Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **461,** 1760–1785 (2016)
Advance Access publication 2016 June 7

doi:10.1093/mnras/stw1318

Galaxy assembly, stellar feedback and metal enrichment: the view from the GAEA model

Michaela Hirschmann,[1,2]★ Gabriella De Lucia[2]★ and Fabio Fontanot[2]

[1]*Sorbonne Universités, UPMC-CNRS, UMR7095, Institut d'Astrophysique de Paris, F-75014 Paris, France*
[2]*INAF – Astronomical Observatory of Trieste, via G.B. Tiepolo 11, I-34143 Trieste, Italy*

Accepted 2016 June 1. Received 2016 May 27; in original form 2015 December 14

ABSTRACT

One major problem of current theoretical models of galaxy formation is given by their inability to reproduce the apparently 'anti-hierarchical' evolution of galaxy assembly: massive galaxies appear to be in place since $z \sim 3$, while a significant increase of the number densities of low-mass galaxies is measured with decreasing redshift. In this work, we perform a systematic analysis of the influence of different stellar feedback schemes, carried out in the framework of GAEA, a new semi-analytic model of galaxy formation. It includes a self-consistent treatment for the timings of gas, metal and energy recycling, and for the chemical yields. We show this to be crucial to use observational measurements of the metallicity as independent and powerful constraints for the adopted feedback schemes. The observed trends can be reproduced in the framework of either a strong ejective or preventive feedback model. In the former case, the gas ejection rate must decrease significantly with cosmic time (as suggested by parametrizations of the cosmological 'FIRE' simulations). Irrespective of the feedback scheme used, our successful models always imply that up to 60–70 per cent of the baryons reside in an 'ejected' reservoir and are unavailable for cooling at high redshift. The same schemes predict physical properties of model galaxies (e.g. gas content, colour, age, and metallicity) that are in much better agreement with observational data than our fiducial model. The overall fraction of passive galaxies is found to be primarily determined by internal physical processes, with environment playing a secondary role.

Key words: methods: numerical – galaxies: abundances – galaxies: evolution – galaxies: formation – galaxies: high-redshift – galaxies: stellar content.

1 INTRODUCTION

Within the framework of the currently favoured dark energy dominated dark matter paradigm (ΛCDM), large-scale structure develops from small-scale density fluctuations, in a bottom–up fashion driven by gravitational forces (Blumenthal et al. 1984). The baryonic component follows the evolution of dark matter: gas that remains trapped in the potential wells of dark matter haloes is shock heated to the virial temperature, and later cools and condenses in a disc, where it forms stars. These eventually explode as supernovae (SN), heating the surrounding gas and driving spectacular galactic outflows (Rees & Ostriker 1977; White & Rees 1978). Our understanding of the various baryonic processes at play is far from complete, and complicated by the fact that they are entangled in a complex network of actions, back-reactions and self-regulations.

One naive expectation of the hierarchical scenario is that the assembly history of galaxies should parallel that of their parent dark matter haloes: most massive galaxies should assemble *later* than their lower mass counterparts, and this should manifest in a significant evolution of the galaxy stellar mass function (GSMF). In fact, a number of observational studies have highlighted an 'anti-hierarchical' (because in marked contrast with these expectations) trend in the evolution of this basic metric (e.g. Fontana et al. 2004; Pozzetti et al. 2007; Marchesini et al. 2009; Muzzin et al. 2013 and references therein): the low-mass end of the observed GSMF appears to evolve more rapidly than the high mass end, i.e. the most massive galaxies assemble *earlier* than their lower mass counterparts. While the discrepancy at the massive end is heavily affected by uncertainties in the measurements of galaxy stellar masses, that at low masses is robust and has proved to be of difficult solution in the framework of hierarchical model of galaxy formation (e.g. Fontanot et al. 2009; Weinmann et al. 2012). When including strong stellar feedback, hierarchical models are able to reproduce the observed GSMF in the local Universe, but systematically overpredict the number densities of sub-M_* galaxies at higher redshift(e.g. Guo et al. 2011, see also fig. 4 in Somerville & Davé 2015). What appears to be needed is a mechanism that is able to decouple the growth of low-mass galaxies (that occurs late) from that of their host dark matter haloes (that occurs early). The obvious

★E-mail: hirschma@iap.fr (MH); delucia@oats.inaf.it (GDL)

suspect is *feedback*, from massive stars and/or from active galactic nuclei (AGN).

Observational signatures in broad emission/absorption lines of AGN (e.g. Moe et al. 2009; Dunn et al. 2010; Maiolino et al. 2012; Cicone et al. 2014; Harrison et al. 2014; Brusa et al. 2015) and cavities in the hot, X-ray emitting gas surrounding radio galaxies (e.g. McNamara & Nulsen 2007; Cattaneo et al. 2009; McNamara, Rohanizadegan & Nulsen 2011; Fabian 2012) highlight the presence of radiative and mechanical feedback from central accreting black holes. These powerful sources of energy have long been suspected to play an important role in the 'cooling-flow' problem. Early models accounted for this form of feedback by simply assuming that cooling would be suppressed above some critical halo mass or circular velocity (e.g. Kauffmann et al. 1999; De Lucia, Kauffmann & White 2004). More sophisticated schemes have been implemented in the last decade, both in semi-analytic models of galaxy formation and hydrodynamical simulations (e.g. Springel, Di Matteo & Hernquist 2005a; Bower et al. 2006; Croton 2006; Choi et al. 2015; Steinborn et al. 2015 and references therein). While these are successful (under certain conditions) in offsetting the cooling flow in massive haloes and lead to massive central galaxies dominated by old stellar populations, some problems remain. E.g. the predicted AGN activity levels of local galaxies exhibit trends as a function of stellar mass and halo mass that are in contrast with those measured (Fontanot et al. 2011), and the metallicities predicted for the most massive galaxies are significantly lower than those observed (De Lucia & Borgani 2012). Some cosmological simulations also tend to overestimate the massive end of the GSMF due to an inefficient suppression of star formation at the centre of galaxy groups and clusters (Puchwein & Springel 2013; Genel et al. 2014; Hirschmann et al. 2014a; Khandai et al. 2015).

Observations also reveal large, supernova-driven, bi-polar galactic outflows with velocities up to $500\,\mathrm{km\,s^{-1}}$ in nearby galaxies (e.g. Heckman, Armus & Miley 1990; Heckman et al. 2000; Martin 2005; Veilleux, Cecil & Bland-Hawthorn 2005; Rubin et al. 2014), and in high-z star-forming objects (e.g. Steidel et al. 1996; Pettini et al. 2000; Weiner et al. 2009; Martin et al. 2013; Newman et al. 2013). These stellar-driven winds are expected to play an important role in regulating star formation in galaxies at and below the exponential cut-off of the GSMF. The energy release originates from a combination of exploding SN type II and Ia, and from massive stars in form of ionizing radiation, radiation pressure and stellar winds. The details of the energetics and outflows, as well as the fate of the outflowing gas, are not well understood neither from the observational, nor from the theoretical point of view. As a consequence, modelling of this particular physical process is necessarily schematic. Nevertheless, theoretical studies show that galactic winds affect significantly a number of galaxy physical properties such as their star formation rates (SFRs), metal and gaseous content, size, gas kinematics as well as the stellar mass assembly (ratio of *in situ* formed to accreted stars), and the baryonic and metal content of the intergalactic and intracluster medium (e.g. De Lucia et al. 2004; Davé, Oppenheimer & Finlator 2011; Hirschmann et al. 2012a; Lackner et al. 2012; Aumer et al. 2013; Hopkins et al. 2014; Hirschmann et al. 2015).

It is interesting to note that the above mentioned overproduction of sub-M_* galaxies is connected to a long-standing problem in hydrodynamical simulations: gas cools efficiently at high redshift and in small compact haloes. The originating clumps merge via dynamical friction leading to a significant transfer of angular momentum from baryons to the dark matter. This results in spiral galaxies with relatively large bulges and compact discs (the 'angular momentum catastrophe'). Significant effort has been devoted to improve the adopted stellar feedback scheme in hydrodynamical simulations (e.g. Aumer et al. 2013; Stinson et al. 2013; Hopkins et al. 2014; Schaye et al. 2015; Murante et al. 2015). These studies are now able to obtain baryon conversion efficiencies close to those expected for $\sim M_*$ galaxies. Very recent cosmological hydrodynamical simulations have also been successfully 'calibrated' to reproduce the observed GSMF in the local Universe (e.g. the Illustris and the EAGLE simulations – Genel et al. 2014; Furlong et al. 2015. The latter also successfully reproduce the observed evolution of the GSMF).

Significant efforts to improve the adopted stellar feedback scheme have also been made in the framework of semi-analytic models of galaxy formation (Knebe et al. 2015), and a few schemes have been shown to successfully reproduce the observed evolution of the GSMF (Henriques et al. 2013; Lagos, Lacey & Baugh 2013; Lu, Mo & Wechsler 2015; White, Somerville & Ferguson 2015). Albeit representing important improvements with respect to previous results, none of these models appears to be completely satisfactory with respect to e.g. the predicted colour distributions (Henriques et al. 2013) or metal evolution (White et al. 2015). All these models are based on the instantaneous recycling approximation, where energy and metals are assumed to be released immediately after star formation, i.e. neglecting the finite lifetime of stars and its dependence on stellar mass. This represents an important limitation, and prevents the use of observational data on the gaseous and stellar metallicity content of galaxies as independent constraints.

In this work, we will carry out a systematic analysis of the influence of different stellar feedback models, with a focus on the mechanisms/schemes required to reproduce the observed evolution of the GSMF. Our work is based on an evolution of the semi-analytic model presented in De Lucia & Blaizot (2007), and includes the detailed chemical enrichment scheme introduced in De Lucia et al. (2014b). Although close to the latest developments of the L-Galaxies model (Guo et al. 2011; Henriques et al. 2015) in terms of basic structure and modelling for some specific physical processes, our model has differentiated significantly from this parallel branch for the treatment of various processes and modification of the chemical enrichment and feedback scheme. To emphasize these differences, we will refer in the following to our updated model as the GAEA[1] (GAlaxy Evolution and Assembly) model.[2]

Our paper is structured as follows. In Section 2 we describe the main features of our GAEA model, and describe in detail the different feedback schemes that we have tested. Sections 3 and 4 focus on the impact of stellar feedback on the predicted evolution of the GSMF, and analyse the corresponding predictions as for the ejection and recycling rates of gas. We then discuss the influence of the same feedback schemes on the gaseous and metal content of galaxies, colour and SFR distributions, and stellar populations in Sections 5, 6 and 7. Finally, we discuss our results in Section 8, and summarize our findings in Section 9.

2 THE GAEA GALAXY FORMATION MODEL

Our galaxy formation model corresponds to that described in De Lucia et al. (2014b), with modifications of the stellar feedback

[1] Named after Gaea, one of the Greek primordial deities. Personification of the Earth and creator of the Universe.

[2] Note that galaxy catalogues for the GAEA model will be made publicly available under http://gavo.mpa-garching.mpg.de/Millennium/

scheme and of some of the model parameters, as described in the following sections. The model is based on that described in detail by De Lucia & Blaizot (2007, and references therein), but has been modified to follow more accurately processes on the scale of the Milky Way satellites as described in De Lucia & Helmi (2008) and Li, De Lucia & Helmi (2010). In particular, our model assumes an earlier re-ionization: it is assumed to start at $z = 15$ and to last for about 0.12 Gyr, while in the De Lucia & Blaizot model it was assumed to start at $z = 8$ and to last approximately for the same amount of time. In addition, we assume that gas cooling is suppressed below 10^4 K, while in the original model, haloes with virial temperature lower than 10^4 K were able to cool as much gas as a 10^4 K halo with the same metallicity. Finally, we assume that for galaxies residing in haloes with virial mass below $5 \times 10^{10}\,\mathrm{M}_\odot$, most of the new metals (95 per cent) are ejected directly into the hot gas phase, while in the De Lucia & Blaizot model all new metals were immediately mixed with the cold gas in the disc. The models discussed in this paper, also differ from that adopted in De Lucia & Blaizot (2007) for the modelling assumed to estimate disc radii and the critical surface density below which no star formation takes place. For details, we refer to De Lucia & Helmi (2008). We stress that, while these modifications are important at the scale of Milky Way satellites (Li et al. 2010), they have little or no influence on the results shown in this study.

In the GAEA model, the evolution of baryons is traced in four different reservoirs, the stellar component of galaxies, the cold gas in galaxy discs, the hot gas associated with the dark matter haloes, and the ejected gas component. To model the mass and energy transfer between these different reservoirs, the GAEA model includes physically and/or observationally motivated prescriptions for gas cooling, re-ionization, star formation, stellar feedback and gas recycling, metal evolution, black hole growth, AGN feedback, disc instabilities and environmental effects. For most of these prescriptions, we follow the same approach as in De Lucia & Blaizot (2007) except for the modifications discussed above, and with the major exception of chemical enrichment (as described in De Lucia et al. 2014b), and stellar feedback and gas recycling (which are the subject of this study). Below, we describe the simulations used in this study, give a summary of our chemical enrichment model, and describe in detail the stellar feedback schemes tested in this work.

2.1 Dark matter simulation and cosmology

In this study, we take advantage of the dark matter merger trees extracted from the Millennium Simulation (Springel et al. 2005b). The simulation assumes a *WMAP*1 cosmology with $\Omega_\Lambda = 0.75$, $\Omega_\mathrm{m} = 0.25$, $\Omega_\mathrm{b} = 0.045$, n=1, $\sigma_8 = 0.9$, and h=0.73. To test the convergence of our results with resolution, we also run the GAEA model over the merger trees of the Millennium-II Simulation (Boylan-Kolchin et al. 2009). This simulation adopts the same cosmology of the Millennium, but corresponds to a smaller box ($100\,\mathrm{Mpc}\,h^{-1}$ against $500\,\mathrm{Mpc}\,h^{-1}$), five times better spatial resolution (the Plummer equivalent softening of the Millennium II is $1.0\,\mathrm{kpc}\,h^{-1}$) and 125 times better mass resolution (the particle mass is $6.9 \times 10^6\,\mathrm{M}_\odot h^{-1}$). Convergence tests are presented in Appendix B.

Note that more recent measurements of the cosmological parameters suggest a lower value for σ_8 (the latest results from the Planck Collaboration suggest a value of 0.831 ± 0.013). This means that the local Universe from the Millennium and Millennium II simulations could be more evolved than the real one (see e.g. Wang et al. 2008). We have explicitly tested to what extent our results change when adopting a lower value for 8σ. In particular, we used the simulation described in Wang et al. (2008) that adopts $8\sigma = 0.7$. Confirming results from that paper (see also Guo et al. 2013; Gonzalez-Perez et al. 2014), we find that model parameters need to be re-tuned to reproduce the same level of agreement at $z = 0$, but results are qualitatively the same.

2.2 The chemical evolution scheme

The interstellar medium (ISM) of galaxies is enriched by metals released by stellar winds and by SN explosions. The large majority of published work is based on SAMs that assume an instantaneous recycling approximation, i.e. they neglect the dependence of stellar lifetimes on stellar mass, and assume that gas recycling and chemical enrichment occur 'instantaneously' (i.e. within the individual time-step of the code). Our GAEA model adopts the chemical enrichment scheme described in De Lucia et al. (2014b), explicitly accounting for the finite lifetimes of stars, and tracing individual element abundances. While there are some uncertainties in the adopted chemical yields and SN Ia model, relaxing the instantaneous recycling approximation prevents us from considering the metal yield as a free adjustable parameter of the model. As we will discuss in the following, this allows us to use observational data on chemical abundances as independent constraints, and therefore draw more robust conclusions on the gas and metal recycling scheme.

For details on the implementation, we refer to De Lucia et al. (2014b). The model assumes a Chabrier IMF (Chabrier 2003) and stellar lifetimes by Padovani & Matteucci (1993). Stars with masses below $8\,\mathrm{M}_\odot$ are assumed to enrich the ISM mainly through AGB winds, while more massive ones are expected to die as SNII. In our chemical enrichment scheme, the only free parameter is represented by the realization probability of a given SNIa scenario, and is constrained by the observed Fe content of the Milky Way disc (we refer to De Lucia et al. 2014b for details). Our fiducial model assumes a delay time distribution (DTD) corresponding to the single degenerate scenario described in Matteucci & Recchi (2001) and Bonaparte et al. (2013), and the metal yields of Thielemann et al. (2003) for SNIa, of Chieffi & Limongi (2002) for SNII, and those of Karakas (2010) for winds from low- and intermediate-mass stars.

De Lucia et al. (2014b) have applied this chemical scheme to a set of simulated Milky Way size haloes (those completed within the Aquarius project – Springel et al. 2008). They also discuss the dependence on various model ingredients, and the differences with respect to the same physical model assuming an instantaneous recycling approximation. Model predictions were shown to be in fair agreement with the observed chemical composition of the Milky Way, in particular of its disc component, and of its satellites.

When applying this model to a cosmological volume, we found that we had to adjust some model parameters in order to match the normalization of the mass–metallicity relation as measured from the Sloan Digital Sky Survey (SDSS). We discuss how these adjustments affect predictions for Milky Way like galaxies in appendix A. In a recent paper, Yates et al. (2013) have independently included an updated chemical enrichment model into the semi-analytic model by Guo et al. (2011). They show that their implementation is able to reproduce simultaneously the chemical content of Milky Way-like galaxies and the observed mass–metallicity relation in the local Universe. They do not show, however, how their updated chemical model affects some basic predictions like the galaxy luminosity or stellar mass function and their evolution. As the same total amount of feedback energy is required as in Guo et al. (2011), the model presented in Yates et al. reproduces the galaxy stellar mass function

Table 1. Overview of the stellar feedback models and of the corresponding free parameters (in bold face) for gas reheating, ejection and re-incorporation, tested in the framework of the GAEA model. In the PREH model, gas that is pre-heated is added to the 'ejected' reservoir. We refer to Section 2.3 for details.

