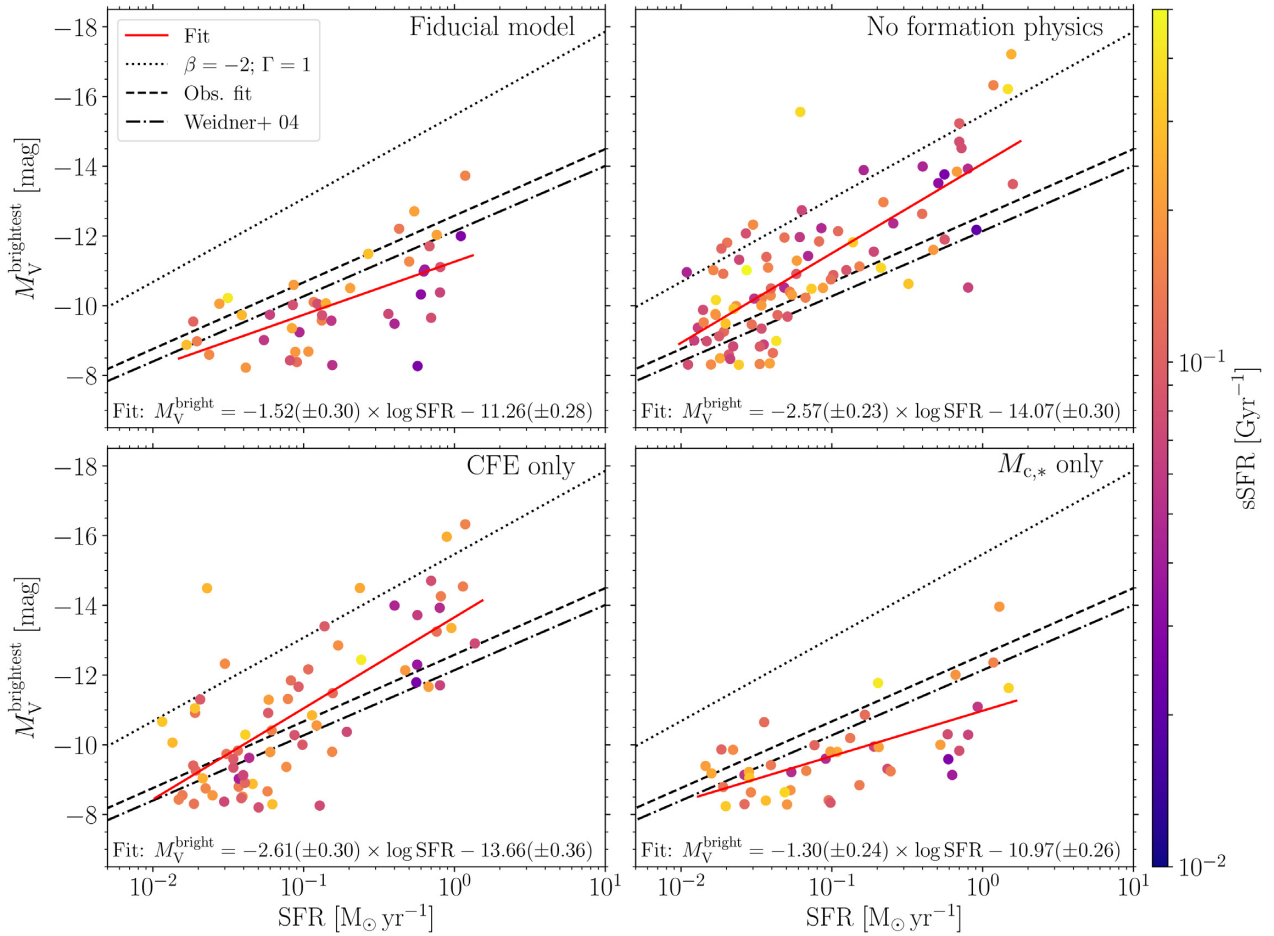
Figure 11. $M_V^{\rm brightest}$–SFR relation (for clusters < 300 Myr) for the simulated galaxies with different cluster formation physics. *Top left*: Fiducial E-MOSAICS model. *Top right*: Constant CFE (10 per cent), $M_{c,*} = \infty$. *Bottom left*: environmentally dependent CFE, $M_{c,*} = \infty$. *Bottom right:* environmentally dependent $M_{c,*}$, constant CFE (10 per cent). The dashed, dash-dotted, and dotted lines as in Fig. 9.

3.6 Cluster age distributions

In this section we investigate the effect of time-varying star and cluster formation rates (CFRs) on cluster age distributions. The age distribution of clusters is considered to be a strong test of cluster mass-loss models, potentially enabling the discrimination between mass/environmentally dependent or independent cluster disruption (e.g. Gieles et al. 2005; Lamers et al. 2005; Whitmore, Chandar & Fall 2007; Bastian et al. 2009; Lamers 2009; Chandar, Fall & Whitmore 2010; Kruijssen et al. 2011, 2012; Bastian et al. 2012; Miholics, Kruijssen & Sills 2017, see Adamo & Bastian 2018 for a review). However, the effect of time-varying SFRs and CFRs on the interpretation of cluster age distributions has not been investigated in realistic galaxy simulations in cosmological environments.

In the upper panel of Fig. 12, we show the cluster age distributions of galaxies from the zoom-in simulations of Milky Way-mass galaxies. We apply a cluster mass limit of $> 5 \times 10^3 \, \rm M_\odot$ at all ages and use clusters within $R_{\rm lim}$ from the centre of the potential. We limit the sample to galaxies with at least 100 surviving clusters younger than $10^{9.25}$ yr (similar to the number of clusters in M31 more massive than $5 \times 10^3 \, \rm M_\odot$; Fouesneau et al. 2014; Johnson et al. 2016), leaving 12 galaxies. For the age distributions we use eight bins with widths of 0.375 dex between $10^{6.25}$ and $10^{9.25}$ yr, where the distributions are normalized at $10^{7.5}$ yr by fitting a power-law relation to the four bins between 10^7 and $10^{8.5}$ yr. The typical

time-steps for stellar particles at $z \approx 0$ are ~ 1 Myr, and therefore even in the youngest age bins the numerical sampling of disruption by tidal heating is adequate. We find a large scatter between the simulations, with some galaxies showing very flat age distributions and others where the number of clusters decreases rapidly at ages $> 10^8$ yr. We discuss the cause of this scatter between galaxies below. The solid line in the figure shows the total cluster age distribution for all of the galaxies combined. For the four bins in the range log(age/ yr) = {7, 8.5} (where observations are complete) we find a slope of -0.39 ± 0.04 for the total population, with a range between -1.18 and 0.17 for individual galaxies.