Model (colour)	Re-heating [$\dot{M}_{\rm reheat}$]	Ejection [$\dot{M}_{\rm eject}$]	Re-incorporation [$\dot{M}_{\rm reinc}$]
1. Fiducial (lila)	$\frac{4}{3}\epsilon_{\rm reheat}\cdot\frac{E_{\rm tot,SN}}{V_{\rm vir}^2}\cdot\dot{M}_{\rm star}$, $\boldsymbol{\epsilon_{\rm reheat}=0.02}$	Centrals: all re-heated gas ejected, Satellites: re-heated gas added to hot halo gas	$\gamma\frac{M_{\rm eject}}{t_{\rm dyn}}$, $\boldsymbol{\gamma=0.1}$
2. Lag13 (dark blue)	$\epsilon_{\rm reheat}\left(\frac{\Sigma_{\rm gas}}{1600\,{\rm M_\odot pc^{-2}}}\right)^{-0.6}\cdot\left(\frac{f_{\rm cold}}{0.12}\right)^{0.8}\dot{M}_{\rm star}$, $\boldsymbol{\epsilon_{\rm reheat}=0.2}$	All re-heated gas ejected	$\gamma\frac{M_{\rm eject}}{t_{\rm dyn}}$, $\boldsymbol{\gamma=0.1}$
3. Hop12 (light blue)	$\epsilon_{\rm reheat}\cdot\left(\frac{\Sigma_{\rm gas}(r)}{10\,{\rm M_\odot pc^{-2}}}\right)^{-0.5}\left(\frac{V_{\rm circ}(r)}{100\,{\rm km\,s^{-1}}}\right)^{-1.1}\cdot\dot{M}_{\rm star}$, $\boldsymbol{\epsilon_{\rm reheat}=1.0}$	Centrals: all re-heated gas ejected, Satellites: re-heated gas added to hot halo gas	$\gamma\frac{M_{\rm eject}}{t_{\rm dyn}}$, $\boldsymbol{\gamma=0.1}$
4. Guo11 (turquoise)	$\epsilon_{\rm reheat}\cdot\left[0.5+\left(\frac{V_{\rm max}}{70\,{\rm km\,s^{-1}}}\right)^{-3.5}\right]\cdot\dot{M}_{\rm star}$, $\boldsymbol{\epsilon_{\rm reheat}=0.7}$	$\frac{\dot{E}_{\rm FB}-0.5\dot{M}_{\rm reheat}V_{\rm vir}^2}{0.5V_{\rm vir}^2}$ with $\dot{E}_{\rm FB}=\epsilon_{\rm eject}\cdot\left[0.5+\left(\frac{V_{\rm max}}{70\,{\rm km\,s^{-1}}}\right)^{-3.5}\right]\cdot 0.5\dot{M}_{\rm star}V_{\rm SN}^2$, $\boldsymbol{\epsilon_{\rm eject}=0.35}$	$\gamma\frac{M_{\rm eject}}{t_{\rm dyn}}$, $\boldsymbol{\gamma=0.5}$
5. Hen13 (green)	$\epsilon_{\rm reheat}\cdot\left[0.5+\left(\frac{V_{\rm max}}{336\,{\rm km\,s^{-1}}}\right)^{-0.46}\right]\cdot\dot{M}_{\rm star}$, $\boldsymbol{\epsilon_{\rm reheat}=1.0}$	As in Guo11 with $\dot{E}_{\rm FB}=\epsilon_{\rm eject}\cdot\left[0.5+\left(\frac{V_{\rm max}}{405\,{\rm km\,s^{-1}}}\right)^{-0.92}\right]\cdot 0.5\dot{M}_{\rm star}V_{\rm SN}^2$, $\boldsymbol{\epsilon_{\rm eject}=0.25}$	$\gamma\frac{M_{\rm eject}}{t_{\rm reinc}}$, with $t_{\rm reinc}=\frac{10^{10}\,{\rm M_\odot}}{M_{\rm vir}}$, $\boldsymbol{\gamma=1.0}$
6. zDEP (yellow)	$\epsilon_{\rm reheat}\cdot\left[0.5+(1+z)^3\left(\frac{V_{\rm max}}{70\,{\rm km\,s^{-1}}}\right)^{-3.5}\right]\cdot\dot{M}_{\rm star}$, $\boldsymbol{\epsilon_{\rm reheat}=0.7}$	As in Guo11 with $\dot{E}_{\rm FB}=\epsilon_{\rm eject}\cdot\left[0.5+(1+z)^3\left(\frac{V_{\rm max}}{70\,{\rm km\,s^{-1}}}\right)^{-3.5}\right]\cdot 0.5\dot{M}_{\rm star}V_{\rm SN}^2$, $\boldsymbol{\epsilon_{\rm eject}=0.15}$	$\gamma\frac{M_{\rm eject}}{t_{\rm dyn}}$, $\boldsymbol{\gamma=0.5}$
7. FIRE (orange)	$\epsilon_{\rm reheat}(1+z)^{1.25}\cdot\left(\frac{V_{\rm max}}{60\,{\rm km\,s^{-1}}}\right)^{\alpha}\cdot\dot{M}_{\rm star}$, $V_{\rm max}<60\,{\rm km\,s^{-1}}\rightarrow\alpha=-3.2$, $V_{\rm max}>60\,{\rm km\,s^{-1}}\rightarrow\alpha=-1.0$, $\boldsymbol{\epsilon_{\rm reheat}=0.3}$	As in Guo11 with $\dot{E}_{\rm FB}=\epsilon_{\rm eject}(1+z)^{1.25}\cdot\left(\frac{V_{\rm max}}{60\,{\rm km\,s^{-1}}}\right)^{\alpha}\cdot 0.5\dot{M}_{\rm star}V_{\rm SN}^2$, $\boldsymbol{\epsilon_{\rm eject}=0.1}$	$\gamma\frac{M_{\rm eject}}{t_{\rm reinc}}$, with $t_{\rm reinc}=\frac{10^{10}\,{\rm M_\odot}}{M_{\rm vir}}$, $\boldsymbol{\gamma=1.0}$
8. PREH (red)	$\frac{4}{3}\epsilon_{\rm reheat}\cdot\frac{E_{\rm tot,SN}}{V_{\rm vir}^2}\cdot\dot{M}_{\rm star}$, $\boldsymbol{\epsilon_{\rm reheat}=0.02}$	$\dot{M}_{\rm gas,preheat}+\dot{M}_{\rm reheat}$ with $\dot{M}_{\rm gas,preheat}=(1-f_{\rm preheat})f_{\rm bar}\dot{M}_{\rm DM,infall}$, $f_{\rm preheat}=\frac{M_{\rm vir}}{M_{\rm crit}}$, $M_{\rm crit}=10^{12}\,{\rm M_\odot}$;	$\gamma\frac{M_{\rm eject}}{t_{\rm reinc}}$, with $t_{\rm reinc}=\frac{10^{10}\,{\rm M_\odot}}{M_{\rm vir}}$, $\boldsymbol{\gamma=0.5}$

at $z=0$ (Yates, private communication), but the expectation is that it over-predicts the number density of low and intermediate mass galaxies at higher redshift (as the Guo et al. model does). As we will show in the following, we have identified a stellar feedback scheme that is able to match the local galaxy mass function and its evolution to higher redshift, as well as the metallicity content of galaxies, down to the scale of the ultra-faint dwarfs of the Milky Way.

2.3 Stellar feedback models

The energy supplied by massive stars in the form of SN and stellar winds represents the engine that drives the galactic scale outflows observed in actively star forming galaxies, both in the local Universe and at high redshift (see Section 1). Unfortunately, the observational measurements available refer to material that is still relatively deep within the gravitational potential well of the halo so that it is difficult to translate the estimated outflow rates into rates at which mass, metals, and energy escape from the galaxies and are transported into the intergalactic medium. As a consequence, galaxy formation models rely on analytic parametrizations or 'sub-grid' models based on our best understanding of the theory and/or on observational measurements.

In order to test the influence of different scalings, we have implemented eight different schemes for stellar feedback and gas recycling in our GAEA model. Four of these have been introduced in the framework of previously published semi-analytic models. For each scheme, we have varied the stellar feedback (re-heating and ejection) and reincorporation efficiencies (see Table 1) so as to match the exponential cut-off of the stellar mass function at $z=0$, and by simultaneously trying to obtain a good match with the observed mass–metallicity relation in the local Universe and the measured evolution of the galaxy stellar mass function at higher redshift. All other model parameters have been left unchanged (so they correspond to those used in De Lucia et al. 2014b and references therein). As mentioned above, we have modified some model parameters for our fiducial model (see Appendix A) with respect to De Lucia et al. (2014b) to recover the correct normalization of the mass–metallicity relation. We want to emphasize that for our fiducial model, we were not able to find any combination of parameters to match the observed GSMF and its evolution, highlighting the difficulty in reproducing the anti-hierarchical trend in galaxy stellar mass growth, which is

the main motivation of this work to explore different models for stellar feedback.

Table 1 summarizes the different stellar feedback schemes discussed and analysed in this work, and lists the corresponding parameter values for the stellar feedback and reincorporation efficiencies. As detailed below, we assume that stellar winds and SN are able to reheat some fraction of the cold gas in the disc. This 'reheated' gas is assumed to leave the galaxy, but remains bound to its parent halo and becomes part of its hot gas component, unless the energy is large enough to escape the potential well of the halo. This 'ejected' component is stored in a reservoir, and is not available for cooling until it is 're-incorporated' into the hot gas component. A feedback scheme is therefore fully characterized by a reheating rate, an ejection rate and a re-incorporation rate. Each rate is characterized by some efficiency ($\epsilon_{\rm reheat}$ and $\epsilon_{\rm eject}$ for the reheating and ejection rates) or factor (γ for the re-incorporation). In the following, we describe in detail each of the schemes tested in this work. In all cases, we assume that when galaxies are accreted on larger systems (i.e. when they become satellites), they are instantaneously stripped of their hot and ejected gas reservoirs. The latter is added to the ejected component associated with the central galaxy. As a consequence, gas re-incorporation is only modelled on to central galaxies of haloes.

2.3.1 *Energy-driven winds due to SNae (fiducial)*

The 'fiducial' feedback scheme implemented in De Lucia et al. (2014b) corresponds to the 'energy-driven' scheme described in De Lucia et al. (2004). The reheated gas mass rate $\dot{M}_{\rm reheat}$ is calculated as:

$$\dot{M}_{\rm reheat} = \frac{4}{3}\epsilon_{\rm reheat}\frac{E_{\rm tot,SN}}{V_{\rm vir}^2} \times \dot{M}_{\rm star}, \quad (1)$$

where $\dot{M}_{\rm star}$ is the SFR in one individual time-step of integration, $V_{\rm vir}$ the virial velocity of the halo (subhalo for satellite galaxies), $\epsilon_{\rm reheat}$ the re-heating efficiency, and $E_{\rm tot,SN}$ the total amount of energy released by SNIa and SNII. The latter quantity is self-consistently calculated within our non instantaneous recycling scheme, i.e. accounting for finite stellar lifetimes. We assume that all reheated gas of central galaxies is ejected out of the halo ($\dot{M}_{\rm eject} = \dot{M}_{\rm reheat}$), while the reheated gas of satellite galaxies is added to the hot gas content of the parent halo (that can only cool on to the central galaxy). We have verified that varying the exponent in the denominator of equation (1) with values from -1 (e.g. assuming 'momentum driven winds') to -4 hardly changes the results discussed in the paper for this particular model.

A certain fraction of the ejected gas mass is assumed to be re-accreted back on to the (parent) halo on a halo dynamical time-scale $t_{\rm dyn} = R_{\rm vir}/V_{\rm vir}$ ($R_{\rm vir}$ and $V_{\rm vir}$ are the virial radius and velocity of the parent halo, respectively):

$$\dot{M}_{\rm reinc} = \gamma\frac{M_{\rm eject}}{t_{\rm dyn}}, \quad (2)$$

where γ is an adimensional parameter.

As mentioned above, we had to modify the model parameters with respect to those adopted in De Lucia et al. (2014b) in order to match the observed metallicity content of galaxies in the local Universe. Specifically, we had to slightly reduce the feedback efficiency (0.02 against 0.05 used in De Lucia et al. 2014b) and the re-incorporation factor (0.1 against 0.5). In Appendix A, we discuss how these parameter modifications affect results for Milky Way-like galaxies.

2.3.2 *Propagation of pressurized bubbles due to SNae (Lag13)*

Lagos et al. (2013) discuss a dynamical model for the propagation and evolution of pressurized SNae bubbles in a multiphase ISM. In particular, they derive parametrizations for the mass loading of the cold gas, i.e. the amount of cold gas that is expelled out of the disc and is thus temporarily not available for cooling. Specifically, they argue that the often assumed dependence on the circular velocity or virial velocity alone is a poor description of the outflow process. Instead, they propose a dependence on the gas surface density $\Sigma_{\rm gas}$ and on the cold gas fraction $f_{\rm cold}$ of the galaxy. In particular, their parametrization of the outflow rate (reheating rate in our terminology) reads as:

$$\dot{M}_{\rm reheat} = \epsilon_{\rm reheat}\left(\frac{\Sigma_{\rm gas}}{1600\,{\rm M_\odot pc^{-2}}}\right)^{-0.6}\left(\frac{f_{\rm cold}}{0.12}\right)^{0.8}\dot{M}_{\rm star}, \quad (3)$$

where $\epsilon_{\rm reheat}$ is the reheating efficiency. The gas surface density is computed as follows:

$$\Sigma_{\rm Gas} = \frac{M_{\rm cold}}{\pi R_{\rm gas}^2} \text{ with } R_{\rm gas} = \frac{J_{\rm gas}}{2M_{\rm cold}V_{\rm max}}, \quad (4)$$

where $R_{\rm gas}$ is the scale radius of the gaseous disc, $M_{\rm cold}$ the cold gas mass, $J_{\rm gas}$ the angular momentum of the gas, and $V_{\rm max}$ the maximum circular velocity. We assume the angular momentum of the gas and the maximum circular velocity to be equal to those of the parent dark matter (sub)haloes (the same assumptions are made for other models that use the same quantities). Following the implementation of this model in the GALFORM semi-analytic code (Lagos et al. 2013), we assume the reheated gas (both from central and satellite galaxies) to be ejected so that it is not available for cooling until it is again reincorporated into the hot gas component. We model gas re-incorporation as in the fiducial model described above (this implies that the re-incorporated gas will be able to cool only on the central galaxy at later times, i.e. gas ejected from satellite galaxies will never be available to the same galaxies at any later time). Lagos et al. (2013) tested this feedback scheme in the framework of the GALFORM model, and found that it predicts optical and NIR luminosity functions in good agreement with data, both at low and high redshifts, in particular with respect to the shallow measured faint end slopes.

2.3.3 *Stellar-driven galactic-scale winds parametrized from isolated galaxy simulations (Hop12)*

Hopkins, Quataert & Murray (2011) and Hopkins, Quataert & Murray (2012a) performed high-resolution simulations of isolated galaxies including a new model for stellar feedback. The key physical processes considered include heating from both type I and type II supernovae, stellar winds from AGBs, heating from the shocked stellar winds, H II photoionization, and radiation pressure. In Hopkins, Quataert & Murray (2012b), they parametrize the resulting gas outflow rates, finding they are well approximated by a momentum-driven wind scaling ($\sim V_{\rm circ}^{-1.1}$) with an additional dependence on the gas surface density $\Sigma_{\rm gas}$. The gas re-heating rate $\dot{M}_{\rm reheat}$ (gas ejection out of the galaxy), depending on the radial distance from the galaxy centre r, is given by:

$$\dot{M}_{\rm reheat} = \epsilon_{\rm reheat}\left(\frac{\Sigma_{\rm gas}(r)}{10\,{\rm M_\odot pc^{-2}}}\right)^{-0.5}\left(\frac{V_{\rm circ}(r)}{100\,{\rm km\,s^{-1}}}\right)^{-1.1}\dot{M}_{\rm star}, \quad (5)$$

where $\epsilon_{\rm reheat}$ is the re-heating efficiency. For simplicity, we use the maximum circular velocity and the gas surface density measured

at the scale radius of the gaseous disc, as computed for the Lag13 model (see equation 4). Like in the fiducial model, we assume that all gas re-heated from centrals galaxies leaves the halo and is temporarily unavailable for cooling, while re-heated gas from satellite galaxies is assumed to remain associated with the hot gas within the parent dark matter halo. Gas re-incorporation is modelled as in the fiducial model. This particular model has not yet been tested in the framework of any semi-analytic model.

2.3.4 *Strong stellar feedback for dwarf galaxies (Guo11)*

The galaxy formation model of Guo et al. (2011) is based on that of Croton (2006) and De Lucia & Blaizot (2007), with some important modifications relative to the modelling of satellite galaxies, and stellar feedback. In order to reproduce the shallow slope of the observed present-day stellar mass function, these authors modified the stellar feedback scheme of De Lucia & Blaizot (2007) allowing higher ejection efficiencies in dwarf galaxies. In particular, the reheated gas mass rate $\dot{M}_{\rm reheat}$ is computed as:

$$\dot{M}_{\rm reheat} = \epsilon_{\rm reheat}\left[0.5 + \left(\frac{V_{\rm max}}{70\ {\rm km\,s^{-1}}}\right)^{-3.5}\right] \times \dot{M}_{\rm star}, \quad (6)$$

The rate of energy injection by massive stars into disc and halo gas is parametrized as:

$$\dot{E}_{\rm FB} = \epsilon_{\rm eject}\left[0.5 + \left(\frac{V_{\rm max}}{70\ {\rm km\,s^{-1}}}\right)^{-3.5}\right] \times 0.5\dot{M}_{\rm star}V_{\rm SN}^2, \quad (7)$$

where $0.5V_{\rm SN}^2$ is the mean kinetic energy of SN ejecta per unit mass of stars formed. In our model, the latter quantity is self-consistently calculated following our non instantaneous chemical enrichment scheme. The additional scaling with $(0.5 + V_{\rm max}^{-3.5})$ is motivated by the observation that dwarf galaxies have lower metallicities and less dust than their more massive counterparts. Therefore, it is plausible that radiative losses during the thermalization of ejecta are substantially smaller than in more massive systems.

Given this energy input into the disc and halo gas, the total amount of material that can be escape the halo can be estimated following energy conservation arguments:

$$\dot{M}_{\rm eject} = \frac{\dot{E}_{\rm FB} - 0.5\dot{M}_{\rm reheat}V_{\rm vir}^2}{0.5V_{\rm vir}^2}, \quad (8)$$

where $V_{\rm vir}$ is the escape velocity from the parent halo (or subhalo for satellite galaxies).

An important difference of this feedback scheme compared to the fiducial model, is that all reheated gas from centrals and satellites is initially added to the hot gas component of the parent halo. Then, the ejection rate is computed from this updated hot gas reservoir, that also contains gas heated by e.g. AGN feedback and newly infalling gas. As consequence, the ejected gas can be larger than the reheated (by stellar feedback alone) gas. As we will see below, this will have important consequences on the predicted SFRs.

In the context of our GAEA model, we adopt the same gas re-incorporation scheme as in our fiducial model and use the same feedback parameters as in Guo11, except for the reheating efficiency $\epsilon_{\rm reheat}$, which is slightly reduced in order to match the observed stellar metallicity content of galaxies in the local Universe.

2.3.5 *Halo-mass dependent scaling for the gas re-incorporation (Hen13)*

Henriques et al. (2013) use a Monte Carlo Markov Chain approach to explore the parameter space of the Guo et al. (2011) model, focusing on matching the observed evolution of the GSMF. They find that no simple modification of the parameter set is able to reproduce the observed galaxy number densities from $z = 0$ to $z \sim 3$. The suggested solution is a modification of the time-scale for gas re-incorporation that scales with halo mass. Specifically, they use the same parametrization for the reheating and ejection rate adopted in Guo et al. (2011) with different parameters:

$$\dot{M}_{\rm reheat} = \epsilon_{\rm reheat}\left[0.5 + \left(\frac{V_{\rm max}}{336\ {\rm km\,s^{-1}}}\right)^{-0.46}\right] \times \dot{M}_{\rm star}, \quad (9)$$

$$\dot{E}_{\rm FB} = \epsilon_{\rm eject}\left[0.5 + \left(\frac{V_{\rm max}}{405\ {\rm km\,s^{-1}}}\right)^{-0.92}\right] \times 0.5\dot{M}_{\rm star}V_{\rm SN}^2. \quad (10)$$

The time-scale for gas re-incorporation is assumed to depend on the inverse of halo mass:

$$M_{\rm reinc} = \gamma\frac{M_{\rm eject}}{t_{\rm reinc}},\ \text{with}\ t_{\rm reinc} = \frac{10^{10}\,{\rm M_\odot}}{M_{\rm vir}} \times yr. \quad (11)$$

This model reproduces the observed evolution of the stellar mass and luminosity functions from $z = 0$ to $z \sim 3$, by construction.

2.3.6 *Early stellar feedback (zDEP)*

Stinson et al. (2013) and Kannan et al. (2014) introduced the so-called early stellar feedback, likely originating from radiation pressure from young massive stars at high redshifts. Their cosmological simulations, carried out using the Gasoline code and including an SN feedback modelled in the form of 'blast-waves', were shown to successfully reproduce the cosmic evolution of baryon conversion efficiencies in low-mass haloes.

Inspired by the success of these sub-grid models and in an attempt to mimic their effect, we introduce an 'ad hoc' model, where we assume feedback efficiencies to depend on redshift (higher reheating and ejection rates are assumed at higher redshifts). Starting from the Guo11 feedback scheme, we implement the following parametrizations, by adding a dependence on $(1 + z)^3$ for both the ejection and re-heating efficiencies:

$$\dot{M}_{\rm reheat} = \epsilon_{\rm reheat}\left[0.5 + (1+z)^3\left(\frac{V_{\rm max}}{70\ {\rm km\,s^{-1}}}\right)^{-3.5}\right]\dot{M}_{\rm star}, \quad (12)$$

and

$$\dot{E}_{\rm FB} = \epsilon_{\rm eject}\left[0.5 + (1+z)^3\left(\frac{V_{\rm max}}{70\ {\rm km\,s^{-1}}}\right)^{-3.5}\right]0.5\dot{M}_{\rm star}V_{\rm SN}^2. \quad (13)$$

In this scheme, the reincorporation of gas is modelled as in the fiducial model. We stress that, although the proposed scenario is physically motivated, the adopted parametrizations are 'empirical', and obtained by forcing the model to reproduce the observed evolution of the low-mass end of the GSMF.

2.3.7 *Stellar feedback parametrized from cosmological zoom simulations (FIRE)*

Hopkins et al. (2014) developed a sophisticated sub-resolution model accounting for individual sources of stellar feedback in the

form of energy and momentum input from SN explosions, radiative feedback (photo-heating and radiation pressure), and stellar winds ('the Feedback In Realistic Environments' – FIRE). Performing fully cosmological simulations, the authors showed that these models are successful in reproducing the cosmic evolution of baryon conversion efficiencies in low-mass haloes. Their simulations reveal that the high-redshift galaxy evolution is dominated by strong bursts of star formation, followed by powerful and highly non-linear gusts of galactic outflows that sweep up large fractions of the gas of the interstellar and circumgalactic medium of galaxies. At low redshift, sufficiently massive galaxies switch into a continuous and quiescent mode of star formation that does not drive outflows. In a recent work, Muratov et al. (2015) have parametrized the outflow rates of gas in these simulations, finding an explicit dependence on redshift at a given circular velocity (see their equations 4 and 5), or equivalently, a dependence on stellar mass (see their eq. 8).

Adopting the parametrizations given in Muratov et al. (2015), we assume:

$$\dot{M}_{\rm reheat} = \epsilon_{\rm reheat}(1+z)^{1.25}\left(\frac{V_{\rm max}}{60\,{\rm km\,s^{-1}}}\right)^{\alpha} \times \dot{M}_{\rm star}. \quad (14)$$

For circular velocities $V_{\rm max} < 60\,{\rm km\,s^{-1}}$, the exponent α is found to be -3.2, while for circular velocities $V_{\rm max} > 60\,{\rm km\,s^{-1}}$, α has a value of -1.0. This means that the FIRE simulations predict 'momentum-driven' winds for massive galaxies, but 'stronger' winds for less massive ones.

We then assume:

$$\dot{E}_{\rm FB} = \epsilon_{\rm eject}(1+z)^{1.25}\left(\frac{V_{\rm max}}{60\,{\rm km\,s^{-1}}}\right)^{\alpha} \times 0.5\dot{M}_{\rm star}V_{\rm SN}^2, \quad (15)$$

and, as in the Guo11 model, we calculate the ejected gas mass from the hot gas reservoir following energy conservation arguments (equation 8). We have verified that the alternative parametrization given for gas outflows in Muratov et al. (2015), including an explicit dependence on stellar mass, leads to results very similar to those obtained with the above parametrization.

Muratov et al. (2015) have not quantified the time-scales over which gas is re-incorporated back on to the halo. We find that, in this feedback model, gas re-incorporation needs to be delayed as proposed in the Hen13 scheme in order to match the present-day GSMF (particularly around the exponential cut-off).

2.3.8 *Preventive feedback (PREH)*

The last feedback scheme we have implemented is based on what we call a 'preventive feedback' model. The circumgalactic/intergalactic medium may be pre-heated (to some level of entropy) by early feedback processes, such that the amount of infalling (for the first time) pristine gas is reduced in lower mass haloes with respect to the universal baryon fraction. Physical mechanisms for pre-heating are not well constrained: they are likely related to various phenomena including e.g. stellar/AGN-driven winds (Mo & Mao 2002, 2004) or intergalactic turbulence (Zhu, Feng & Fang 2011). Inspired by the work by Lu & Mo (2007) and Lu et al. (2015), we have tested the following scheme: we reduce the amount of infalling gas by a factor $f_{\rm preheat}$, linearly scaling with the halo virial mass $M_{\rm vir}$ for galaxies with halo masses below $M_{\rm crit} = 10^{12}\,{\rm M}_{\odot}$ at all redshifts:

$$f_{\rm preheat} = \frac{M_{\rm vir}}{M_{\rm crit}} \quad (16)$$

The rate of newly infalling gas $\dot{M}_{\rm gas,infall}$ is then given by:

$$\dot{M}_{\rm gas,infall} = f_{\rm preheat} \times f_{\rm bar} \times \dot{M}_{\rm DM,infall}, \quad (17)$$

where $f_{\rm bar}$ is the universal baryon fraction and $\dot{M}_{\rm DM,infall}$ the dark matter accretion rate. For estimating $f_{\rm preheat}$, we have not added any additional redshift dependence or more complex dependence on halo mass (as in Lu et al. 2015).

For simplicity, we add the amount pre-heated gas to the ejected gas component. Therefore, this gas will be accreted on to the halo at later times, following the re-incorporation model adopted in Hen13 (equation 11). Finally, in this scheme, we model the heating and ejection due to stellar feedback as in our fiducial scheme.

3 THE 'ANTI-HIERARCHICAL' EVOLUTION OF GALAXIES

In this section, we focus on a few basic predictions of our GAEA model, and illustrate to what extent they are affected by different parametrizations for stellar feedback and gas recycling. The trends discussed in this section, i.e. the evolution of the GSMF and of the luminosity function, are often referred to as 'anti-hierarchical', suggesting conflicts with expectations from the currently favoured cosmological model for structure formation.

3.1 The evolution of the galaxy stellar mass function

Fig. 1 shows the evolution of the GSMF for the different models used in this study (lines with different colours) from $z = 3$ to present (different panels), compared to various observational measurements (black symbols and dotted lines). Observational data clearly illustrate the 'anti-hierarchical' trend mentioned above: the massive end appears to be in place already at high redshifts (z = 2–3), while the number densities of low-mass galaxies increase significantly towards $z = 0$.

The fiducial model strongly overestimates the number density of low-mass galaxies, particularly at high redshifts, i.e. too many low-mass galaxies form at early times. This has been a well-known problem for galaxy formation models for at least one decade now (e.g. Fontana et al. 2004; Fontanot et al. 2009; Weinmann et al. 2012). Increasing the feedback efficiency does not improve the agreement with observational measurements in this model because it reduces the number density of galaxies around the knee of the mass function.

We find the same qualitative behaviour for the Lag13, Hop12 and Guo11 models, although these are in better agreement (by construction) with the observational measurements at $z = 0$. This means that the different dependencies on other physical quantities such as gas surface density, cold gas fraction or circular velocity, at least in the form of the patametrizations considered here, are not adequate to efficiently decouple the evolution of baryons from that of low-mass dark matter haloes, in particular at early times.

Assuming longer time-scales for gas re-cycling (Hen13, FIRE), a stronger energy input at early times, i.e. an explicit redshift dependence of the ejected and reheated gas (zDEP, FIRE), or some form of pre-heating (PREH), can significantly reduce the number density of low-mass galaxies ($<10^{11}\,{\rm M}_{\odot}$) by up to one order of magnitude, particularly at high redshifts. All these modifications are able to bring model predictions in fairly good agreement with the observed evolution of the GSMF over the redshift range considered.

We stress that a crucial element for the success of the feedback schemes Hen13, zDEP and FIRE is that large fractions of gas (larger than the amounts reheated by stellar feedback) can be ejected outside the haloes. When assuming a redshift dependent ejection rate and/or longer gas re-cycling time-scales in the framework of the fiducial feedback scheme, where the amount of gas that can be ejected is

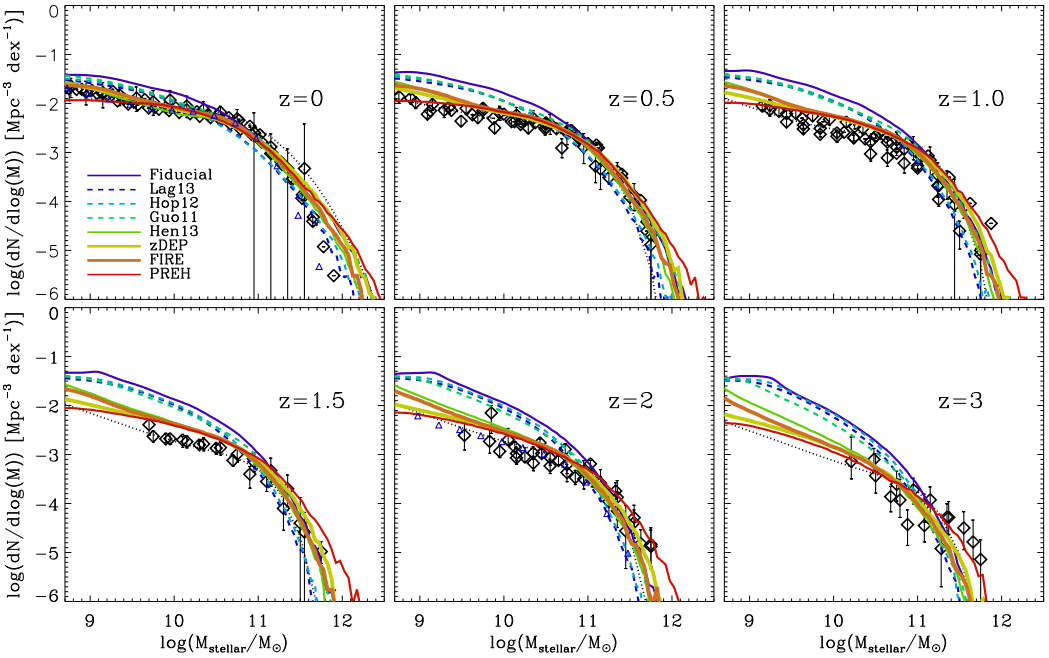

Figure 1. Redshift evolution of the GSMF for the different feedback models considered in this study (lila lines: fiducial, dark blue lines: Lag13, light blue lines: Hop12, turquoise lines: Guo11, green lines: Hen13, yellow: zDEP, orange: FIRE, red: PREH), compared to observational measurements (Bell et al. 2003; Drory et al. 2004; Bundy, Ellis & Conselice 2005; Fontana et al. 2006; Panter et al. 2007; Marchesini et al. 2007; Pérez-González et al. 2008; Ilbert et al. 2010; Muzzin et al. 2013, black symbols and black dotted lines). Both strong ejective feedback models (Hen13, zDEP, FIRE) and the preventive feedback scheme (PREH) are successful in reproducing the measured evolution of the low-mass end of the GSMF.

limited to that reheated (according to the energy driven formulation adopted), we do not reproduce the evolution of the GSMF.

Our results indicate that both a strong ejective (Hen13, zDEP, FIRE) and a preventive form (PREH) of feedback are capable of reproducing the observed evolution of the GSMF. To what extent these models are able to reproduce other galaxy physical properties, and how we can discriminate which scheme performs better, will be discussed below in Sections 5 and 7.

3.2 The evolution of the galaxy luminosity functions

It is well known that the conversion from observable properties (typically, luminosities in different bands) to stellar masses carries relatively large uncertainties, both statistical and (more dangerous because less under control) systematics (e.g. see discussion in Marchesini et al. 2009; De Lucia, Muzzin & Weinmann 2014a). On the other hand, galaxy formation models can predict observables, specifically luminosities, albeit making a number of assumptions (the same needed to convert observed luminosities into physical properties) on e.g. the stellar population models, the initial mass function, dust attenuation.

In Fig. 2, we show the *K*-band (top panels) and *V*-band (bottom panels) galaxy luminosity functions at $z = 0$, 1, 2 from the models (lines of different colours), compared to observational measurements by Cole et al. (2001); Kochanek et al. (2001); Pozzetti et al. (2003); Saracco et al. (2006); Caputi et al. (2007) and Marchesini et al. (2012) (black dotted and dash–dotted lines and grey shaded areas). Luminosities are computed assuming a Chabrier initial mass function (consistently with the choice adopted for the chemical enrichment) and stellar population synthesis models by Bruzual & Charlot (2003), as described in De Lucia et al. (2004).

At $z = 0$, the strong ejective feedback models (zDEP, FIRE, and Hen15) still slightly overestimate the number densities at the low-luminous end of the *K*-band luminosity function. The PREH model appears to provide a better match to the observational data. As for the galaxy stellar mass function, the Hen13, zDEP, FIRE and PREH models predict a relatively shallow faint end slope, in fairly good agreement with observations at $z = 1$ and $z = 2$.

For the *V*-band luminosity function (measured down to $M = -17$ at $z = 2$), the PREH model even tends to slightly underestimate the number densities of low-luminous galaxies at $z = 1$–2, while the Hen13 and FIRE models still slightly overpredict the number densities of low-luminosity galaxies at $z = 2$. The zDEP model provides the best match to the observed *V*-band luminosity function over the redshift range considered.

Turning to the bright end of the luminosity function, we find an underestimation of luminous galaxies in the Hop12 and Lag13 models, irrespective of the redshift. This is relevant around L_*. Only at the highest redshift considered ($z \sim 2$), these models predict number densities for the most luminous galaxies that are in agreement with observational determinations. All other models tend instead to overestimate the bright end, particularly for the *V*-band luminosity function. These overabundant *V*-band luminous galaxies in the models have typically experienced recent merger driven starbursts. It is likely that our treatment of dust attenuation is inadequate for these systems, which might explain the relatively good agreement with the observed number densities of the most luminous galaxies in the *K*-band, and the poorer agreement for the *V*-band.

3.3 The evolution of the baryon conversion efficiencies

Fig. 3 shows the redshift evolution of the baryon conversion efficiency versus the dark matter halo mass for the different feedback models used in this study (lines of different colours), compared to predictions from the subhalo abundance matching approach

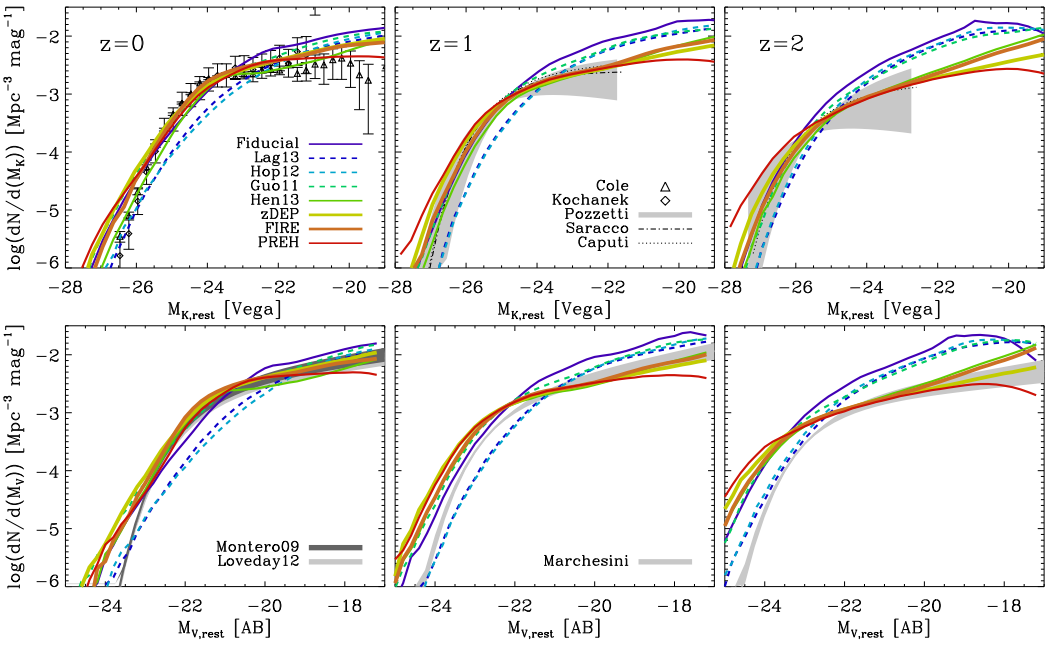

Figure 2. Evolution of the *K*-band (top row) and *V*-band luminosity function (bottom row) in different models (coloured lines as in Fig. 1), compared to observational measurements by Cole et al. (2001); Kochanek et al. (2001); Pozzetti et al. (2003); Saracco et al. (2006); Caputi et al. (2007) and Marchesini et al. (2012) (black symbols and grey shaded areas).