We compare the cluster age distributions from the simulations with the observed distributions in M31 (Fouesneau et al. 2014; Johnson et al. 2016), M83 (Silva-Villa et al. 2014; Adamo & Bastian 2018), and M51 (Messa et al. 2018a,b), also using a mass limit of $> 5 \times 10^3 \, \rm M_\odot$. This sample comprises observed galaxies most similar to our sample of simulated Milky Way-mass galaxies. At ages $< 10^7$ yr, the observations may be contaminated by unbound stellar associations (Gieles & Portegies Zwart 2011; Bastian et al. 2012; Ward & Kruijssen 2018), while at ages $> 10^{8.5}$ yr cluster populations are incomplete due to luminosity limits. Within the region for which observations are complete (10^7–$10^{8.5}$ yr), the age distributions of the simulations and observations are in good agreement, showing a similar level of scatter between different
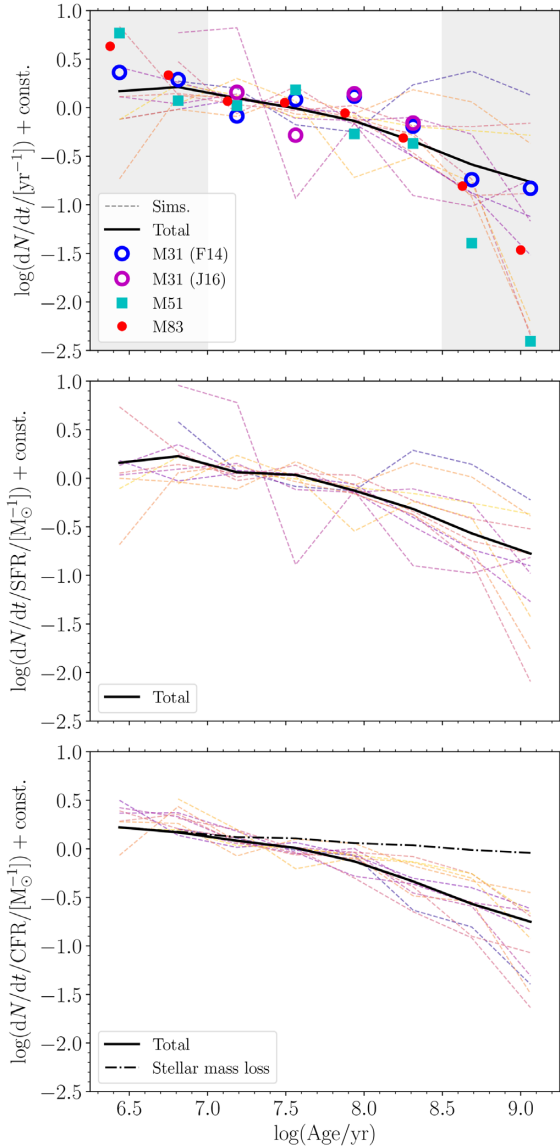
Figure 12. *Upper:* Cluster age distributions for 12 of the target L^* galaxies in the Milky Way-mass zooms (dashed lines, with a different colour for each galaxy). For comparison we show cluster age distributions from M83 (using the distribution from Silva-Villa et al. 2014; Adamo & Bastian 2018), M31 (Fouesneau et al. 2014 and Johnson et al. 2016), and M51 (Messa et al. 2018a,b), all using a cluster mass limit of $> 5 \times 10^3 \, M_\odot$. The shaded regions show ages where observed cluster populations may be contaminated by unbound associations ($< 10 \, \mathrm{Myr}$) and are typically incomplete due to luminosity limits ($> 300 \, \mathrm{Myr}$, for masses $> 5 \times 10^3 \, M_\odot$). *Middle:* Cluster age distributions normalized by the star formation rate (SFR). *Lower:* Cluster age distributions normalized by the cluster formation rate (CFR) for masses $> 5 \times 10^3 \, M_\odot$. The dash-dotted line shows the total cluster age distribution if only cluster mass-loss due to stellar evolution is included.

galaxies and consistent slopes of the age distributions (-0.07 and -0.14 for the Fouesneau et al. 2014 and Johnson et al. 2016 M31 catalogues, respectively, -0.43 for M51 and -0.33 for M83). We note that mass-loss by tidal shocks is underestimated in the simulations, due to the lack of an explicit model for the cold, dense phase of the star-forming interstellar medium in the EAGLE model (see Pfeffer et al. 2018). The extent to which this affects the age distribution predictions from the simulations depends upon what

cluster age this mechanism becomes important at $z \approx 0$ (which likely depends upon the local environment within which young clusters reside). However, a similar result, where cluster mass-loss is mainly important after a few hundred megayears, was found in simulations of isolated galaxies by Kruijssen et al. (2011, using the same MOSAICS dynamical evolution model, but using a galaxy formation model with a simple model for the cold, dense phase of the interstellar medium, see Pelupessy, van der Werf & Icke 2004).

The cluster age distribution is a function of both cluster formation and evolution. Therefore, variations in the SFR or CFR will also lead to variations in the age distributions. We assess the impact of these effects in the bottom two panels of Fig. 12. In the middle panel of Fig. 12, we show the cluster age distribution normalized by the total SFR in each temporal bin for each simulated galaxy. Overall, the distributions for the galaxies show similar behaviour to the standard distributions (upper panel), with little reduction in scatter between the galaxies (the scatter about the total relation decreased from 0.43 to 0.34 dex). This is expected, since the SFR in each galaxy is relatively constant over the period investigated. Over $10^{9.2} \, \mathrm{yr}$, the SFR in each temporal bin typically varies by less than a factor of two from the median for all bins. The majority of the impact of variations in the SFR occurs at $> 10^{8.5} \, \mathrm{yr}$, as can be seen by the slight reduction of scatter in the $10^9 \, \mathrm{yr}$ bin between the upper and middle panels in the figure. The slope of the SFR-normalized age distribution is also very similar to the standard age distribution (-0.35 ± 0.07, with a range between -1.14 and 0.15). Therefore, for the galaxies investigated, any variations in SFR induce little impact in the cluster age distributions.