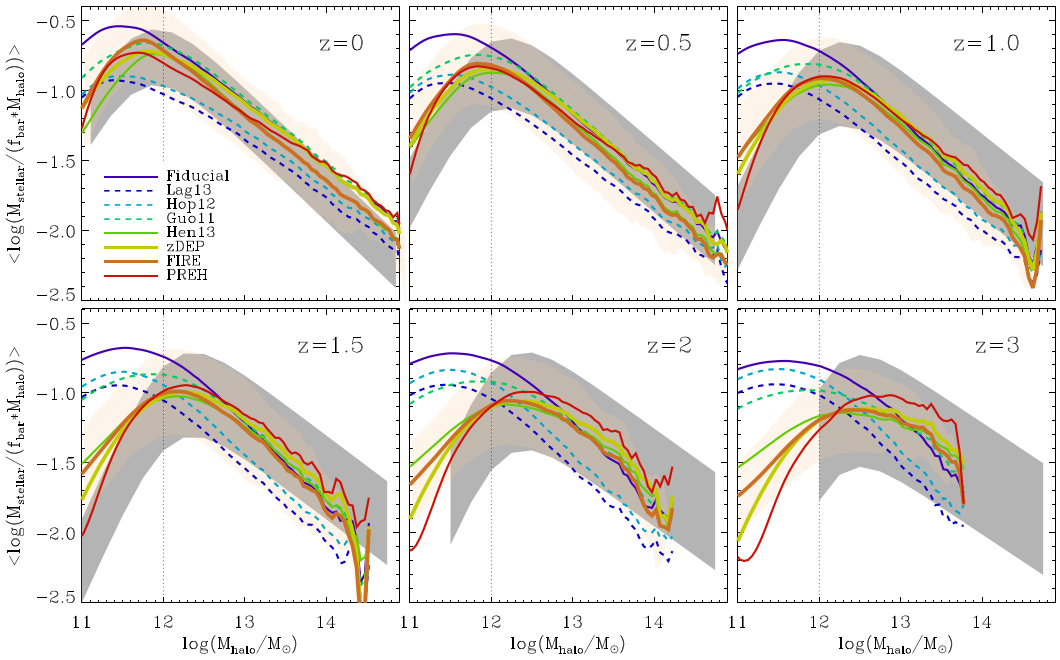

Figure 3. Mean Baryon conversion efficiencies ($M_{\rm stellar}/(f_{\rm bar} \times M_{\rm halo})$) versus halo mass at different redshifts, as predicted by the different models used in this study (coloured lines as in Fig. 1; the orange dashed area illustrates the 1σ-scatter corresponding to the FIRE model). The Hen13, zDEP, FIRE and PREH models (green, yellow, orange, red solid lines), that match the GSMF, are also consistent with predictions from the subhalo abundance matching method by Moster, Naab & White (2013) (grey shaded areas). The vertical dotted lines correspond to halo mass of $\sim 10^{12}\,{\rm M}_{\odot}$. At the resolution of the Millennium, these are resolved with less than 1000 particles.

discussed in Moster et al. (2013) (grey shaded areas, see also Behroozi, Wechsler & Conroy 2013). The baryon conversion efficiency is defined as the ratio between the total stellar mass $M_{\rm stellar}$ within a given halo and the total baryonic mass expected assuming a universal baryon fraction ($f_{\rm bar} \times M_{\rm halo}$ – we assume here $f_{\rm bar} = 0.18$). At the mass resolution of the Millennium simulation, haloes with mass $\sim 10^{12}\,{\rm M}_{\odot}$ (marked by vertical dotted lines in Fig. 3) are resolved with less than 1000 particles. For the figure shown, we are therefore pushing into the resolution limit of our simulation. We have, however, verified that results based on the Millennium II are consistent with those shown. These convergence tests are discussed in Appendix B.

The subhalo abundance matching method gives a maximum baryon conversion efficiency of 20–30 per cent, for halo masses of roughly $10^{12}\,M_{\odot}$. For higher and lower masses, the conversion efficiencies drop to much smaller values, down to only a few per cent. In addition, the peak of the baryon conversion efficiency is found to shift towards lower halo masses with decreasing redshift. In other words, the halo mass scale at which most of the star formation takes places increases at higher redshift.

Irrespective of the redshift, the fiducial, Lag13, Hop12 and Guo11 models predict by far too high conversion efficiencies in haloes less massive than $10^{12}\,M_{\odot}$. In addition, they cannot capture the correct time evolution of the peak: the fiducial, Lag13 and Hop12 models predict a maximum conversion efficiency at halo masses of $\sim 10^{11.5}\,M_{\odot}$, while the Guo11 model predicts a maximum conversion efficiency at $\sim 10^{12}\,M_{\odot}$. The peak hardly evolves with redshift in all these models. The Hen13, zDEP, FIRE and PREH models predict baryon conversion efficiencies in low-mass haloes that are in good agreement with expectations, and the correct evolution as a function of redshift. This confirms results from the previous sections: both strong ejective (Hen13, zDEP, FIRE) and preventive (PREH) feedback models can capture the observed trends. On the basis of the results presented so far, however, we cannot discriminate among the proposed solutions.

4 ORIGIN OF THE 'ANTI-HIERARCHICAL' TREND

In order to understand the origin of the 'anti-hierarchical' trends illustrated in the previous section, and the physical mechanisms responsible for the success or failure of the proposed feedback schemes, we analyse the circulation of baryons between the different phases (i.e. the amount of gas heated, ejected, re-incorporated or cooled and being converted into stars), for the different parametrizations adopted in this study.

4.1 Evolution of the gas outflows

The first and second row of Fig. 4 show the redshift evolution of the mean reheated and ejected gas mass rates (due to stellar feedback) for the different feedback schemes tested in this work (illustrated by lines of different colours). Results are shown for galaxies in four different bins of stellar mass (different columns). Since ejection and re-heating rates scale linearly with the SFRs (see equations summarized in the second and third column in Table 1), we have normalized them to these values so that the figures show what are usually referred to as 'mass-loading' factors. To compute the averages shown in the figure, we have selected galaxies on the basis of their galaxy stellar mass at $z = 0$, and traced their main progenitor (typically the most massive progenitor at each node of the merger tree) backwards in time.

In the strong ejective feedback models (Hen13, zDEP and FIRE), the mass-loading of the re-heated gas is by up to two orders of magnitude larger than in the fiducial, Lag13, Hop12, Guo11, and PREH models, particularly at redshift $z > 2$. As noted above, our preventive feedback scheme (PREH) adopts the same modelling for the reheating and ejection as our fiducial model. Therefore, the predicted evolution are very similar but both the reheating and the ejection rates are systematically lower in the former scheme, at all redshifts for low-mass galaxies, and at high redshift for more massive systems: less gas is available for cooling at high redshift in small haloes, so less stars are formed and less gas can be reheated.

When looking at the ejection rate of gas, the differences between the feedback schemes are even more extreme. For the fiducial, Lag13, Hop12 and the PREH model, the (ejection) mass loading is rather low and hardly evolves with redshift. Instead, for the Guo11, Hen13, zDEP and FIRE models, the mass-loading can be by up to three orders of magnitude larger. This is a direct consequence of the higher reheating rates, and is enhanced by the assumption that also a fraction of the hot gas associated with the parent halo (and not only the gas reheated from the cold disc component) can be driven out of the halo.

The FIRE and zDEP models are characterized by a very strong decrease of the mass-loading factor with time, resulting in significantly lower ejection rates at $z = 0$ than e.g. in the Hen13 model. In this model, the mass-loading is in fact high, but nearly constant as a function of cosmic time. We will come back to consequences of these different gas flows on e.g. the metal enrichment of model galaxies in Section 7. The results discussed so far indicate that high ejection rates at early times represent a possible (but not necessary) condition for reproducing the observed evolution of the GSFM.

4.2 Evolution of the gas re-accretion rate

The third row of Fig. 4 shows how the re-accretion gas rates (these are, in all models, proportional to the amount of gas in the ejected reservoir) vary as a function of redshift. As a direct consequence of the larger amount of ejected material in the Guo11, Hen13, zDEP and FIRE models, the mean gas re-incorporation rates are larger than in the fiducial, Lag13 and Hop12 models, but at low redshift for the most massive galaxies considered. Here, the re-accretion rates predicted by the FIRE model are comparable to, or even lower than those found in the fiducial scheme.

The high average re-accretion rates predicted by the PREH model are a result of our specific treatment for the gas that is prevented to fall on to dark matter haloes. For simplicity, we assign this gas to the same reservoir as the ejected component (both are not available for cooling). This obviously leads to high 're-accretion' rates although the term is inappropriate in this case since most of these baryons have never been inside the halo before.

The figure also shows the impact of the adopted gas recycling scheme: models assuming re-incorporation time-scales inversely proportional to halo mass (Hen13, FIRE, and PREH) predict a different redshift dependence of the gas re-accretion rates than the other models assuming the gas to be re-incorporated over a halo dynamical time-scale. Independently of the galaxy stellar mass considered, a halo mass dependence of the re-incorporation time-scales leads to a strong suppression of early gas re-accretion rates with respect to the fiducial reincorporation scheme. Gas re-accretion is delayed to progressively lower redshifts for galaxies of decreasing stellar mass, resembling the 'anti-hierarchical' trends discussed earlier. In contrast, models based on the fiducial gas re-incorporation scheme do not reproduce such a behaviour, and predict nearly constant re-incorporation rates as a function of redshift.

4.3 Evolution of the ejected baryon fraction

The results shown in the previous sections demonstrate that the model that successfully reproduce the measured evolution of the GSMF (Hen13, zDEP, FIRE, and PREH) are characterized by rather different behaviours of the ejection and re-incorporation rates.

One crucial, common feature of these models is the existence of large amounts of baryons in the ejected component, where they are not available to cooling. The fourth row of Fig. 4 shows the

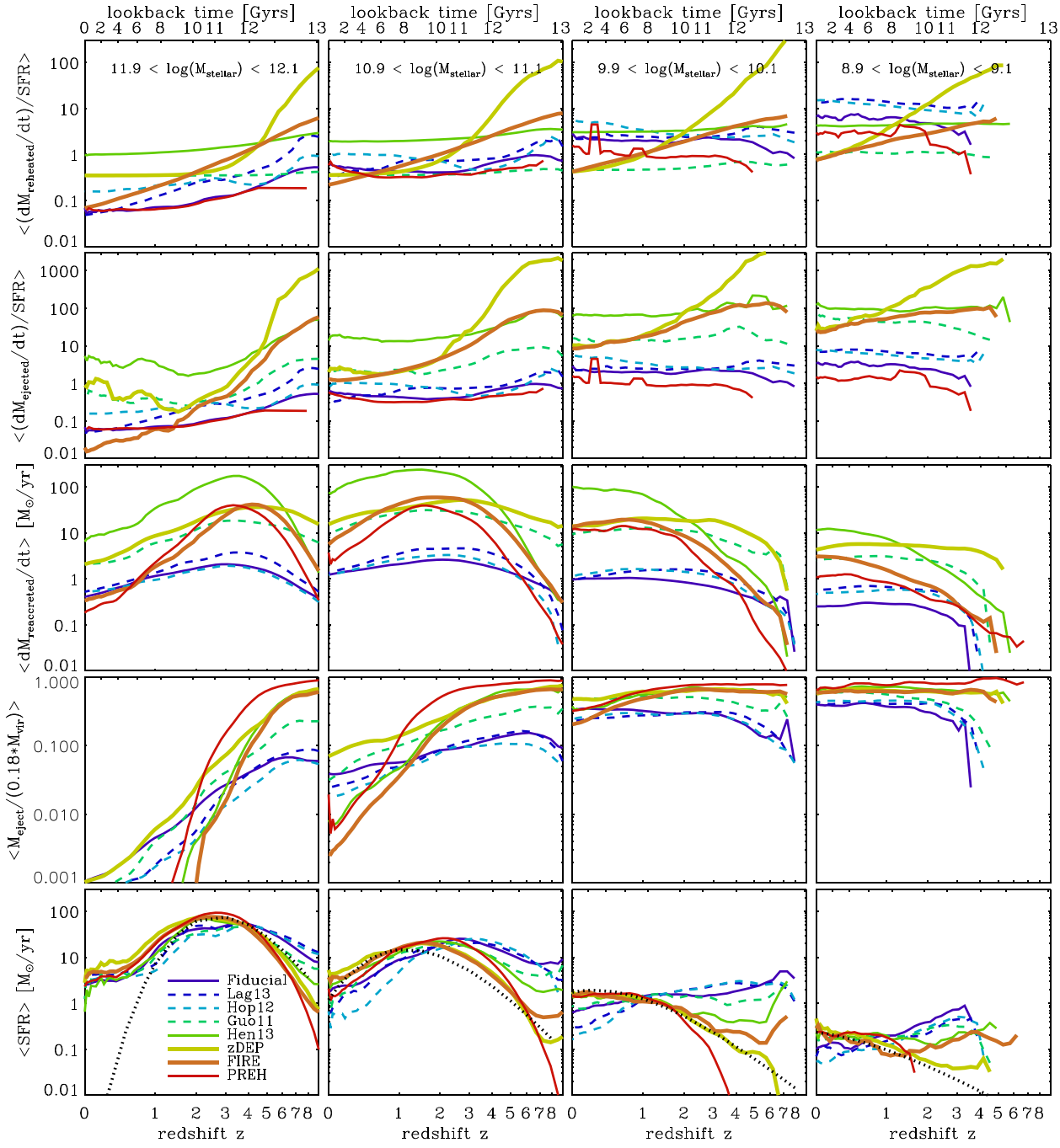

Figure 4. Redshift evolution of the mean mass-loading for the reheated (first row), ejected (second row), and re-accreted rate (third row). The fourth row shows the evolution of the mean fraction of baryons in the ejected gas phase and the bottom row the evolution of the mean SFRs. Different columns correspond to galaxies of different present-day stellar mass. Lines of different colours (as in Fig. 1) illustrate the different feedback schemes considered in this work. Black dotted lines in the bottom row show estimates based on subhalo abundance matching methods by Moster et al. (2013).

evolution of the mean 'baryon ejected fractions' as a function of redshift. These are defined as the fractions of baryons residing in the ejected component with respect to the expected baryonic mass ($= M_{\rm eject}/(f_{\rm bar} * M_{\rm halo})$). We stress that for the PREH model, the ejected component comprises (by construction) gas prevented from infall, i.e. that was never within a halo before.

Hen13, zDEP, FIRE, and PREH require the vast majority of baryons/gas to reside in the 'ejected' reservoir. For low-mass galaxies (10^9 $M_\odot$), the ejected fractions are always very high (larger than 60 per cent). For more massive galaxies, the fraction of baryons in the ejected component is very large at high redshift and decreases with decreasing redshift, more rapidly so for more massive galaxies. For the most massive galaxies considered, the ejected fractions are lower than those predicted by the fiducial model for $z < 2$–3 in the Hen13, FIRE and PREH models.

Therefore, to falsify predictions from both successful ejective (zDEP, FIRE) and preventive (PREH) feedback models, observations of the (diffuse, weakly ionized or neutral) gas in the circumgalactic (CGM) and/or intergalactic medium (IGM) are essential. Diffuse warm gas (10^4–10^5 K) is very difficult to

detect, but high-resolution spectroscopy in the rest-frame UV has started to probe the diffuse gas and metals in the CGM using absorption line measurements along the line of sight of quasars (e.g. Rudie et al. 2012; Prochaska, Hennawi & Simcoe 2013; Tumlinson et al. 2013; Peeples et al. 2014). This provides more stringent constraints on the gas and metals that have been ejected by the winds or have been prevented from infall due to pre-heating invoked by our models. In case of preventive feedback, for example, the CGM/IGM should mainly consist of pristine gas, while for ejective feedback schemes, we would expect much higher levels of metal enrichment.

4.4 Evolution of SFRs

As a direct consequence of the large ejected gas fractions in the Hen13, zDEP, FIRE, and PREH models, star formation is suppressed at early times for galaxies of all masses. The bottom row of Fig. 4 shows the mean star formation histories of galaxies in each of the stellar mass bins considered, and compare predictions from the feedback schemes analysed in this study with estimates based on the subhalo abundance matching method by Moster et al. (2013). For galaxies less massive than 10^{11} $M_{\odot}$, the SFRs in the models are *increasing* with decreasing redshift, while the opposite behaviour is predicted by the fiducial, Lag13, Hop12, and Guo11 models. For the most massive galaxies, the SFRs peak at higher redshift than for their lower mass counterparts in the Hen13, zDEP, FIRE and PREH models, while such a trend is weaker or insignificant for the other models.

The behaviour described above is in fairly good agreement with predictions from the subhalo abundance matching approach described by Moster et al. (2013). For low-mass[3] galaxies (10^{9}–10^{10} $M_{\odot}$), the zDEP model provides the best match to the estimates by Moster et al. (2013), while the FIRE and Hen13 models predict larger SFRs at high redshifts. In the PREH model, the formation of these low-mass galaxies appears to be simply shifted towards later times.

For the most massive galaxies ($M_{\rm stellar} \sim 10^{12}$ $M_{\odot}$), the SFRs tend to be overestimated below $z \sim 1$ with respect to estimates by Moster et al. (2013), for all feedback schemes considered in this study. This is due to the fact that, within the new schemes proposed, the AGN feedback model adopted becomes inefficient. We will come back to that issue in the following.

5 COLD GAS IN THE INTERSTELLAR MEDIUM

Due to the large variations of gas reheating, ejection and reincorporation rates discussed in the previous section, the stellar feedback schemes considered in this study provide significantly different predictions for the amount and properties of the interstellar medium in model galaxies. Our current working version of the GAEA model follows the *total* cold gas content. An updated version of our model, that models explicitly the transition from atomic to molecular hydrogen, is in preparation (Xie et al., in preparation).

5.1 Cold gas fractions

Fig. 5 illustrates how the cold gas fraction varies as a function of the galaxy stellar mass at $z = 0$, 1, 2 for the different feedback models used in this work (lines of different colours), and compares model predictions to observational data (black and grey symbols). The orange shaded region illustrates the 1σ scatter for the FIRE model. Cold gas fractions are defined as:

$$F_{\rm cold} = \frac{M_{\rm cold}}{M_{\rm cold} + M_{\rm stellar}}, \quad (18)$$

where $M_{\rm cold}$ is the cold gas mass (in our model this is associated only with the galaxy disc), and $M_{\rm stellar}$ is the galaxy stellar mass. The model predictions plotted here refer only to 'star-forming' galaxies, selected according to the criterion suggested by Franx et al. (2008): sSFR $> 0.3/t_{\rm Hubble}$. In fact, it is worth noting that observational measurements are available for relatively small (and likely biased) samples of galaxies, particularly at high redshift. In addition, the scatter is rather large (±0.3 dex), in particular for massive galaxies.