In the bottom panel of Fig. 12 we show the cluster age distributions normalized by the CFR (for masses $> 5 \times 10^3 \, M_\odot$) in each temporal bin for the simulated galaxies. Again, the slope of the CFR-normalized age distribution is similar to the standard age distribution, but with a reduced range (-0.37 ± 0.06, with a range between -0.75 and -0.09). Unlike the SFR, normalizing by the CFR has a large effect on the age distributions, significantly reducing the galaxy-to-galaxy scatter in the distributions from 0.43 to 0.21 dex about the total distribution. When normalized by the CFR, the age distributions follow a very similar evolution for the galaxies, with some divergence at $> 10^8 \, \mathrm{yr}$ reflecting the varying cluster mass-loss rates between the galaxies (due to, e.g. differing potentials, gas densities, or galactocentric radii of cluster formation). Though the SFR varies by up to a factor of two between bins, the CFR can vary by up to a factor of six, and therefore the CFR is not simply following changes in the SFR.[8] In addition to the SFR, the CFE (i.e. the fraction of star formation in bound clusters) may also vary with time by up to a factor of two, and thus together account for about half of the variance in the CFR($M > 5 \times 10^3 \, M_\odot$). The rest of the variation (factor of three) can be attributed to the stochasticity of cluster formation at the high-mass end of the cluster mass function (seven of the galaxies have $M_{c,*} < 10^5 \, M_\odot$, and therefore in these galaxies only the high-mass end of the cluster population satisfies our mass cut of $M > 5 \times 10^3 \, M_\odot$). This effect is lessened in galaxies with larger cluster populations, although even for those with >1000 clusters above the mass limit, the variation in the mass-limited CFR, in addition to that of the SFR and CFE, is a factor of two. For the observed galaxies in Fig. 12, M31 may

---

[8] A similar level of variation occurs when comparing the initial age distribution, dN_{init}/dt (i.e. without including cluster mass-loss), rather than CFR($M > 5 \times 10^3 \, M_\odot$), thus it is not simply due to stochasticity in the masses of clusters that form.

be most affected by this effect since it has the fewest clusters with $M > 5 \times 10^3 \, M_\odot$, though this may be alleviated somewhat by including lower mass clusters (i.e. the completeness limit in the PHAT survey for clusters younger than 300 Myr is $\sim 1000 \, M_\odot$; Johnson et al. 2015).

In Fig. 12, we also show the contribution of stellar-evolutionary mass-loss (dash-dotted line) to the total cluster age distribution (solid line). The stellar mass-loss relation has a significantly flatter slope (≈ -0.1) than the total cluster age distribution. Stellar mass-loss dominates for ages $\lesssim 10^7$ yr, with dynamical mass-loss (two-body relaxation and tidal shocks) becoming dominant only at older ages.

Therefore, temporal variations in the CFE and CFR should also be considered when using cluster age distributions to test models for cluster mass-loss, since they may impart non-negligible variations in the cluster age distributions.

4 DISCUSSION AND CONCLUSIONS

In this paper, we present a comparison of the YSC populations in the E-MOSAICS simulations with observed populations in nearby galaxies. The aim of this work is to both test the cluster formation models in realistic simulations of galaxy formation and, by varying the cluster formation physics (the CFE and exponential truncation of the cluster mass function), obtain an insight into the formation processes at work in observed cluster populations.

We find that, due to a combination of spatially varying (non-uniform) star formation and the metallicity-dependent density threshold for star formation in the EAGLE model, the CFE in the simulated galaxies is elevated above the (input) Kruijssen (2012) relation at $\Sigma_{SFR} < 10^{-2.5} \, M_\odot \, \mathrm{yr}^{-1} \, \mathrm{kpc}^{-2}$ (Section 3.1; Fig. 2). In fact, our sample of simulated galaxies could be fit by a power-law relation in CFE$-\Sigma_{SFR}$ (also see Goddard et al. 2010), despite the input relation. A similar effect might be present in observed cluster populations, though this could also be caused by a non-linear $\Sigma_{gas}-\Sigma_{SFR}$ relation (Johnson et al. 2016).

For cluster populations in the simulated galaxies that can be fit with a Schechter (1976) mass function, the formulation of the Reina-Campos & Kruijssen (2017) model implemented in E-MOSAICS performs well in reproducing observed truncation masses (Section 3.2; Fig. 4). However, more observations are needed to test if the scatter in $M_{c,*}$ seen between galaxies (i.e. at a given Σ_{SFR}) in the simulations is consistent with observed galaxies. Due to the higher gas fractions of galaxies at higher redshifts, the E-MOSAICS model predicts that, at the epochs of GC formation ($z \gtrsim 1$), the $M_{c,*}-\Sigma_{SFR}$ relation should be elevated above the present day relation (Fig. 5).

In Section 3.3, we investigate the power-law index of the cluster mass function at low cluster masses, for evolved cluster populations. We find that in galaxies with high SFR surface densities ($\Sigma_{SFR} \gtrsim 10^{-1} \, M_\odot \, \mathrm{yr}^{-1} \, \mathrm{kpc}^{-2}$), mass function indices become flatter ($\beta > -2$) due to cluster disruption and increased relative mass-loss towards low cluster masses. In galaxies with lower SFR surface densities, mass function indices are similar to the initial index ($\beta = -2$) or may potentially become steeper ($\beta < -2$) due to the degeneracy between the mass function index and upper truncation mass ($M_{c,*}$). The results are consistent with the findings of Cook et al. (2016), who investigated observed cluster luminosity functions, suggesting that variable cluster mass function index is not necessary to explain their findings.

In Section 3.5, we test the relation between the brightest cluster within a galaxy and its SFR. This observed relation presents a useful test for the simulations since it is sensitive to cluster formation physics, yet was not used to test or calibrate the cluster formation models. We find that the fiducial E-MOSAICS cluster formation model reproduces both the slope of the $M_V^{\mathrm{brightest}}$–SFR relation and the scatter around the relation (Fig. 9). If an upper age limit (< 300 Myr) is applied to the clusters, we find that a number of the simulated galaxies fall significantly below the observed relation, which tend to be those with low sSFRs. This result implies that the $M_V^{\mathrm{brightest}}$–SFR relation may partially arise due to the preferential selection of galaxies on the star-forming main sequence of galaxies. Selecting samples of galaxies at fixed sSFRs may therefore result in offset $M_V^{\mathrm{brightest}}$–SFR sequences, which is a prediction offering a new avenue to test cluster formation models.

Some previous work has suggested that young cluster formation proceeds in an environmentally independent manner (Whitmore et al. 2007; Fall & Chandar 2012; Chandar, Fall & Whitmore 2015; Chandar et al. 2017). However, in Sections 3.4 and 3.5.2 we find that only the fully environmentally dependent cluster formation (fiducial) model can explain both the $T_L(U)$–Σ_{SFR} and $M_V^{\mathrm{brightest}}$–SFR relations. If cluster formation is environmentally independent (constant CFE and a power-law mass function with $\beta = -2$), both the slope of the $M_V^{\mathrm{brightest}}$–SFR relation and its absolute offset are inconsistent with the observed relation. Similarly, a constant CFE cannot simultaneously explain galaxies at both high and low specific U-band cluster luminosities [$T_L(U)$], while an environmental dependence in $M_{c,*}$ is required to explain galaxies with $T_L(U) < 1$. We therefore conclude that an environmental dependence in cluster formation, both for the CFE and the upper truncation of the cluster mass function ($M_{c,*}$), is required to reproduce observed young cluster populations. This shows that models in which a constant fraction of stars form in clusters are inconsistent with observations. The importance of environmentally dependent cluster formation in reproducing GC populations has similarly been discussed in previous work with the E-MOSAICS simulations (Pfeffer et al. 2018; Usher et al. 2018; Reina-Campos et al. 2019).

Finally, we compare the cluster age distributions in 12 (out of 25) of our sample of Milky Way-mass galaxies with those in observed nearby disc galaxies (Section 3.6). We find that, for ages where observed cluster populations are generally complete (between 10^7 and $10^{8.5}$ yr), the scatter in the age distributions between the simulated galaxies is similar to that in the observed galaxies. However, for the environmentally dependent cluster formation model, the variation of the CFR (not SFR) over Myr time-scales may impart significant variation in the cluster age distributions (Fig. 12). Therefore, the use of cluster ages as a discriminator of cluster mass-loss mechanisms should be approached with caution unless the variation of the CFR with time can be accounted for. The variation of the CFR is mainly driven by an underlying time-variation of the CFE. However, even for galaxies with >1000 clusters, additional (potentially dominant) variations may be imparted simply due to the stochasticity of cluster formation in time, with larger variations for less numerous cluster populations.

The results presented in this paper reinforce the understanding of cluster formation as an environmentally dependent process. We make predictions for future comparisons to observations ($M_{c,*}$ at high and low gas fractions, Section 3.2; the $M_V^{\mathrm{brightest}}$–SFR relation at varying sSFRs, Section 3.5), which will be useful in both further testing the models implemented in E-MOSAICS, as well as cluster formation theories in general. This work also highlights the importance of realistic simulations of galaxies with a diverse range of properties and environments when testing models of cluster formation. Comparing the simulations and observations at

the extremes of the galaxy population will enable a strong test of cluster formation theories, by testing them outside of the range in which they were developed, and allow further insight into the star and galaxy formation mechanisms which shape YSC populations across cosmic history.

ACKNOWLEDGEMENTS

REFERENCES

Adamo A., Bastian N., 2018, in Stahler S., ed., ASP Conf. Ser. Vol. 424, The Birth of Star Clusters. Astron. Soc. Pac., San Francisco, p. 91

Adamo A., Östlin G., Zackrisson E., Hayes M., Cumming R. J., Micheva G., 2010, MNRAS, 407, 870

Adamo A., Östlin G., Zackrisson E., 2011, MNRAS, 417, 1904

Adamo A., Kruijssen J. M. D., Bastian N., Silva-Villa E., Ryon J., 2015, MNRAS, 452, 246

Annibali F., Tosi M., Aloisi A., van der Marel R. P., 2011, AJ, 142, 129

Bahé Y. M. et al., 2016, MNRAS, 456, 1115

Barnes A. T., Longmore S. N., Battersby C., Bally J., Kruijssen J. M. D., Henshaw J. D., Walker D. L., 2017, MNRAS, 469, 2263

Bastian N., 2008, MNRAS, 390, 759

Bastian N., Trancho G., Konstantopoulos I. S., Miller B. W., 2009, ApJ, 701, 607

Bastian N. et al., 2012, MNRAS, 419, 2606

Baumgardt H., Parmentier G., Anders P., Grebel E. K., 2013, MNRAS, 430, 676

Bigiel F., Leroy A., Walter F., Brinks E., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2846

Bik A., Lamers H. J. G. L. M., Bastian N., Panagia N., Romaniello M., 2003, A&A, 397, 473

Billett O. H., Hunter D. A., Elmegreen B. G., 2002, AJ, 123, 1454

Booth C. M., Schaye J., 2009, MNRAS, 398, 53

Bournaud F., Duc P.-A., Emsellem E., 2008, MNRAS, 389, L8

Bourne N. et al., 2017, MNRAS, 467, 1360

Calzetti D. et al., 2015, AJ, 149, 51

Chabrier G., 2003, PASP, 115, 763

Chandar R., Fall S. M., Whitmore B. C., 2010, ApJ, 711, 1263

Chandar R., Fall S. M., Whitmore B. C., 2015, ApJ, 810, 1

Chandar R., Fall S. M., Whitmore B. C., Mulia A. J., 2017, ApJ, 849, 128

Charlot S., Fall S. M., 2000, ApJ, 539, 718

Chomiuk L., Povich M. S., 2011, AJ, 142, 197

Conroy C., Gunn J. E., 2010, ApJ, 712, 833

Conroy C., Gunn J. E., White M., 2009, ApJ, 699, 486

Cook D. O. et al., 2012, ApJ, 751, 100

Cook D. O., Dale D. A., Lee J. C., Thilker D., Calzetti D., Kennicutt R. C., 2016, MNRAS, 462, 3766

Cook D. O. et al., 2019, MNRAS, 484, 4897

Crain R. A. et al., 2015, MNRAS, 450, 1937

Crain R. A. et al., 2017, MNRAS, 464, 4204

Dolag K., Borgani S., Murante G., Springel V., 2009, MNRAS, 399, 497

Dowell J. D., Buckalew B. A., Tan J. C., 2008, AJ, 135, 823

Efremov Y. N., Elmegreen B. G., 1998, MNRAS, 299, 588

Elmegreen B. G., 1989, ApJ, 338, 178

Elmegreen B. G., 2002, ApJ, 564, 773

Elmegreen B. G., Efremov Y. N., 1997, ApJ, 480, 235

Elmegreen B. G., Elmegreen D. M., 2001, AJ, 121, 1507

Elmegreen B. G., Falgarone E., 1996, ApJ, 471, 816

Fall S. M., Chandar R., 2012, ApJ, 752, 96

Fedotov K., Gallagher S. C., Konstantopoulos I. S., Chandar R., Bastian N., Charlton J. C., Whitmore B., Trancho G., 2011, AJ, 142, 42

Fonnesbeck C., Patil A., Huard D., Salvatier J., 2015, PyMC: Bayesian Stochastic Modelling in Python, Astrophysics Source Code Library, record ascl:1506.005

Forbes D. A. et al., 2018, Proc. R. Soc. A, 474, 20170616

Fouesneau M. et al., 2014, ApJ, 786, 117

Furlong M. et al., 2015, MNRAS, 450, 4486

Furlong M. et al., 2017, MNRAS, 465, 722

Galleti S., Federici L., Bellazzini M., Fusi Pecci F., Macrina S., 2004, A&A, 416, 917

Gieles M., 2009, MNRAS, 394, 2113

Gieles M., Portegies Zwart S. F., 2011, MNRAS, 410, L6

Gieles M., Bastian N., Lamers H. J. G. L. M., Mout J. N., 2005, A&A, 441, 949

Gieles M., Larsen S. S., Scheepmaker R. A., Bastian N., Haas M. R., Lamers H. J. G. L. M., 2006a, A&A, 446, L9