At fixed stellar mass, all models predict decreasing cold gas fractions with decreasing redshift – at least for galaxies more massive than 10^{10} $M_{\odot}$. However, only for the Hen13, zDEP, FIRE, PREH, Guo11 models, gas depletion time-scales are long enough that the

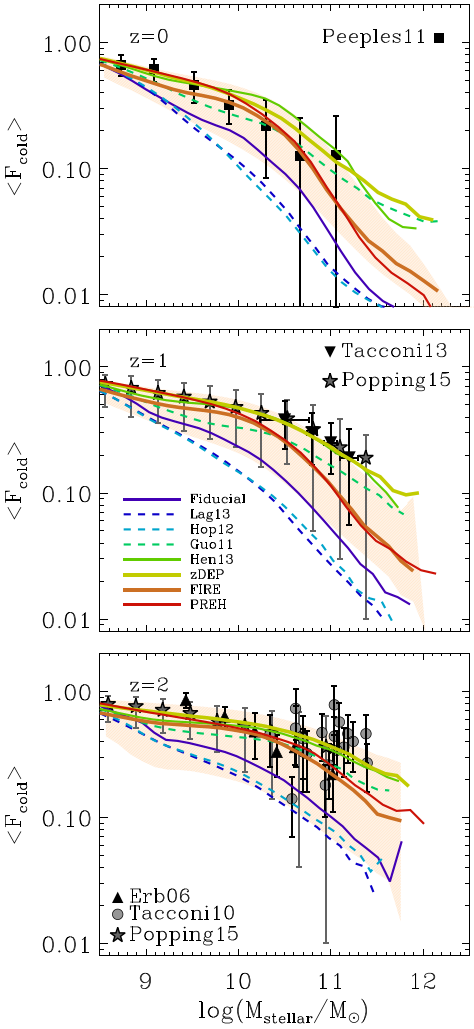

Figure 5. Mean cold gas fractions ($\langle M_{\rm cold}/(M_{\rm cold} + M_{\rm stellar})\rangle$) of star-forming galaxies as a function of galaxy stellar mass at $z = 0$, 1, 2 (from top to bottom). Model predictions (coloured lines and shaded areas as in Fig. 3) are compared to observational data (shown as black symbols) by Peeples & Shankar (2011); Tacconi & et al. (2013); Tacconi et al. (2010); Popping et al. (2015) and Erb et al. (2006).

[3] Note that the 10^{9} $M_{\odot}$ stellar mass bin is at the limit of the Millennium resolution. Using merger trees from the better resolved Millennium-II simulation, we have verified that the results are unchanged.

cold gas fractions are consistent with observational estimates at $z = 0$, 1, 2. In contrast, the fiducial, Lag13 and Hop12 models predict gas fractions that are significantly below the observational estimates, at all galaxy masses, although the scatter (comparable to that of the FIRE model) is fairly large (± 0.2 dex). Interestingly, stellar feedback affects strongly the cold gas content of even very massive galaxies, that are usually believed to be mainly affected by AGN feedback only.

The cold gas content of a galaxy is determined by the combined effect of star formation, reheating, and cooling from the hot gas reservoir associated with the parent halo. The latter quantity depends crucially on the balance between ejection and re-incorporation. We find that these two processes are the main responsible for the differences found among the schemes considered: in the fiducial, Lag13, and Hop12 schemes, the cold gas reservoir of galaxies is quickly consumed via star formation at early times ($z > 3$), and only little gas is ejected and re-incorporated at later times. The other models are characterized by larger ejection rates at high redshift. This prevents early star formation (that would lock large amount of gas). Later large re-incorporation rates lead to the large predicted cold gas fractions that are shown in Fig. 5. Again, a very good agreement with observed data, down to rather small galaxy stellar masses, can be achieved by either ejecting larger fractions than reheated gas (i.e. ejecting also a fraction of the hot gas in the halo), or preventing gas from infall.

The slightly lower cold gas fractions for galaxies more massive than $10^{11}\,M_{\odot}$ in the FIRE and PREH models with respect to the Hen13, zDEP and Guo11 models, are due to the lower gas re-accretion rates, caused by the halo-mass-dependent scaling for gas recycling. The large gas re-accretion rates (and thus, cold gas fractions) in the Hen13 model is a consequence of the large and almost constant ejection rates.

5.2 Cold gas metallicity

Fig. 6 shows how the cold gas metallicity of model star-forming galaxies (selected as mentioned in the previous section) varies as a function of galaxy stellar mass at $z = 0$, 1, 2 for the different feedback models considered (lines of different colours), and compares model predictions with observational data. The 1σ scatter is shown only for the FIRE model (shaded area), and is representative for all models.

The absolute value and evolution of the gaseous metallicity provide strong constraints on the stellar feedback scheme, due to the self-consistent modelling of chemical enrichment in our GAEA model. It is worth reminding that the absolute normalization (and even the shape) of the observed mass–metallicity relation is strongly dependent on the choice of metallicity calibration (Kewley & Ellison 2008). Therefore, it is important that different line diagnostics adopted by different surveys and/or at different redshifts are cross-calibrated consistently (e.g. Maiolino et al. 2008). In this way, the *relative* evolution of the metal content at fixed stellar mass is expected to provide a stronger constrain than the absolute value at a given cosmic epoch.

The PREH, fiducial, Lag13, Hop12, and Guo11 models predict very little evolution in the relation between galaxy stellar mass and the metallicity of the cold gas component, over the redshift range considered. This is due to the rather low reheating and ejection rates at early times in these models, that determine high rates of star formation at high redshift and therefore a rapid enrichment of the cold gas component of galaxies.

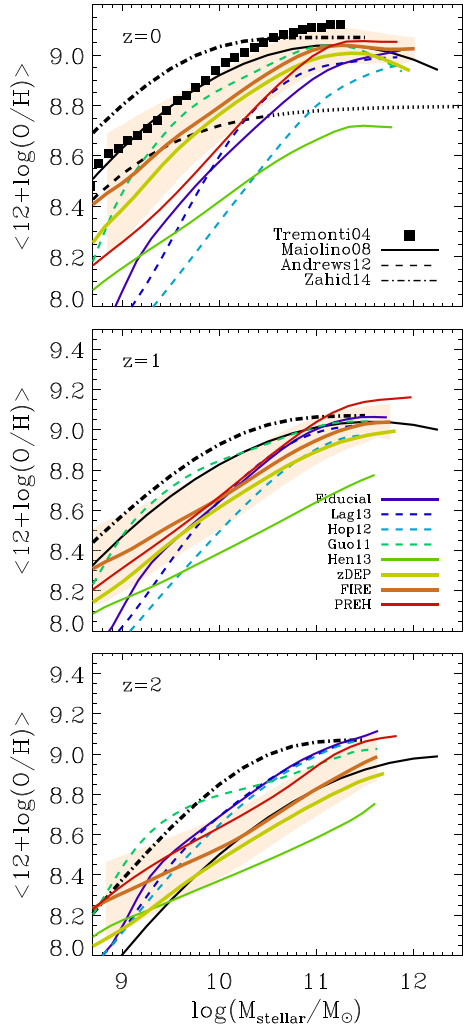

Figure 6. Mean cold gas metallicity of star-forming galaxies as a function of galaxy stellar mass at $z = 0$, 1, 2 (from top to bottom). Model predictions (coloured lines and shaded areas as in Fig. 3) are compared to observational data (shown as black lines and symbols) by Tremonti et al. (2004); Maiolino et al. (2008); Andrews & Martini (2013), and Zahid et al. (2014).

In the Hen13 model, due to the significantly higher gas re-heating rates, more enriched material is blown out of the galaxies (and maybe even out of the haloes) so that the metal enrichment of the cold gas is delayed towards lower redshift. However, since the high reheating gas rates are almost constant with time, the metal enrichment remains strongly suppressed at lower redshifts, leading to an underestimation of the cold gas metallicity with respect to observational measurements.

In contrast with the predictions obtained by our Hen13 implementation, Henriques et al. (2015) argue that the proposed feedback and re-incorporation scheme predicts a realistic metallicity content for present-day galaxies. This apparent contradiction is due to the fact that the model discussed in Henriques et al. (2015) assumes an unrealistically high metal yield (0.047) within their instantaneous recycling approximation. We will come back to this issue Section 8.1.

In the zDEP and FIRE based ejective feedback models, the enrichment of the cold gas is delayed at high redshifts as in the Hen13 model, but the cold gas metallicity is strongly increasing towards $z = 0$, in fairly good agreement with observational constraints from

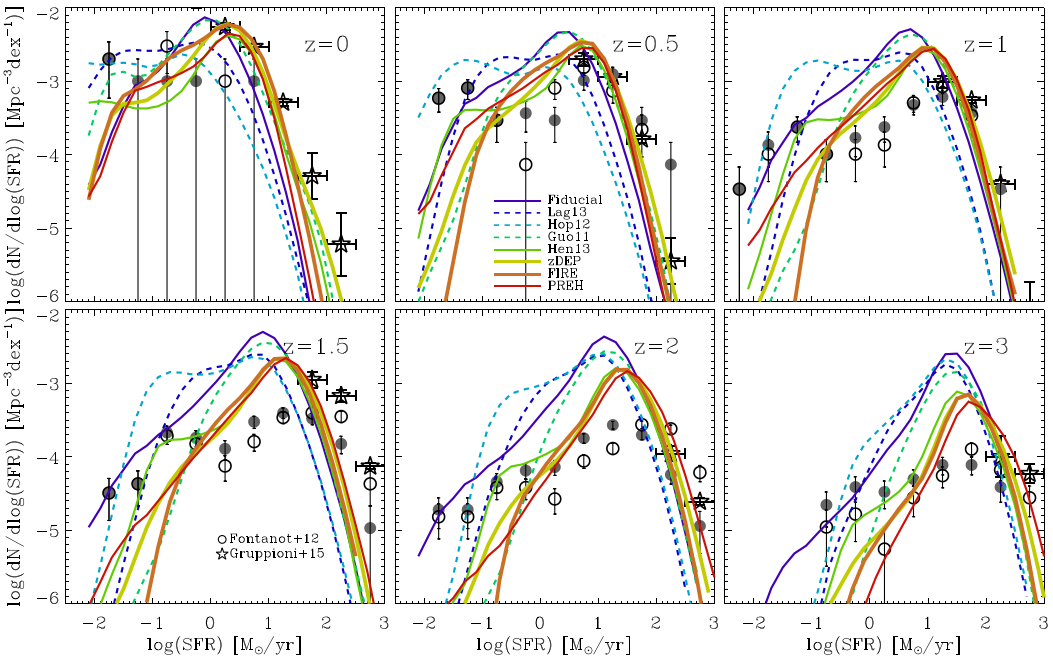

Figure 7. Evolution of the SFR function in the different feedback models (coloured lines as in Fig. 1), compared to observational measurements (black circles and stars taken from Fontanot et al. 2012; Gruppioni et al. 2015; grey circles indicate SFR measurements derived from SED fitting only, and are taken from Fontanot et al. 2012).

e.g. Maiolino et al. (2008). This is due to the rather strongly declining gas reheating rate with decreasing redshift (see top row in Fig. 4) such that much smaller amounts of metal enriched gas are driven out of the galaxies at low redshifts with respect to the Hen13 model.

An interesting behaviour is found for galaxies more massive than $10^{11}\,M_\odot$: for the zDEP, Guo11 and Hen13 models, the cold gas metallicity is slightly decreasing with increasing stellar mass – in tension with the observed trend. This is most likely a consequence of the higher re-accretion rates (and thus cooling rates) of relatively metal-poor gas (that tends to dilute the gas metallicity) with respect to the other models (see third row in Fig. 4).

In all models, we find that the slope of the relation hardly changes with decreasing redshift, while the relation tends to flatten for galaxies more massive than $M_{\rm stellar} < 10^{10.5}\,M_\odot$. This trend is consistent with observational measurements by Maiolino et al. (2008). Only for the zDEP and FIRE models we find also an increase in the normalization of the mass–metallicity relation. A more detailed investigation of the evolution of the mass–metallicity relation will be the subject of a forthcoming paper.

6 STAR FORMATION

The strong impact of stellar feedback on the amount and composition of the cold gas content implies rather different star formation histories in our model galaxies, as discussed in Section 4. In an attempt to discriminate among the different schemes analysed in this paper, we now present a more quantitative comparison between model predictions and the observed amount of star formation as a function of stellar mass.

6.1 Evolution of the SFR function

Fig. 7 shows the evolution of the SFR function for the different feedback schemes analysed (lines with different colours), compared to observational measurements by Fontanot et al. (2012), and Gruppioni et al. (2015) (black circles and stars). Gruppioni et al. (2015) computed the SFRs based on a combination of SED fitting and infrared *Herschel* data coming from the PEP and HerMEs projects (covering a range from 70 to 500 μm). The SFR functions discussed in Fontanot et al. (2012) are computed from the GOODS-MUSIC catalogues (Santini et al. 2009), and they are based either on a combination of UV and 24 μm(for the high-SFR end of the function) or on SED fitting (in the low-SFR regime)[4]. The agreement between these different observational data sets is relatively good, particularly at $z < 1.5$.

We adopted a stellar mass cut in the models of $10^{10}\,M_\odot$ in order to be consistent with the mass-limited samples of Fontanot et al. (2012). This mass cut only affects (decreases) the number densities of galaxies with SFRs below the peak, e.g. SFR $<1-10\,M_\odot$/yr depending on redshift. The comparison with the data by Gruppioni et al. (2015) (IR-flux limited, not mass limited) is, however, still fair, since the sample includes only galaxies with relatively high SFRs and the number densities of these is not affected by the adopted mass cut.

The Hen13, zDEP, FIRE and PREH models are in relatively good agreement with observational measurements for SFRs larger than $\sim 1 M_\odot\,{\rm yr}^{-1}$, up to $z = 1$. At higher redshifts, the number densities of galaxies with SFRs in the range $1–100\,M_\odot{\rm yr}^{-1}$ is overestimated by up to 0.5 dex. The overestimation is even larger for the other models. At all redshifts, the Hen13, zDEP, FIRE models, and in particular the PREH model predict larger number densities than the other models for galaxies with SFRs above $>100\,M_\odot{\rm yr}^{-1}$, in better agreement with observations. Only at $z = 1.5$, the number

[4] These estimates correspond to the open circles in Fig. 7, while solid symbols refer to the alternative choice of using SED fitting for all GOODS-MUSIC sources

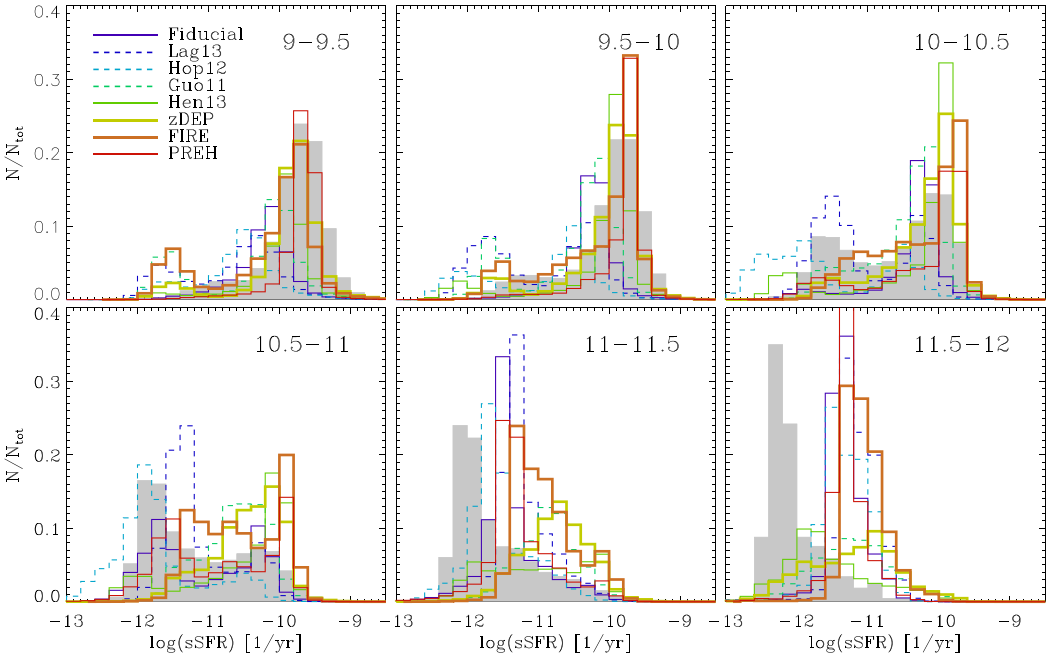

Figure 8. Distributions of specific SFRs for different stellar mass bins (different panels) as predicted by the different stellar feedback models considered in this study (coloured lines as in Fig. 1). Model predictions are compared with observational measurements from SDSS (grey shaded histograms, Yang et al. 2007).

density of these highly star-forming galaxies appears to be strongly underestimated with respect to data.

For galaxies with SFRs lower than $\sim 30\,M_\odot yr^{-1}$, the Hen13, zDEP, FIRE and PREH models predict by up to one order of magnitude smaller number densities than the other models, particularly at high redshift. This clearly originates from the stronger suppression of star formation at high redshift in low-mass galaxies (see bottom row in Fig. 4). For galaxies with SFRs lower than ($1\,M_\odot yr^{-1}$), the Hen13, zDEP, FIRE, and PREH models underestimate the measured number densities at all redshifts. We note, however, that these measurements are based only on SED fitting, and therefore carry large uncertainties.

6.2 Specific SFRs

Fig. 8 shows the present-day distributions of the specific SFR (SFR/$M_{\rm stellar}$), as predicted by the different feedback schemes considered in this study (lines of different colours). Each panel corresponds to a different bin of galaxy stellar mass, as indicated by the legend, and model predictions are compared to observational measurements by the SDSS (grey shaded histograms).

For low-mass galaxies ($M_{\rm stellar} < 10^{10.5}\,M_\odot$, top row), the Hen13, zDEP, FIRE and PREH models predict more galaxies with higher specific SFRs than the other models (whose specific SFRs of star-forming galaxies are generally too low), in better agreement with observations. This is clearly a consequence of the delayed star formation and increasing SFR histories for low-mass galaxies in our strong ejective (Hen13, zDEP and FIRE) and preventive feedback models (PREH, see bottom row in Fig. 4).

For intermediate-mass galaxies $10^{10.5} < M_{\rm stellar} < 10^{11}\,M_\odot$ (bottom left panel), the Hen13, zDEP, FIRE and PREH models tend to predict too large sSFRs, while the others (in particular the Lag13 and the Hop12 models) provide a better match to the observational data, with a relatively large fraction of passive galaxies.

For more massive galaxies ($M_{\rm stellar} > 10^{11}\,M_\odot$), the specific SFRs are overestimated in *all* models, although the disagreement is worse for the Hen13, zDEP, FIRE and PREH models. We ascribe this failure to the adopted scheme for radio-mode AGN feedback that, in the framework of our new feedback schemes, appears to be inefficient in suppressing the star formation in massive galaxies. A simple parameter change (an increase of the AGN feedback efficiency) cannot solve this disagreement. We will come back to that issue in future work.