Gieles M., Larsen S. S., Bastian N., Stein I. T., 2006b, A&A, 450, 129

Ginsburg A., Kruijssen J. M. D., 2018, ApJ, 864, L17

Girardi L., Bressan A., Bertelli G., Chiosi C., 2000, A&AS, 141, 371

Gnedin O. Y., Hernquist L., Ostriker J. P., 1999, ApJ, 514, 109

Goddard Q. E., Bastian N., Kennicutt R. C., 2010, MNRAS, 405, 857

Haydon D. T., Kruijssen J. M. D., Hygate A. e. P. S., Schruba A., Krumholz M. R., Chevance M., Longmore S. N., 2018, preprint (arXiv:1810.10897)

Hollyhead K., Adamo A., Bastian N., Gieles M., Ryon J. E., 2016, MNRAS, 460, 2087

Hughes M. E., Pfeffer J., Martig M., Bastian N., Crain R. A., Kruijssen J. M. D., Reina-Campos M., 2019, MNRAS, 482, 2795

Johnson L. C. et al., 2015, ApJ, 802, 127

Johnson L. C. et al., 2016, ApJ, 827, 33

Johnson L. C. et al., 2017, ApJ, 839, 78

Jordán A. et al., 2007, ApJS, 171, 101

Kang Y., Bianchi L., Rey S.-C., 2009, ApJ, 703, 614

Kennicutt R. C., Jr., 1998, ApJ, 498, 541

Kennicutt R. C., Evans N. J., 2012, ARA&A, 50, 531

Kim J.-H. et al., 2018, MNRAS, 474, 4232

Kruijssen J. M. D., 2012, MNRAS, 426, 3008

Kruijssen J. M. D., 2014, Class. Quantum Gravity, 31, 244006

Kruijssen J. M. D., 2015, MNRAS, 454, 1658

Kruijssen J. M. D., Bastian N., 2016, MNRAS, 457, L24

Kruijssen J. M. D., Longmore S. N., 2014, MNRAS, 439, 3239

Kruijssen J. M. D., Pelupessy F. I., Lamers H. J. G. L. M., Portegies Zwart S. F., Icke V., 2011, MNRAS, 414, 1339

Kruijssen J. M. D., Pelupessy F. I., Lamers H. J. G. L. M., Portegies Zwart S. F., Bastian N., Icke V., 2012, MNRAS, 421, 1927

Kruijssen J. M. D., Pfeffer J. L., Crain R. A., Bastian N., 2019a, MNRAS, 486, 3134

Kruijssen J. M. D., Pfeffer J. L., Reina-Campos M., Crain R. A., Bastian N., 2019b, MNRAS, 486, 3180
Krumholz M. R., McKee C. F., 2005, ApJ, 630, 250
Krumholz M. R., McKee C. F., Bland-Hawthorn J., 2019 , ARA&A, 57, 227
Lagos C. d. P. et al., 2015, MNRAS, 452, 3815
Lagos C. d. P. et al., 2016, MNRAS, 459, 2632
Lamers H. J. G. L. M., 2009, Ap&SS, 324, 183
Lamers H. J. G. L. M., Gieles M., Bastian N., Baumgardt H., Kharchenko N. V., Portegies Zwart S., 2005, A&A, 441, 117
Larsen S. S., 2002, AJ, 124, 1393
Larsen S. S., 2009, A&A, 494, 539
Larsen S. S., Richtler T., 1999, A&A, 345, 59
Larsen S. S., Richtler T., 2000, A&A, 354, 836
Lewis A. R. et al., 2015, ApJ, 805, 183
Li Y., Mac Low M.-M., Klessen R. S., 2004, ApJ, 614, L29
Li H., Gnedin O. Y., Gnedin N. Y., Meng X., Semenov V. A., Kravtsov A. V., 2017, ApJ, 834, 69
Longmore S. N. et al., 2014, in Beuther H., Klessen R. S., Dullemond C. P., Henning T., eds, Protostars and Planets VI. University of Arizona Press, Tucson, p. 291
Maji M., Zhu Q., Li Y., Charlton J., Hernquist L., Knebe A., 2017, ApJ, 844, 108
Marasco A., Crain R. A., Schaye J., Bahé Y. M., van der Hulst T., Theuns T., Bower R. G., 2016, MNRAS, 461, 2630
Marigo P., Girardi L., 2007, A&A, 469, 239
Marigo P., Girardi L., Bressan A., Groenewegen M. A. T., Silva L., Granato G. L., 2008, A&A, 482, 883
McConnachie A. W., Irwin M. J., Ferguson A. M. N., Ibata R. A., Lewis G. F., Tanvir N., 2005, MNRAS, 356, 979
McCrady N., Graham J. R., 2007, ApJ, 663, 844
Mengel S., Tacconi-Garman L. E., 2007, A&A, 466, 151
Messa M. et al., 2018a, MNRAS, 473, 996
Messa M. et al., 2018b, MNRAS, 477, 1683
Miholics M., Kruijssen J. M. D., Sills A., 2017, MNRAS, 470, 1421
Oppenheimer B. D. et al., 2016, MNRAS, 460, 2157
Pelupessy F. I., van der Werf P. P., Icke V., 2004, A&A, 422, 55
Pfeffer J., Kruijssen J. M. D., Crain R. A., Bastian N., 2018, MNRAS, 475, 4309
Pfeffer J., Bastian N., Crain R. A., Kruijssen J. M. D., Hughes M. E., Reina-Campos M., 2019, MNRAS, 487, 4550
Planck Collaboration et al., 2014, A&A, 571, A1
Portegies Zwart S. F., McMillan S. L. W., Gieles M., 2010, ARA&A, 48, 431
Prieto J. L., Gnedin O. Y., 2008, ApJ, 689, 919
Rahmati A., Schaye J., Bower R. G., Crain R. A., Furlong M., Schaller M., Theuns T., 2015, MNRAS, 452, 2034
Rahmati A., Schaye J., Crain R. A., Oppenheimer B. D., Schaller M., Theuns T., 2016, MNRAS, 459, 310
Reina-Campos M., Kruijssen J. M. D., 2017, MNRAS, 469, 1282
Reina-Campos M., Kruijssen J. M. D., Pfeffer J., Bastian N., Crain R. A., 2018, MNRAS, 481, 2851
Reina-Campos M., Kruijssen J. M. D., Pfeffer J. L., Bastian N., Crain R. A., 2019, MNRAS, 486, 5838
Renaud F., Bournaud F., Duc P.-A., 2015, MNRAS, 446, 2038
Rosas-Guevara Y. M. et al., 2015, MNRAS, 454, 1038
Ryon J. E. et al., 2014, AJ, 148, 33
Salpeter E. E., 1955, ApJ, 121, 161
Sánchez-Blázquez P. et al., 2006, MNRAS, 371, 703
Schaye J., 2004, ApJ, 609, 667
Schaye J., Dalla Vecchia C., 2008, MNRAS, 383, 1210
Schaye J. et al., 2015, MNRAS, 446, 521
Schechter P., 1976, ApJ, 203, 297
Schiminovich D. et al., 2007, ApJS, 173, 315
Silva-Villa E., Larsen S. S., 2011, A&A, 529, A25
Silva-Villa E., Adamo A., Bastian N., 2013, MNRAS, 436, L69
Silva-Villa E., Adamo A., Bastian N., Fouesneau M., Zackrisson E., 2014, MNRAS, 440, L116
Springel V., 2005, MNRAS, 364, 1105
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Toomre A., 1964, ApJ, 139, 1217
Trayford J. W. et al., 2015, MNRAS, 452, 2879
Trujillo-Gomez S., Reina-Campos M., Kruijssen J. M. D., 2019, MNRAS, 488, 3972
Turner M. L., Schaye J., Crain R. A., Theuns T., Wendt M., 2016, MNRAS, 462, 2440
Turner M. L., Schaye J., Crain R. A., Rudie G., Steidel C. C., Strom A., Theuns T., 2017, MNRAS, 471, 690
Usher C., Pfeffer J., Bastian N., Kruijssen J. M. D., Crain R. A., Reina-Campos M., 2018, MNRAS, 480, 3279
Ward J. L., Kruijssen J. M. D., 2018, MNRAS, 475, 5659
Weidner C., Kroupa P., Larsen S. S., 2004, MNRAS, 350, 1503
Whitmore B. C., Chandar R., Fall S. M., 2007, AJ, 133, 1067
Whitmore B. C., Chandar R., Bowers A. S., Larsen S., Lindsay K., Ansari A., Evans J., 2014, AJ, 147, 78
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
Zhang Q., Fall S. M., 1999, ApJ, 527, L81

APPENDIX A: RELATIONSHIPS BETWEEN THE CFE, PRESSURE, AND Σ_{SFR}

In Fig. A1 we compare the relationships between the CFE, natal gas pressure, and Σ_{SFR} for the simulations. The top panel of the figure shows the CFE as a function of the natal gas pressure, which is the fundamental relationship underpinning the models. The galaxies closely trace the relation between CFE and pressure used at the particle level in the simulations (dash-dotted line in the figure; see Pfeffer et al. 2018), with a small amount of scatter caused by pressure fluctuations within the galaxies. In this parameter space, metallicity plays no role and galaxies of all metallicities follow the relation.

In the middle panel of Fig. A1 we again compare the CFE as a function of Σ_{SFR} as in Fig. 2, but with symbols coloured by metallicity. The scatter in this parameter space is significantly larger than for CFE versus pressure (top panel), demonstrating that scatter in the relation between Σ_{SFR} and pressure largely determines the scatter in the CFE–Σ_{SFR} relation. We show this explicitly in the bottom panel of Fig. A1, where we compare the natal gas pressure with Σ_{SFR} for the galaxies. The black dashed line in this panel shows the expected relationship, which describes the star formation relation used in the EAGLE model (Schaye & Dalla Vecchia 2008; Schaye et al. 2015). At $\Sigma_{SFR} \gtrsim 10^{-2}\,\mathrm{M_\odot\,yr^{-1}\,kpc^{-2}}$, most galaxies follow the Schaye & Dalla Vecchia (2008) relation. However at $\Sigma_{SFR} \lesssim 10^{-2.5}\,\mathrm{M_\odot\,yr^{-1}\,kpc^{-2}}$, the galaxies are systematically offset from the relation (and this is also the case for a smaller fraction of galaxies at higher Σ_{SFR}). This offset does not depend on the definition for R_{lim} (Section 2.2) and occurs for both limits of $1.5R_{1/2}$ and the radius containing 68 per cent of the recent star formation. This deviation is largely caused by highly concentrated or substructured star formation, such that star formation occurs in a much smaller area compared to the area for which Σ_{SFR} is calculated. The inverse is not the case (too high pressures for a given Σ_{SFR}), since smaller apertures can always be chosen to bring the galaxies into better agreement with the expected P–Σ_{SFR} relation (at the expense of star particle and cluster numbers). Such an effect can be mitigated by calculating a mass-weighted surface density (see Johnson et al. 2016). In Fig. A2, we show the natal gas pressure compared with the mass-weighted Σ_{SFR} for the simulated galaxies (computed within R_{lim}, as for Fig. A1). We calculate
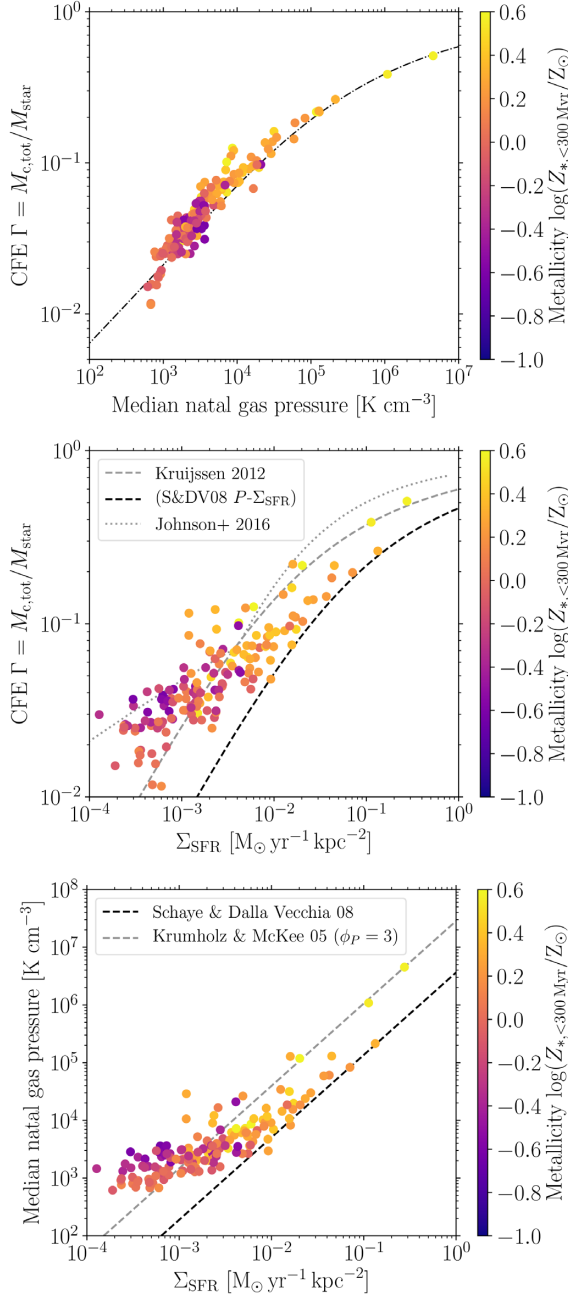
Figure A1. Relationships between the CFE, median natal gas pressure of star formation, and star formation rate surface density (Σ_{SFR}) of the simulated galaxies, for star particles younger than 300 Myr. Symbols in each panel are coloured by the median metallicity of star particles younger than 300 Myr. *Top:* CFE as a function of the median natal gas pressure. The dash-dotted line shows the relation used at the particle level in the simulations (see figure 3 of Pfeffer et al. 2018). *Middle:* CFE as a function of the star formation rate surface density (Σ_{SFR}). The line styles are as in Fig. 2, showing the predictions of the fiducial Kruijssen (2012) model (grey dashed line), the relation shifted to match the pressure–Σ_{SFR} relation adopted in the EAGLE simulations (Schaye & Dalla Vecchia 2008; Schaye et al. 2015) and the same model with a modified Σ_{gas}–Σ_{SFR} relation (Johnson et al. 2016) (grey dotted line). *Bottom:* The relationship between the median natal pressure and Σ_{SFR} for the galaxies. The grey dashed line shows the relation adopted for the star formation law in EAGLE (i.e. the expected relation for the simulations Schaye & Dalla Vecchia 2008), while the black dashed line shows the relation adopted in the fiducial Kruijssen (2012) CFE model (assuming $\phi_P = 3$; Krumholz & McKee 2005).
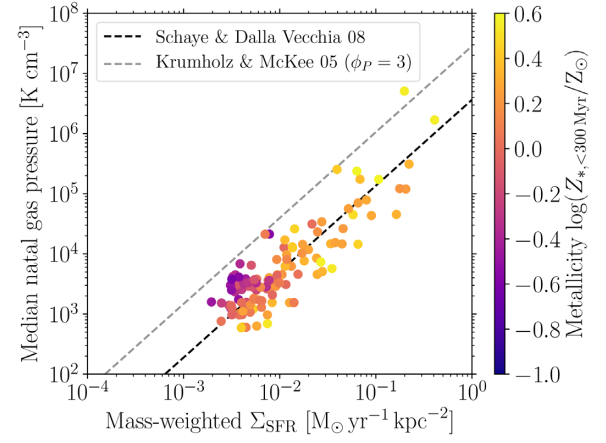