6.3 The galaxy main sequence

Observations reveal a rather tight correlation between (specific) SFRs of star-forming galaxies and their stellar mass (Daddi et al. 2007; Elbaz et al. 2007; Karim et al. 2011), at both low and high redshifts. Some studies (Daddi et al. 2007; Weinmann et al. 2012; Granato et al. 2015) highlighted that the state-of-the-art galaxy formation models fail to reproduce this observational constraint, by underpredicting the normalization of this correlation and by predicting too shallow slopes for galaxies with stellar mass lower than $10^{10}\,M_\odot$ (Weinmann et al. 2012).

Fig. 9 shows model predictions for the mean specific SFRs of star-forming galaxies (selected as discussed in Section 5) as a function of stellar mass, as predicted by the different feedback schemes used in this study (lines of different colours; the orange shaded area indicates the 1σ-scatter of the FIRE model) at $z = 0$, 1, 2 (from top to bottom) compared to observational data (black symbols, lines and grey shaded area; Daddi et al. 2007; Elbaz et al. 2007; Salimet al. 2007; Santini et al. 2009).

In all models, in agreement with the observed trend, the specific SFRs at a given stellar mass are decreasing with decreasing redshift. The fiducial, Lag13, Hop12, Guo11 models exhibit the known problem of a too low normalization, and nearly flat or even negative slope of the relation for low-mass galaxies.

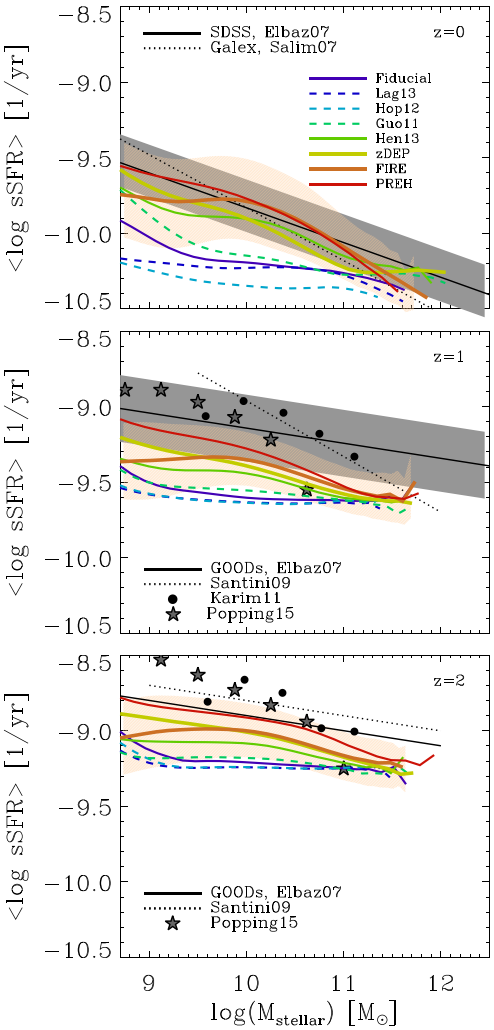

Figure 9. Main sequence of star-forming galaxies, i.e. mean specific SFRs are plotted versus the galaxy stellar mass at $z = 0$, 1, 2 (from top to bottom). Model predictions (coloured lines and shaded areas as in Fig. 3) are compared to observational data (black lines with shaded areas and black symbols – Daddi et al. 2007; Elbaz et al. 2007; Salimet al. 2007; Santini et al. 2009).

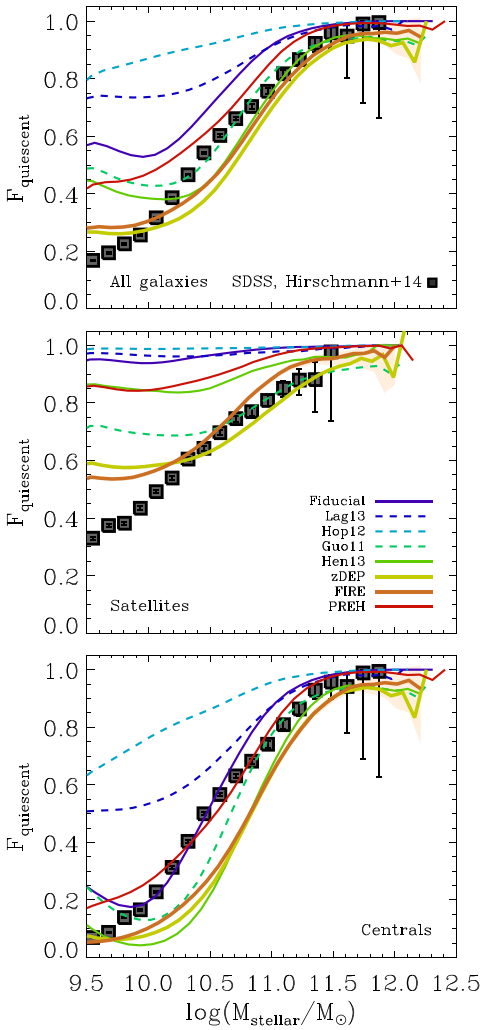

Figure 10. Quiescent galaxy fractions versus galaxy stellar mass for all (top panel), satellite (middle panel) and central galaxies (bottom panel). Model predictions (coloured lines and shaded areas as in Fig. 3) are compared to observational measurements by Hirschmann et al. (2014b), based on SDSS (grey symbols). Galaxies are selected in both models and simulations assuming the same luminosity cut as in Hirschmann et al. (2014b)

In contrast, the Hen13, zDEP, FIRE and PREH models predict larger specific SFRs at any given mass, and steeper relations. The change in normalization with respect to the other models is due to the fact that there is overall more gas available for star formation. For low-mass galaxies, star formation histories are significantly delayed (see Section 4, bottom row of Fig. 4), leading to steeper slopes in the these models at redshift zero. The slope of the relation flattens with increasing redshift, in qualitative agreement with observational measurements.

Although model predictions from the Hen13, zDEP, FIRE and PREH models are definitely in better agreement with data than for the other feedback schemes tested in this work, they still tend to underestimate the specific SFs at high redshifts. The largest specific SFRs and the steepest correlations are obtained with the PREH model.

6.4 Quiescent galaxy fractions

One of the long-standing problems of currently used galaxy formation models is given by their tendency to largely overestimate the fraction of quiescent galaxies, in particular for low-mass satellite galaxies (see e.g. Kimm et al. 2009; Weinmann et al. 2009, 2010; De Lucia et al. 2012; Hirschmann et al. 2014b). When first noticed, this model failure was ascribed to the simplified models for environmental processes: most models (including the GAEA model) assume that the hot gas reservoir associated with each model galaxy is instantaneously stripped as it is accreted on to a larger system (i.e. when the galaxy becomes a satellite). This prevents further cooling and, combined with the typically efficient stellar feedback, drives a fast suppression of the star formation in the model galaxies that transit from star forming to passive on very short time-scales. While more sophisticated treatments of this process have led to some improvements, model predictions still appear unable to entirely reproduce the observed trends in the local Universe (see e.g. Hirschmann et al. 2014b; Wang et al. 2014).

Fig. 10 shows the present-day quiescent fractions of all galaxies (top panel), of satellite (middle panel) and of central galaxies (bottom panel) compared to observational measurements by Hirschmann et al. (2014b). For both data and models, we have

applied the same luminosity cut of Hirschmann et al. (2014b). To distinguish between star-forming and quiescent galaxies, we again used the criterion suggested by Franx et al. (2008). We note that the predicted distributions of sSFR do not agree very well with observational data (in particular for massive galaxies, which have too high sSFRs). The adoption of a different separation criterion for quiescent galaxies would affect the comparison shown in Fig. 10. In particular, assuming a lower sSFR limit for selecting quiescent galaxies, would result in a slight underestimation of the quiescent fractions for massive galaxies.

The top panel of Fig. 10 illustrates that the Hen13, zDEP, FIRE, PREH and Guo11 models predict quiescent fractions consistent with observations for galaxies more massive than $10^{10}\,M_{\odot}$. In contrast, the fiducial, Lag13 and Hop12 models overestimate the fraction of quiescent galaxies, particularly at low masses. For galaxies less massive than $10^{10}\,M_{\odot}$, the quiescent galaxy fractions are overestimated in all models to a very different degree.

In the middle and bottom panels, we split the galaxy populations into centrals and satellites. Regarding centrals, the fiducial, PREH and Guo11 models predict realistic quiescent fractions, while the strong ejective models (Hen13, zDEP and FIRE) predict slightly too few quiescent galaxies – an effect already discussed in Fig. 8 and likely pointing towards more fundamental modifications of the modelling adopted for radio-mode AGN feedback.

For satellites, the quiescent fractions are found to be very strongly dependent on the stellar feedback scheme. We find reduced quiescent fractions in the strong ejective schemes (Hen13, zDEP, FIRE) and in the Guo11 models with respect to the other models. We stress that, in all adopted schemes, we assume an instantaneous stripping of the hot gas reservoir associated with infalling galaxies. Therefore, the model quiescent satellite fractions are mainly regulated by: (i) the amount of cold gas at the time of infall, and (ii) the rate at which the cold gas gets re-heated and is, thus, blown out of the satellites (later on, it can be re-incorporated only on to the central galaxy of the halo).

At the time of infall, the (central) galaxies in the strong ejective feedback models (Hen13, zDEP, FIRE) are more star-forming, i.e. they have a larger cold gas content, which allows them to sustain star formation for longer time-scales. The quiescent fractions predicted by the strong ejective feedback schemes vary significantly: the more gas gets re-heated (as e.g. in the Hen13 model, see top row in Fig. 4), the less gas is available for further star formation, the faster is the remaining cold gas consumed, and the higher are the quiescent fractions. This explains the larger quiescent fractions in the Hen13 model (characterized by almost constant re-heating rates) with respect to the FIRE and zDEP models (characterized by decreasing re-heating rates towards $z = 0$). The rather low quiescent fractions in the Guo11 model can be explained with similar arguments: the re-heating rates in this model are almost constant and rather low with respect to the Hen13 model or even the fiducial model.

We stress that *all* our models are based on the simplified assumption of an instantaneous hot gas stripping of satellites. Therefore, in the framework of our models, the fractions of passive galaxies (both centrals and satellites) are primarily determined by internal physical processes, with environmental processes playing only a secondary role. The fact that the most successful models are still overestimating the lowest mass quiescent satellites suggests that environmental effects become important at these mass scales. We postpone a more detailed analysis to a future work (Hirschmann et al., in preparation).

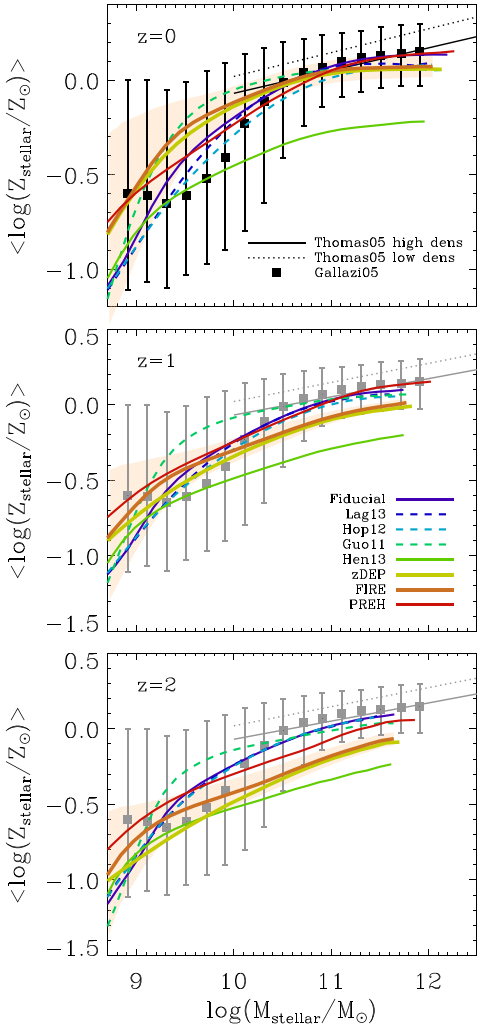

Figure 11. Mean stellar metallicity versus galaxy stellar mass at $z = 0$, 1, 2 (from top to bottom). Model predictions (coloured lines and shaded areas as in Fig. 3) are compared to observational data of the present-day Universe (black/grey lines with shaded areas and black/grey symbols, Gallazzi et al. 2005; Thomas et al. 2005).

7 STELLAR POPULATIONS

In this section, we discuss the impact of the different stellar feedback schemes adopted in this study on the stellar populations of our model galaxies, particularly with respect to their stellar metallicities, ages and colours.

7.1 Stellar metallicity

Fig. 11 shows the stellar mass-stellar metallicity relation at $z = 0$, 1, 2 for our different feedback schemes, compared to the present-day observed relation (black symbols and lines, corresponding to measurements by Gallazzi et al. 2005 and Thomas et al. 2005, respectively). For reference, we show the $z = 0$ observational measurements also in the panels corresponding to higher redshifts. At $z = 2$, the fiducial, Lag12, Hop11, Guo11 models predict that galaxies more massive than $10^{10}\,M_{\odot}$ have already reached supersolar metallicities, and their metal content is hardly evolving down to $z = 0$. The PREH model shows a similar behaviour: since it includes the same re-heating and ejection scheme as in the fiducial model, we expect a similar metal enrichment history.

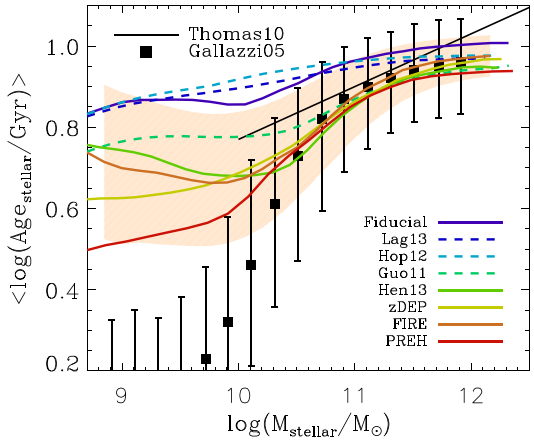

Figure 12. Present-day mean stellar ages as a function of the galaxy stellar mass in the different feedback models considered in this study (coloured lines and shaded areas as in Fig. 3), compared to observations by Thomas et al. (2010) and Gallazzi et al. (2005) (black solid line and symbols).

Strong ejective feedback (Hen13, zDEP, and FIRE illustrated by green, yellow and orange solid lines) can have a significant impact on the evolution of the stellar metallicity: early star formation is suppressed and, in addition, metal-enriched gas is ejected out of the galaxy. As a consequence, metal enrichment of the stellar component is delayed, so that all galaxies in the strong ejective feedback models have sub-solar metallicity at $z = 2$. In the FIRE and zDEP models, the stellar metallicity is strongly increasing with decreasing redshift, approaching supersolar metallicity for the most massive galaxies, in good agreement with the present-day observed relation. In contrast, the Hen13 model predicts a too strongly delayed metal enrichment so that even the most massive galaxies are always below solar levels.

This different behaviour of the strong ejective feedback schemes (Hen13, zDEP, and FIRE) can be explained considering the larger re-heating and ejection rates in the Hen13 model with respect to the FIRE and zDEP models below $z \sim 2$ (see two first rows in Fig. 4). Since the metal flows are assumed to follow the gas flows, more metals are ejected from the cold, star-forming gas phase and transferred to the hot gas phase at late times in the Hen13 model than in the FIRE and zDEP models. Due to mixing with less metal enriched gas in the hot gas phase (or in the ejected phase), the metallicity of the cooled gas is typically lower than that of the cold gas. This tends to dilute the metallicity of the cold gas component, and therefore keeps the metallicity of the stars formed later low.

7.2 Galaxy stellar ages

Fig. 12 shows the present-day stellar mass-age relation for the different feedback models (lines of different colours; the orange shaded area indicates the 1σ-scatter for the FIRE model), compared to observational measurements by Thomas et al. (2010) and Gallazzi et al. (2005) (black solid line and symbols). Both model stellar ages and observational measurements are luminosity weighted (r-band, SDSS).

Galaxies in the fiducial, Lag13 and Hop12 models are generally too old compared to observational measurements. In the Guo11 feedback scheme, galaxies are on average younger, and in good agreement with observations for galaxies more massive than $10^{10.5}\,M_\odot$. Less massive galaxies are still too old compared to data. The younger stellar ages in the Guo11 model are a natural consequence of the strongly suppressed star formation at early cosmic times, compared to the other models.

The trend is even more extreme for the Hen13, zDEP, FIRE and PREH models, resulting in fairly realistic stellar ages for galaxies with masses above $10^{10}\,M_\odot$. For less mass galaxies ($<10^{10}\,M_\odot$), stellar populations tend to be still too old compared to measurements by Gallazzi et al. (2005), although the scatter in both observations and models at the low mass is rather large (± 0.2 dex).

The PREH and zDEP models predict the youngest low-mass galaxies with respect to the other feedback models. This is due to the stronger early star formation suppression (see bottom panel in Fig. 4) with respect to the other successful FIRE or Hen13 models. Interestingly, for the FIRE and Hen13 low-mass galaxies, stellar ages are increasing with decreasing stellar mass, a trend that is neither visible in the zDEP and PREH models and, nor in the observations.

7.3 Galaxy colours

We conclude our sections on model results by showing the predicted colour distributions from the different feedback schemes used in our study. While these are direct observables, assumptions need to be made in the models to predict them from physical properties (e.g. on stellar population models, dust attenuation, etc.). Model predictions for the u-i colour distributions are shown in Fig. 13, and compared with measurements from the SDSS (grey shaded histograms).

All models predict a rather strong colour bi-modality for galaxies less massive than $10^{11}\,M_\odot$, while in the data this is evident only for galaxies in the mass range $10^{9.5}$–$10^{10.5}\,M_\odot$. The colours of blue model galaxies tend to be generally too blue (see top row). For galaxies with stellar mass $<10^{11}\,M_\odot$, the Hen13, zDEP, FIRE, and PREH models predict a larger fraction of blue galaxies than the other models, consistently with the lower stellar ages discussed above. This results in a better agreement with observational measurements for masses $<10^{10}\,M_\odot$, in particular when considering the FIRE and zDEP models. For galaxies with masses between 10^{10} and $10^{11}\,M_\odot$, the fraction of blue galaxies is, however, clearly overestimated with respect to data.

For galaxies with mass larger than $10^{11}\,M_\odot$, model colours are generally not red enough compared to observational measurements, a trend that appears even worse for the Hen13, zDEP, FIRE and PREH models. This is consistent with the non negligible levels of star formation in massive galaxies discussed earlier (see e.g. Fig. 4 bottom, left panel). These are low enough to not affect significantly stellar masses, ages and metallicities, but have a larger effect on colours (especially in the blue bands).

8 DISCUSSION

In the last decade, a number of theoretical studies have pointed out the existence of a *fundamental problem* with the evolution of low mass ($\leq 10^9\,M_\odot$) galaxies in hierarchical models. This problem has different manifestations: (i) models tend to systematically and significantly overpredict the number density of low-mass galaxies (e.g. Guo et al. 2011; Hirschmann et al. 2012b; Lu et al. 2014); (ii) low-mass galaxies, and in particular satellite galaxies, tend to be too passive with respect to observational data (e.g. Kimm et al. 2009; Weinmann et al. 2010). The work carried out in the last decade has pointed out that these problem cannot be overcome by simple modifications of the satellite treatment like a non instantaneous stripping of the hot gas reservoir associated with infalling galaxies (see e.g.

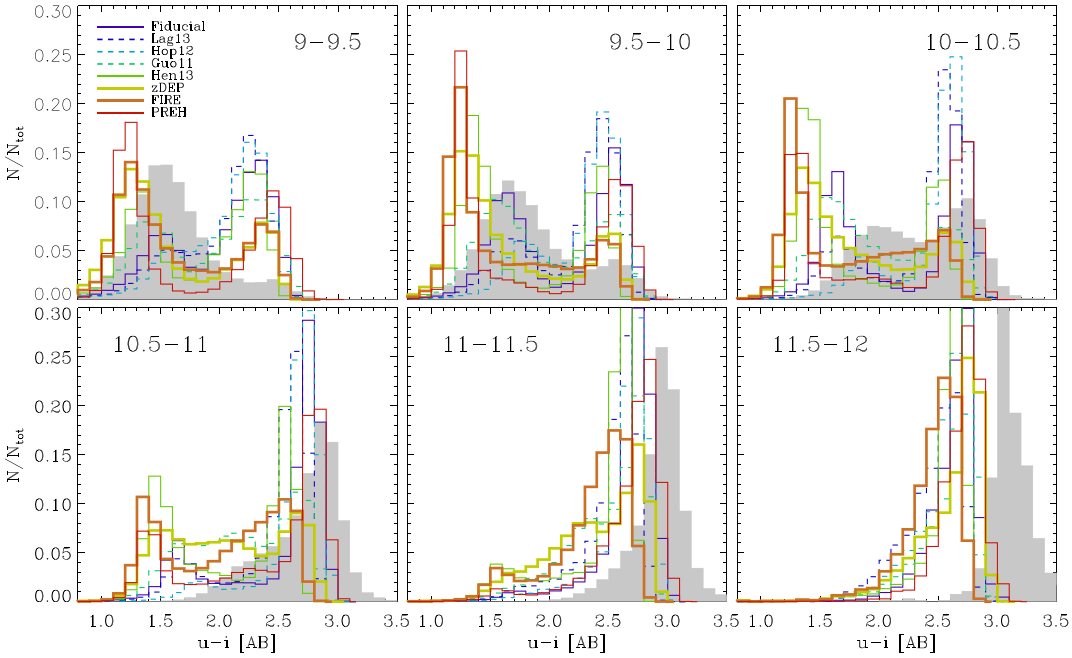

Figure 13. Present-day *u-i* colour distributions for different stellar mass bins (different panels) as predicted by the different stellar feedback models (coloured lines as in Fig. 1), compared to SDSS measurements (grey shaded histograms, Yang et al. 2007).

Hirschmann et al. 2014b), or a more efficient stellar feedback or a stronger scaling with halo circular velocity(e.g. Guo et al. 2011).

8.1 Comparison with previous work

A few solutions to the problems mentioned above have been proposed in the framework of different theoretical models of galaxy formation and evolution. We discuss them below, comparing results from these studies to ours.

8.1.1 L-Galaxies models

Although different in a number of details, our GAEA model originates from the same model as the most recent versions of that built by the 'Munich' group: L-Galaxies.

The overestimation of the low-mass galaxies at $z > 0$ in the Guo11 feedback scheme adopted in GAEA is entirely consistent with results presented in Guo et al. (2011), despite significant differences in modelling the metal enrichment and the evolution of satellite galaxies. This scheme also leads to low-mass galaxies that are too red, and massive galaxies that are too blue compared with observational data.

Although both our Guo11 feedback scheme and the original Guo et al. (2011) model predict reduced, and thus more realistic, present-day quiescent satellite fractions at fixed stellar mass, the physical reasons for this are different: the lower quiescent fractions in the Guo11 model in GAEA with respect to our fiducial model are mainly due to the lower re-heating efficiency, which we found to be necessary for predicting a realistic metal content in our present-day galaxies. In contrast, we suspect that in Guo et al. (2011), the reduced quiescent satellite fractions are mostly due to the different treatment of environmental effects (a delayed stripping of the hot gas associated with infalling satellites). In fact, when assuming the same re-heating efficiency as in Guo et al. (2011), we obtain much larger quiescent fractions with GAEA.

In the most recent version of the L-Galaxies model, Henriques et al. (2013, 2015), have proposed a modified gas reincorporation scheme where the re-incorporation time-scales vary with the inverse of the halo mass. This new element introduced in the feedback scheme is able to bring model predictions in agreement with the measured evolution of the GSMF. Our implementation of the Hen13 feedback scheme in the GAEA model is fully consistent with results discussed by Henriques et al. (2013). The physical properties of low-mass galaxies are in better agreement with data: they are younger, bluer, more star-forming, more gas-rich and the fraction of quiescent satellites is reduced. Nevertheless, some inconsistencies with observations remain, e.g. with respect to colour and sSFR distributions, in particular for the most massive galaxies that tend to be too blue and star forming.

We find, however, significant differences between our GAEA implementation of the Hen13 feedback scheme and results by Henriques et al. (2015) in the framework of the L-Galaxies code. In the former, the present-day stellar (and gaseous) metallicity is severely underestimated compared to observational measurements, while Henriques et al. (2015) show a very good agreement with data. Their good match is due to an unrealistically high metal yield (0.047) adopted in the framework of their instantaneous recycling approximation. In the GAEA model, the metal yield is no longer a free parameter anymore, but is determined by the set of metal yields chosen and, although there are some uncertainties in the yields of some elements, it cannot be significantly larger than ∼0.02. In this way, observational measurements of the metal content in the local Universe represent a truly independent constraint with respect to measurements of the GSMF.

We find that, reducing the re-heating and ejection efficiencies in our Hen13 feedback implementation such that massive galaxies reach more realistic supersolar metallicities, the evolution of the low-mass end of the SMF cannot be reproduced anymore, and the number densities of low-mass galaxies are again overestimated with respect to data. Our analysis does not exclude that it is possible

to find a good match with the observed evolution of the GSMF and the measured metal content of galaxies in the local Universe with a lower chemical yield, within the Henriques et al. (2015) scheme. Such an analysis is beyond the aims of this paper, but can be effectively carried out using the Monte Carlo Markov Chain approach used in Henriques et al. (2015). If such a solution is found, it will be interesting to compare predictions from such a model with results shown in our Fig. 4.

8.1.2 Independent semi-analytic models

Lagos et al. (2013) argued that with their feedback model they can reproduce the evolution of the shallow end of the luminosity function (they do not show a comparison between model predictions and the observed GSMF). For our Lag13 implementation in GAEA, the faint end of the luminosity function is also shallower than in our fiducial model. Model results are, however, inconsistent with the observed evolution of the GSMF. However, it should be noticed that the model presented in Lagos et al. (2013) does not follow explicitly the fate of the re-heated gas. One difference between our GAEA implementation and the original Lagos et al. (2013) model is that the latter includes an explicit modelling of the H I and H_2 formation. This, however, is not expected to introduce significant differences in model predictions (Lagos et al. 2011). Therefore, our results are probably due to the overall significant differences between the modelling of physical processes in GALFORM and GAEA, highlighting the strong interplay between stellar feedback and other physical processes driving galaxy evolution.

Our simple implementation of pre-heating agrees with results based on the more sophisticated model discussed in Lu et al. (2015): reducing the amount of newly infalling gas can lead to realistic SFR histories, cold gas fractions and baryon conversion efficiencies in low-mass galaxies. Lu et al. (2015) do not discuss the impact on metallicity, that we find to be a weak point of our pre-heating implementation: this model predicts a too fast enrichment of the cold gas content, in contrast with observational measurements. We might speculate that, in reality, preventive and ejective feedbacks are both at play, likely with different relative importance at different cosmic epochs and at different mass scales.

White et al. (2015) study the impact of preventive and redshift dependent ejective feedback models using the semi-analytic model presented in Somerville et al. (2012). They base their work on 'ad hoc' parametrizations (just like our zDEP model) aimed at matching the number density of low-mass galaxies. Even if their specific feedback implementations differ from ours in detail, they also find that both ejective and preventive feedbacks reduce the number density of low-mass galaxies and predict larger cold gas fractions and specific SFRs, in fairly good agreement with observational constraints. In their preventive model, the gaseous metallicity at a fixed stellar mass is fairly constant over time (like in our implementation), and only the ejective feedback predicts a (slightly) increasing gaseous metal content. Although not in perfect agreement with observational measurements, their results and conclusions are consistent with ours.

Mitchell et al. (2014) show that a redshift dependent reincorporation time-scale provides a viable solution to obtain more realistic stellar mass assembly histories and a negative slope in the sSFR-stellar mass relation (particularly for low-mass galaxies). Their assumption, however, results in a worse match with the GSMF than a simple dependence on halo mass as suggested by Henriques et al. (2013). We can speculate that, within our GAEA model, a redshift dependent reincorporation time-scale (instead of a redshift dependent outflow rate) can, in some form, provide a good match to the observed evolution of the GSMF. In this respect, it would be beneficial to study how the reaccretion time-scale vary ad a function of redshift in the state-of-the-art cosmological hydrodynamical simulations.

8.1.3 Cosmological hydrodynamical simulations

Over the last years, several studies, based on different hydrodynamical simulations codes, have been successful in reproducing the estimated baryon conversion efficiencies and, in a few cases, even in matching the shallow low-mass end of the GSMF (Aumer et al. 2013; Stinson et al. 2013; Hopkins et al. 2014; Furlong et al. 2015). This success was, however, achieved by using fundamentally different sub-resolution models and by making different assumption for stellar feedback. This is just a reflection of the intrinsic difficulty to model stellar feedback, i.e. processes such as stellar winds, radiation pressure, ionizing radiation, expansion of SN bubbles, from 'first principles'.

In agreement with results discussed above, all successful simulations predict that, at high redshift, the vast majority of gas should be prevented from cooling and forming stars, either by ejecting accreted, low angular momentum gas or by preventing gas infall from the IGM. The consequences of having large amounts of non-cooling gas are similar to those discussed earlier: star formation is suppressed at high redshift and delayed towards later times, metal enrichment is delayed, cold gas fractions are increased. In particular, low mass-galaxies are bluer, more star-forming (higher sSFR) and younger (e.g. Aumer et al. 2013; Hirschmann et al. 2013; Hopkins et al. 2014), in overall better agreement with observations.

Some simulations (e.g. Hopkins et al. 2014), however, show that when assuming some form of strong ejective feedback to reproduce the observationally inferred mass assembly, they fail to generate simultaneously realistically looking, thin disc-like galaxies. This tension may point towards other, less 'violent' mechanisms to get rid of large amounts of cold gas in galaxies or to prevent pristine gas from infall in order to suppress early star formation by *simultaneously* allowing for a build-up of a realistic spiral/disc-like structure.

8.2 A successful feedback scheme in GAEA. Limits and ways forward

In our analysis, four of the feedback schemes tested, the Hen13, zDEP, FIRE and PREH models, are reasonably successful in reproducing the observed trends in galaxy mass assembly. This is achieved by 'storing' large amounts of gas in an 'ejected gas reservoir', where it is not available for cooling. These models predict specific SFRs, stellar ages and cold gas fractions that are overall in better agreement with observational measurements than our fiducial feedback scheme used in previous work. When considering other observational constraints, such as the gaseous and stellar metal content and the quiescent satellite fractions, our zDEP and FIRE models appear to perform better. In particular, both the Hen13 and PREH scheme predict too many passive satellite galaxies, and the Hen13 model predicts a too efficient suppression of the metal enrichment so that the metal content of model galaxies is underestimated, particularly in the local Universe. The PREH model can match the observed present-day mass–metallicity relation, but it results in a too fast enrichment of the cold gas content in massive galaxies. We stress that, given the self-consistent treatment of chemical enrichment in the GAEA model, the chemical yield cannot be regarded

as a free parameter and observational measurements of the metal content of different baryonic components represent independent and powerful constraints to distinguish among different feedback schemes.

Both favoured feedback models (FIRE and zDEP) imply an explicit redshift dependence of the mass-loading for the re-heated and ejected gas, at fixed circular velocity. In particular, they predict that a baryon fraction of ∼60–70 per cent is not available to cool on to the central galaxy disc and to form stars at high redshift. Forthcoming measurements of the gas composition 'outside' galaxies, e.g. in the CGM and/or IGM, will help to falsify model predictions. Specifically, the metal content of the gas in the surroundings of a galaxy will provide strong constraints on the feedback schemes: while purely preventive models would predict a rather pristine gas composition, ejective models favour a metal-enriched CGM.

In spite of the success of our FIRE and zDEP models, some inconsistencies with observations remain. In particular, massive galaxies ($>10^{11}\,M_{\odot}$) have still too high SFRs at low redshifts. As mentioned, this is most likely related to the currently adopted model for AGN feedback. We find, however, that simply increasing the corresponding feedback efficiency is not sufficient to bring model predictions in better agreement with data. This points towards a more fundamental revision of the AGN feedback scheme with the inclusion of e.g. the effect of quasar driven winds (e.g. Ostriker et al. 2010), that is currently neglected in the GAEA model.

In addition, in the FIRE model, low-mass galaxies ($<10^{9.5}\,M_{\odot}$) still tend to be too numerous at $z = 0.5$–2, too red and too old. We argue that, at this mass scale, environmental effects become important and a better treatment of satellite galaxies (e.g. a non instantaneous stripping of the hot gas reservoir associated with infalling satellites) might improve the agreement with observational data. We plan to address this in future work. It is interesting to note that these problems are alleviated in the zDEP model and in the PREH model. It remains to be seen if a more sophisticated pre-heating model, coupled with a strong ejective feedback scenario, could provide a better description of the low-mass galaxy populations.

Our analysis does not allow us to further discriminate between the FIRE and zDEP model. As the latter has been constructed 'ad hoc' to reproduce the evolution of the GSMF, we regard the former as our 'reference model'. Although this model is based on results from cosmological hydrodynamical simulations that include an explicit modelling of complex and relevant physical processes, we stress that this is still a 'sub-grid' (semi-analytic) model.

9 SUMMARY

In this study, we carried out a systematic analysis of the influence of different stellar feedback models, with a focus on the mechanisms/schemes required to reproduce the observed evolution of the GSMF. Our work is based on the GAEA model, an evolution of the semi-analytic model presented in De Lucia & Blaizot (2007), and includes the detailed chemical enrichment scheme introduced in De Lucia et al. (2014b). For each feedback model, we have adjusted the corresponding parameters (for stellar feedback and the reincorporation efficiency, see table 1) so as to match the exponential cut-off of the stellar mass function at $z = 0$, and by simultaneously trying to obtain a good match with the observed mass–metallicity relation in the local Universe and the measured evolution of the galaxy stellar mass function (GSMF) at higher redshift.

Our main results can be summarized as follows.

(i) Both a strong ejective (Hen13, zDEP, FIRE) and a preventive form of feedback (PREH) are capable of reproducing the observed evolution of the GSMF. In our successful strong ejective feedback models (zDEP, FIRE), the mass-loading is dependent on redshift, and large fractions of the hot gas associated with the parent dark matter halo can be driven out of the halo (ejected) and made unavailable for cooling for relatively long time-scales. In order to reproduce the observed trends, large amounts of baryons (up to 70 per cent of the baryon budged available) have to be unavailable for cooling, particularly at high redshift. Delayed gas re-incorporation later reduces the large amounts of cold, star-forming gas at high redshifts (as already suggested by Henriques et al. 2013). The same schemes predict bluer colours and larger amounts of cold gas for low-mass galaxies, in overall better agreement with observational measurements than our previous fiducial model.

(ii) Due the full chemical enrichment scheme in our GAEA model, we can use observational measurements of the gaseous and stellar metallicity content of galaxies as independent constraints. A feedback model with strong but constant gas outflows and a delayed gas-recycling (as suggested by Henriques et al. 2013) leads to a too efficient suppression of the metal enrichment, in contrast with observational data. Our preventive feedback scheme (PREH) is able to predict a realistic present-day metal content, but the ISM is enriched too quickly. Only by using a strong ejective feedback with a redshift dependent mass-loading (zDEP, FIRE), can we successfully reproduce both a delayed metal enrichment and a realistic present-day stellar and gaseous metallicity.

(iii) Although significantly improved with respect to previous results, our new models are not without problems. In particular, massive galaxies appear to be too active with respect to observational measurements at low redshift and the number densities of low-mass galaxies are still overestimated over the redshift range $z = 0.5$–2. Dwarf galaxies also appear too old over this redshift range. We argue that these problems require a significant revision of our AGN feedback model, and a more sophisticated treatment of environmental processes. We plan to address these issues in future work.

Despite significant recent progress in reproducing a number of crucial observed trends, current galaxy formation models are still highly degenerate in terms of the adopted stellar and recycling schemes. More and more stringent observational constraints are coming in the next future, for example through precise high redshift measurements of the stellar mass functions, baryon conversion efficiencies and stellar metallicities. Strong constraints on the baryon cycle can come from observations of the diffuse gas and metals in the GCM and IGM.

ACKNOWLEDGEMENTS

Galaxy catalogues for our new GAEA model (with the FIRE feedback scheme) implemented on the Millennium merger trees will be publicly available at http://www.mpa-garching.mpg.de/millennium. We thank Rachel Somerville, Pierluigi Monaco, Stéphane Charlot and Simon White, Bruno Henriques and Rob Yates for fruitful discussions. We thank Volker Springel and Mike Boylan-Kolchin for making available to us the merger trees from the Millennium and Millennium II simulations and Gerard Lemson and Simon White for helping to make our model catalogues publicly available.

MH acknowledges financial support from the European Research Council via an Advanced Grant under grant agreement no. 321323 NEOGAL. GDL and from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 202781. GDL acknowledges support from the MERAC foundation. FF acknowledges financial support from the grants PRIN INAF 2010 'From the dawn of galaxy formation', and PRIN MIUR 2012 'The Intergalactic Medium as a probe of the growth of cosmic structures'.