Figure A2. The relationship between the median natal pressure and Σ_{SFR} computed using a mass-weighted surface density. The mass-weighted Σ_{SFR} is calculated in a grid with regions of $0.7 \times 0.7\,\mathrm{kpc}^2$. The dashed lines showing the Schaye & Dalla Vecchia (2008) and Krumholz & McKee (2005) relations are as in the lower panel of Fig. A1.

the mass-weighted Σ_{SFR} in a grid of $0.7 \times 0.7\,\mathrm{kpc}^2$ regions (large enough such that incomplete sampling of star-forming regions is not important; Kruijssen & Longmore 2014), with each region weighted by the mass of young ($< 300\,\mathrm{Myr}$) stars. Using a mass-weighted Σ_{SFR} brings the low pressure ($P/k \lesssim 5 \times 10^3\,\mathrm{K\,cm^{-3}}$) galaxies in line with the expected P–Σ_{SFR} relation (with the exception of low-metallicity galaxies, see below), demonstrating the effectiveness of the method in accounting for spatially non-uniform star formation in the galaxies.

A secondary effect causing offset in the pressure–Σ_{SFR} relation is due to the metallicity of star formation. The EAGLE model includes a metallicity-dependent density threshold for star formation, such that star formation must occur at higher densities at lower metallicity (Schaye et al. 2015). This threshold is included to model the effect of the thermogravitational collapse of warm, photoionized interstellar gas into a cold, dense phase, which is expected to occur at lower densities and pressures in metal-rich gas (Schaye 2004). This higher density threshold at lower metallicities results (through the lower density limit for star formation imposed by the polytropic equation-of-state implemented at high gas densities) in higher pressures of star formation at a given Σ_{SFR}, which is evident at $\Sigma_{\mathrm{SFR}} \lesssim 5 \times 10^{-3}\,\mathrm{M_\odot\,yr^{-1}\,kpc^{-2}}$ in the lower panel of Figs A1 and A2. The threshold is also responsible for the apparent pressure floor in the figures (at $P/k \sim 10^3\,\mathrm{K\,cm^{-3}}$). Note that this effect of increasing pressure with metallicity is not expected to occur at high Σ_{SFR}, since the densities of star-forming gas in this regime are well above the metallicity-dependent threshold. The effect is also more apparent at higher redshifts in the simulations, where the metallicities of star-forming gas in the galaxies are lower.

In the bottom panel of Fig. A1, we also show the pressure–Σ_{SFR} adopted in the fiducial Kruijssen (2012) model (Krumholz & McKee 2005). This relation is offset to lower Σ_{SFR} by ≈ 0.6 dex when compared to the relation used in the EAGLE model. In the middle panel of Fig. A1, we also show the Kruijssen (2012) relation shifted to higher Σ_{SFR} to account for this offset (black dashed line). The fiducial (grey dashed line) and shifted (black dashed line) relations for the Kruijssen (2012) model thus give an indication of how the uncertainty in the pressure–Σ_{SFR} relation results in an uncertainty in the CFE–Σ_{SFR} relation.
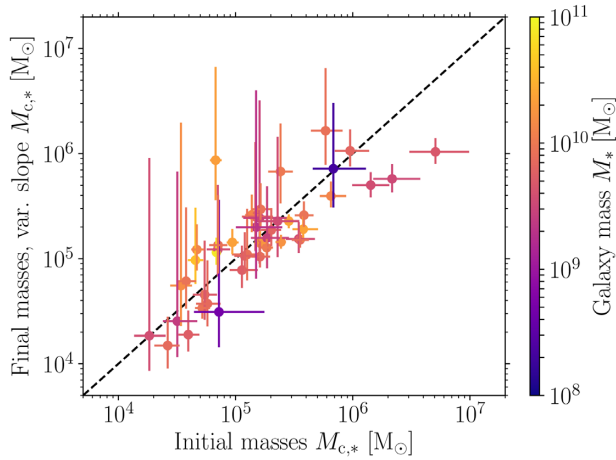
Figure B1. Cluster mass function truncation, $M_{c,*}$, when fitting initial versus final (evolved) cluster masses. The points show the median of the posterior distribution from the MCMC fit, while errorbars show the 16 and 84 per cent confidence intervals. Initial cluster masses are fit with a Schechter (1976) function with a constant lower mass power-law index of $\beta = -2$. Final cluster masses are fit with a Schechter function with a variable lower mass power-law index. Galaxies from both the $z = 0$ and $z = 0.5$ snapshots are included to increase galaxy numbers. The dashed line shows the one-to-one relation.

APPENDIX B: TRUNCATIONS OF INITIAL AND FINAL CLUSTER MASS FUNCTIONS

In Fig. B1, we test how our fits of the cluster mass function truncation are affected by cluster mass-loss in old, evolved cluster populations. We compare the upper exponential truncation mass ($M_{c,*}$) resulting from Schechter (1976) function fits to the initial and evolved cluster populations in the simulated galaxies. For fits to the initial cluster masses, we fix the lower mass power-law index to $\beta = -2$ (the initial value set in the simulations). For fits to the final

cluster masses, we use a prior on the index β between -3 and -0.5. The figure only shows galaxies where the most massive cluster is more massive than $M_{c,*}$ (the median of the posterior distribution), for both initial and final cluster mass fits (see the discussion in Section 3.2). Overall, we find good agreement in $M_{c,*}$ between fitting initial and evolved cluster masses, with potentially an offset to higher initial $M_{c,*}$. However, a small offset to higher initial $M_{c,*}$ is expected simply due to stellar-evolutionary mass-loss (a factor $\sim$1.3–1.4 for clusters with ages 100–300 Myr, or 0.1 dex, based on the mass-loss in the simulations). A number of galaxies have large errorbars for the fits to the final cluster masses due to the degeneracy between the truncation mass $M_{c,*}$ and index β.

APPENDIX C: RELATION OF ϕ_P TO GALACTIC PROPERTIES

In Section 3.2 we find that, at a given Σ_{SFR}, $M_{c,*}$ is elevated above the $z = 0$ relation at higher redshifts. We suggest that this effect may be caused by the higher gas fractions in galaxies at higher redshifts, through the parameter ϕ_P (Krumholz & McKee 2005), and therefore in this section we directly test that suggestion.

In Fig. C1, we compare the median local ϕ_P for recently formed stars (< 300 Myr) with Σ_{SFR} for all galaxies in Figs 4 and 5. We find that ϕ_P does not directly correlate with Σ_{SFR}, but instead with the gas fraction, which we show in Fig. C2. Therefore, at high redshift or low Σ_{SFR}, we typically find $\phi_P \approx 1$, since the galaxies have low (stellar) mass with high gas fractions. The median ϕ_P increases from $\phi_P \approx 1$ at $z = 2$ to $\phi_P \approx 2.5$ at $z = 0$, implying a factor of four increase in the Toomre mass (since $M_T \propto \phi_P^{-1.5}$, see equations 6 and 7 in Pfeffer et al. 2018), and thus in $M_{c,*}$ (at the same pressure and epicyclic frequency κ). We show this effect in Fig. 5 as dashed lines, by scaling the fit to observed local galaxies (Johnson et al. 2016) by the median ϕ_P at each redshift. Therefore, the decrease of the typical ϕ_P with increasing redshift explains the elevation of $M_{c,*}$ at higher redshifts found for the simulations.
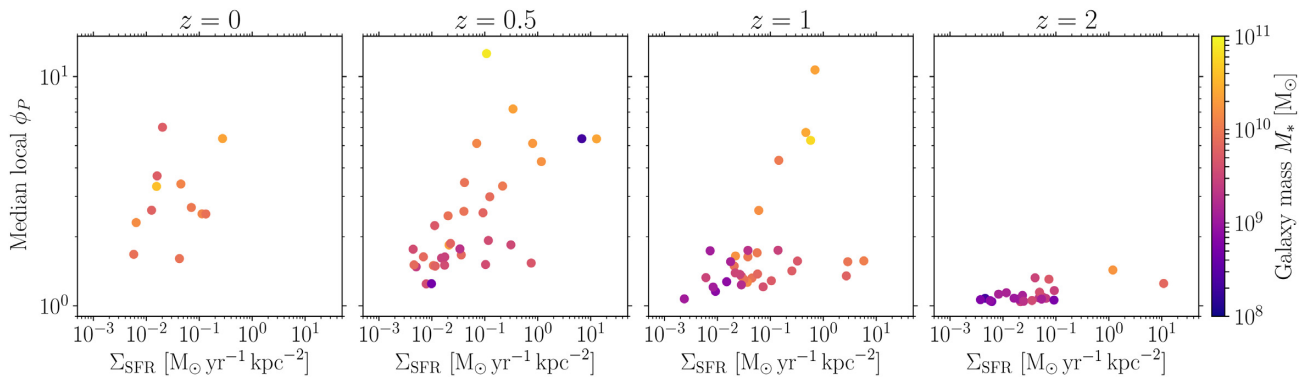

Figure C1. ϕ_P as a function of Σ_{SFR} at different redshifts. Variations in ϕ_P become more important at lower redshifts and at high Σ_{SFR}. At $z = \{0, 0.5, 1, 2\}$ we find medians of $\phi_P = \{2.5, 1.8, 1.3, 1.1\}$, respectively.
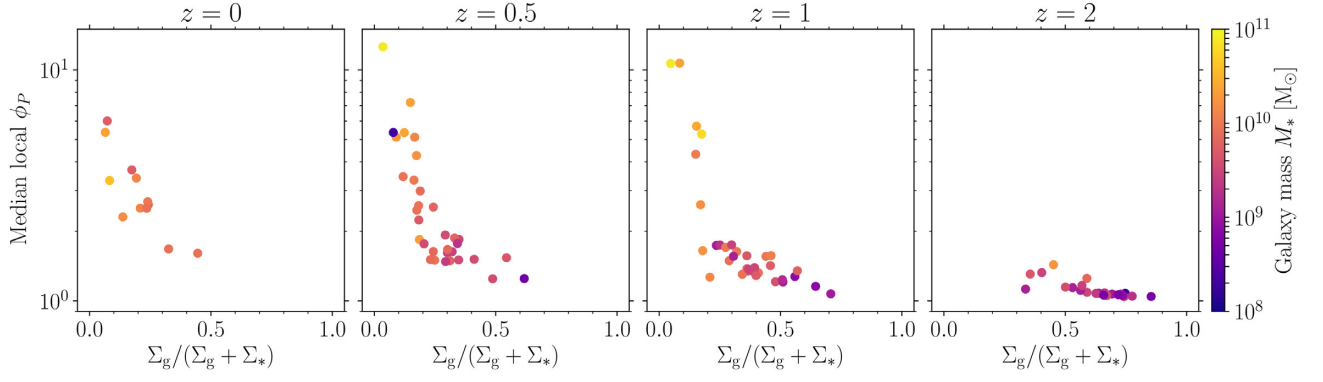
Figure C2. ϕ_P as a function of the gas mass fraction $\Sigma_{\mathrm{g}}/(\Sigma_{\mathrm{g}} + \Sigma_*)$ at redshifts $z = \{0, 0.5, 1, 2\}$. ϕ_P increases at lower redshifts as stellar densities increase relative to gas densities.

This paper has been typeset from a TeX/LaTeX file prepared by the author.