REFERENCES

Andrews B. H., Martini P., 2013, ApJ, 765, 140
Aumer M., White S. D. M., Naab T., Scannapieco C., 2013, MNRAS, 434, 3142
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Bell E. F., McIntosh D. H., Katz N., Weinberg M. D., 2003, ApJ, 585, L117
Blumenthal G. R., Faber S. M., Primack J. R., Rees M. J., 1984, Nature, 311, 517
Bonaparte I., Matteucci F., Recchi S., Spitoni E., Pipino A., Grieco V., 2013, MNRAS, 435, 2460
Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645
Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, MNRAS, 398, 1150
Brusa M. et al., 2015, MNRAS, 446, 2394
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Bundy K., Ellis R. S., Conselice C. J., 2005, ApJ, 625, 621
Caputi K. I. et al., 2007, ApJ, 660, 97
Cattaneo A. et al., 2009, Nature, 460, 213
Chabrier G., 2003, PASP, 115, 763
Chieffi A., Limongi M., 2002, ApJ, 577, 281
Choi E., Ostriker J. P., Naab T., Oser L., Moster B. P., 2015, MNRAS, 449, 4105
Cicone C. et al., 2014, A&A, 562, A21
Cole S. et al., 2001, MNRAS, 326, 255
Croton D. J., 2006, MNRAS, 369, 1808
Daddi E. et al., 2007, ApJ, 670, 156
Davé R., Oppenheimer B. D., Finlator K., 2011, MNRAS, 415, 11
De Lucia G., Blaizot J., 2007, MNRAS, 375, 2
De Lucia G., Borgani S., 2012, MNRAS, 426, L61
De Lucia G., Helmi A., 2008, MNRAS, 391, 14
De Lucia G., Kauffmann G., White S. D. M., 2004, MNRAS, 349, 1101
De Lucia G., Weinmann S., Poggianti B. M., Aragón-Salamanca A., Zaritsky D., 2012, MNRAS, 423, 1277
De Lucia G., Muzzin A., Weinmann S., 2014a, New Astron. Rev., 62, 1
De Lucia G., Tornatore L., Frenk C. S., Helmi A., Navarro J. F., White S. D. M., 2014b, MNRAS, 445, 970
Drory N., Bender R., Feulner G., Hopp U., Maraston C., Snigula J., Hill G. J., 2004, ApJ, 608, 742
Dunn J. P. et al., 2010, ApJ, 709, 611
Elbaz D. et al., 2007, A&A, 468, 33
Erb D. K., Shapley A. E., Pettini M., Steidel C. C., Reddy N. A., Adelberger K. L., 2006, ApJ, 644, 813
Fabian A. C., 2012, ARA&A, 50, 455
Fontana A. et al., 2004, A&A, 424, 23
Fontana A. et al., 2006, A&A, 459, 745
Fontanot F., De Lucia G., Monaco P., Somerville R. S., Santini P., 2009, MNRAS, 397, 1776
Fontanot F., Pasquali A., De Lucia G., van den Bosch F. C., Somerville R. S., Kang X., 2011, MNRAS, 413, 957
Fontanot F., Cristiani S., Santini P., Fontana A., Grazian A., Somerville R. S., 2012, MNRAS, 421, 241
Franx M., van Dokkum P. G., Schreiber N. M. F., Wuyts S., Labbé I., Toft S., 2008, ApJ, 688, 770
Furlong M. et al., 2015, MNRAS, 450, 4486
Gallazzi A., Charlot S., Brinchmann J., White S. D. M., Tremonti C. A., 2005, MNRAS, 362, 41
Genel S. et al., 2014, MNRAS, 445, 175
Gonzalez-Perez V., Lacey C. G., Baugh C. M., Lagos C. D. P., Helly J., Campbell D. J. R., Mitchell P. D., 2014, MNRAS, 439, 264
Granato G. L., Ragone-Figueroa C., Domínguez-Tenreiro R., Obreja A., Borgani S., De Lucia G., Murante G., 2015, MNRAS, 450, 1320
Gruppioni C. et al., 2015, MNRAS, 451, 3419
Guo Q. et al., 2011, MNRAS, 413, 101
Guo Q., White S., Angulo R. E., Henriques B., Lemson G., Boylan-Kolchin M., Thomas P., Short C., 2013, MNRAS, 428, 1351
Harrison C. M., Alexander D. M., Mullaney J. R., Swinbank A. M., 2014, MNRAS, 441, 3306
Heckman T. M., Armus L., Miley G. K., 1990, ApJS, 74, 833
Heckman T. M., Lehnert M. D., Strickland D. K., Armus L., 2000, ApJS, 129, 493
Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R. E., Guo Q., Lemson G., Springel V., 2013, MNRAS, 431, 3373
Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, MNRAS, 451, 2663
Hirschmann M., Naab T., Somerville R. S., Burkert A., Oser L., 2012a, MNRAS, 419, 3200
Hirschmann M., Somerville R. S., Naab T., Burkert A., 2012b, MNRAS, 426, 237
Hirschmann M., De Lucia G., Iovino A., Cucciati O., 2013, MNRAS, 433, 1479
Hirschmann M., Dolag K., Saro A., Bachmann L., Borgani S., Burkert A., 2014a, MNRAS, 442, 2304
Hirschmann M., De Lucia G., Wilman D., Weinmann S., Iovino A., Cucciati O., Zibetti S., Villalobos Á., 2014b, MNRAS, 444, 2938
Hirschmann M., Naab T., Ostriker J. P., Forbes D. A., Duc P.-A., Davé R., Oser L., Karabal E., 2015, MNRAS, 449, 528
Hopkins P. F., Quataert E., Murray N., 2011, MNRAS, 417, 950
Hopkins P. F., Quataert E., Murray N., 2012a, MNRAS, 421, 3488
Hopkins P. F., Quataert E., Murray N., 2012b, MNRAS, 421, 3522
Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, MNRAS, 445, 581
Ilbert O. et al., 2010, ApJ, 709, 644
Kannan R., Stinson G. S., Macciò A. V., Brook C., Weinmann S. M., Wadsley J., Couchman H. M. P., 2014, MNRAS, 437, 3529
Karakas A. I., 2010, MNRAS, 403, 1413
Karim A. et al., 2011, ApJ, 730, 61
Kauffmann G., Colberg J. M., Diaferio A., White S. D. M., 1999, MNRAS, 303, 188
Kewley L. J., Ellison S. L., 2008, ApJ, 681, 1183
Khandai N., Di Matteo T., Croft R., Wilkins S., Feng Y., Tucker E., DeGraf C., Liu M.-S., 2015, MNRAS, 450, 1349
Kimm T. et al., 2009, MNRAS, 394, 1131
Knebe A. et al., 2015, MNRAS, 451, 4029
Kochanek C. S. et al., 2001, ApJ, 560, 566
Lackner C. N., Cen R., Ostriker J. P., Joung M. R., 2012, MNRAS, 425, 641
Lagos C. D. P., Lacey C. G., Baugh C. M., Bower R. G., Benson A. J., 2011, MNRAS, 416, 1566
Lagos C. d. P., Lacey C. G., Baugh C. M., 2013, MNRAS, 436, 1787
Li Y.-S., De Lucia G., Helmi A., 2010, MNRAS, 401, 2036
Lu Y., Mo H. J., 2007, MNRAS, 377, 617
Lu Y. et al., 2014, ApJ, 795, 123
Lu Y., Mo H. J., Wechsler R. H., 2015, MNRAS, 446, 1907
McNamara B. R., Nulsen P. E. J., 2007, ARA&A, 45, 117
McNamara B. R., Rohanizadegan M., Nulsen P. E. J., 2011, ApJ, 727, 39
Maiolino R. et al., 2008, A&A, 488, 463
Maiolino R. et al., 2012, MNRAS, 425, L66
Marchesini D. et al., 2007, ApJ, 656, 42
Marchesini D., van Dokkum P. G., Förster Schreiber N. M., Franx M., Labbé I., Wuyts S., 2009, ApJ, 701, 1765
Marchesini D., Stefanon M., Brammer G. B., Whitaker K. E., 2012, ApJ, 748, 126

Martin C. L., 2005, ApJ, 621, 227
Martin C. L., Shapley A. E., Coil A. L., Kornei K. A., Murray N., Pancoast A., 2013, ApJ, 770, 41
Matteucci F., Recchi S., 2001, ApJ, 558, 351
Mitchell P. D., Lacey C. G., Cole S., Baugh C. M., 2014, MNRAS, 444, 2637
Mo H. J., Mao S., 2002, MNRAS, 333, 768
Mo H. J., Mao S., 2004, MNRAS, 353, 829
Moe M., Arav N., Bautista M. A., Korista K. T., 2009, ApJ, 706, 525
Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121
Murante G., Monaco P., Borgani S., Tornatore L., Dolag K., Goz D., 2015, MNRAS, 447, 178
Muratov A. L., Kereš D., Faucher-Giguère C.-A., Hopkins P. F., Quataert E., Murray N., 2015, MNRAS, 454, 2691
Muzzin A. et al., 2013, ApJ, 777, 18
Newman S. F. et al., 2013, ApJ, 767, 104
Ostriker J. P., Choi E., Ciotti L., Novak G. S., Proga D., 2010, ApJ, 722, 642
Padovani P., Matteucci F., 1993, ApJ, 416, 26
Panter B., Jimenez R., Heavens A. F., Charlot S., 2007, MNRAS, 378, 1550
Peeples M. S., Shankar F., 2011, MNRAS, 417, 2962
Peeples M. S., Werk J. K., Tumlinson J., Oppenheimer B. D., Prochaska J. X., Katz N., Weinberg D. H., 2014, ApJ, 786, 54
Pérez-González P. G. et al., 2008, ApJ, 675, 234
Pettini M., Steidel C. C., Adelberger K. L., Dickinson M., Giavalisco M., 2000, ApJ, 528, 96
Popping G. et al., 2015, MNRAS, 454, 2258
Pozzetti L. et al., 2003, A&A, 402, 837
Pozzetti L. et al., 2007, A&A, 474, 443
Prochaska J. X., Hennawi J. F., Simcoe R. A., 2013, ApJ, 762, L19
Puchwein E., Springel V., 2013, MNRAS, 428, 2966
Rees M. J., Ostriker J. P., 1977, MNRAS, 179, 541
Rubin K. H. R., Prochaska J. X., Koo D. C., Phillips A. C., Martin C. L., Winstrom L. O., 2014, ApJ, 794, 156
Rudie G. C. et al., 2012, ApJ, 750, 67
Salim S. et al., 2007, ApJS, 173, 267
Santini P. et al., 2009, A&A, 504, 751
Saracco P. et al., 2006, MNRAS, 367, 349
Schaye J. et al., 2015, MNRAS, 446, 521
Somerville R. S., Davé R., 2015, ARA&A, 53, 51
Somerville R. S., Gilmore R. C., Primack J. R., Domínguez A., 2012, MNRAS, 423, 1992
Springel V., Di Matteo T., Hernquist L., 2005a, MNRAS, 361, 776
Springel V. et al., 2005b, Nature, 435, 629
Springel V. et al., 2008, MNRAS, 391, 1685
Steidel C. C., Giavalisco M., Pettini M., Dickinson M., Adelberger K. L., 1996, ApJ, 462, L17
Steinborn L. K., Dolag K., Hirschmann M., Prieto M. A., Remus R.-S., 2015, MNRAS, 448, 1504
Stinson G. S., Brook C., Macciò A. V., Wadsley J., Quinn T. R., Couchman H. M. P., 2013, MNRAS, 428, 129
Tacconi L. J. et al., 2010, Nature, 463, 781
Tacconi L. J. et al., 2013, ApJ, 768, 74
Thielemann F.-K. et al., 2003, in Hillebrandt W., Leibundgut B., eds, ESO/MPA/MPE Conf. Workshop Proc., From Twilight to Highlight: The Physics of Supernovae Supernova Nucleosynthesis and Galactic Evolution. Springer-Verlag, p. 331
Thomas D., Maraston C., Bender R., Mendes de Oliveira C., 2005, ApJ, 621, 673
Thomas D., Maraston C., Schawinski K., Sarzi M., Silk J., 2010, MNRAS, 404, 1775
Tremonti C. A. et al., 2004, ApJ, 613, 898
Tumlinson J. et al., 2013, ApJ, 777, 59
Veilleux S., Cecil G., Bland-Hawthorn J., 2005, ARA&A, 43, 769
Wang J., De Lucia G., Kitzbichler M. G., White S. D. M., 2008, MNRAS, 384, 1301
Wang W., Sales L. V., Henriques B. M. B., White S. D. M., 2014, MNRAS, 442, 1363
Weiner B. J. et al., 2009, ApJ, 692, 187
Weinmann S. M., Kauffmann G., van den Bosch F. C., Pasquali A., McIntosh D. H., Mo H., Yang X., Guo Y., 2009, MNRAS, 394, 1213
Weinmann S. M., Kauffmann G., von der Linden A., De Lucia G., 2010, MNRAS, 406, 2249
Weinmann S. M., Pasquali A., Oppenheimer B. D., Finlator K., Mendel J. T., Crain R. A., Macciò A. V., 2012, MNRAS, 426, 2797
White S. D. M., Rees M. J., 1978, MNRAS, 183, 341
White C. E., Somerville R. S., Ferguson H. C., 2015, ApJ, 799, 201
Yang X., Mo H. J., van den Bosch F. C., Pasquali A., Li C., Barden M., 2007, ApJ, 671, 153
Yates R. M., Henriques B., Thomas P. A., Kauffmann G., Johansson J., White S. D. M., 2013, MNRAS, 435, 3500
Zahid H. J., Dima G. I., Kudritzki R.-P., Kewley L. J., Geller M. J., Hwang H. S., Silverman J. D., Kashino D., 2014, ApJ, 791, 130
Zhu W., Feng L.-L., Fang L.-Z., 2011, MNRAS, 415, 1093

APPENDIX A: MILKY-WAY LIKE GALAXIES

As discussed in Section 2.2, our chemical evolution model cwas tuned to reproduce the distribution of Fe in the Milky Way disc, using simulations from the Aquarius project. When running the same model on a cosmological volume, we found that a few parameters needed to be slightly modified in order to match the normalization of the mass–metallicity relation observed in the local Universe. Specifically, we had to reduce both the feedback efficiency (from $\epsilon_{\rm re-heat} = 0.05$ to 0.02), and the re-incorporation factor (from $\gamma = 0.5$ to 0.1).

To verify how much these parameter changes affect the results discussed in our previous work, we have used the Millennium simulation to select Milky Way-like galaxies using criteria as close as possible to those used to select the Aquarius haloes. Specifically, we have considered only central galaxies of haloes with mass between $\sim 6.5 \times 10^{11}\,{\rm M}_{\odot}$ and $\sim 1 \times 10^{12}\,{\rm M}_{\odot}$ with no halo within a sphere of 1 Mpc with mass larger than half its mass.

Fig. A1 shows the distribution of stellar masses, cold gas masses and SFRs obtained for galaxies selected using these criteria in different models. The grey shaded histogram, in particular, refers to the model used in De Lucia et al. (2014b), while the coloured histograms correspond to the fiducial (lila), FIRE (orange), and zDEP (yellow) models used in this paper. The vertical dotted lines mark the observational estimates as given in De Lucia et al. (2014b), and the arrows indicate the median of the distributions. Considering the FIRE and zDEP models, that are our favourite schemes on the basis of the results presented in this paper, we have on average slightly lower stellar masses, larger cold gas masses, and larger SFRs with respect to results from the model used in De Lucia et al. (2014b). We note, however, that there is still a sizeable population (comparable in number to that found on the basis of the original model used in De Lucia et al. 2014b), whose physical properties are comparable to those estimated for our Galaxy.

Fig. A2 shows the distributions of [Fe/H] and [O/Fe] obtained for the Milky Way-like galaxies selected as discussed above, for the same models. In this case the changes are more dramatic (not surprisingly as we needed to introduce these modifications to change the overall normalization of the mass–metallicity relation in the first place): both the FIRE and zDEP models (as well as our fiducial model) predict average values for the [Fe/H] of Milky Way discs that are larger than those obtained in De Lucia et al. (2014b). The opposite is true for [O/Fe], although here the distribution measured for the stars in our Galaxy disc is very large and the median might be not very informative. We plan, in future work, to apply our new favourite scheme to the Aquarius simulations and reconsider the chemical properties of both the Milky Way and its satellites.

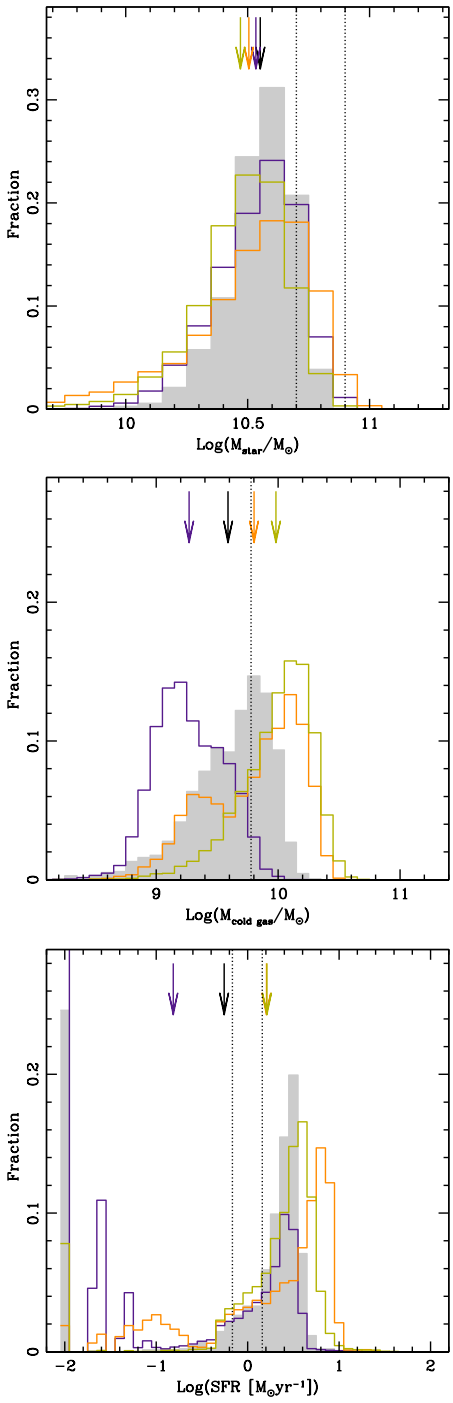

Figure A1. Distributions of galaxy stellar mass (top panel), cold gas mass (middle panel) and SFRs (bottom panel) for Milky Way like galaxies selected following the Aquarius-like criteria described in the text. The grey shaded area shows results obtained using the model described in De Lucia et al. (2014b), while the coloured histograms correspond to the fiducial (lila), FIRE (orange), and zDEP (yellow) models discussed in this work. Arrows correspond to the median of the distributions, while the vertical dotted lines show the (range of) observational estimate(s).

In a recent paper, Yates et al. (2013) have included an independently developed chemical enrichment model into the semi-analytic model by Guo et al. (2011). They show that their implementation reproduces, simultaneously, the chemical properties of Milky Way-like galaxies and the observed mass–metallicity relation in the local

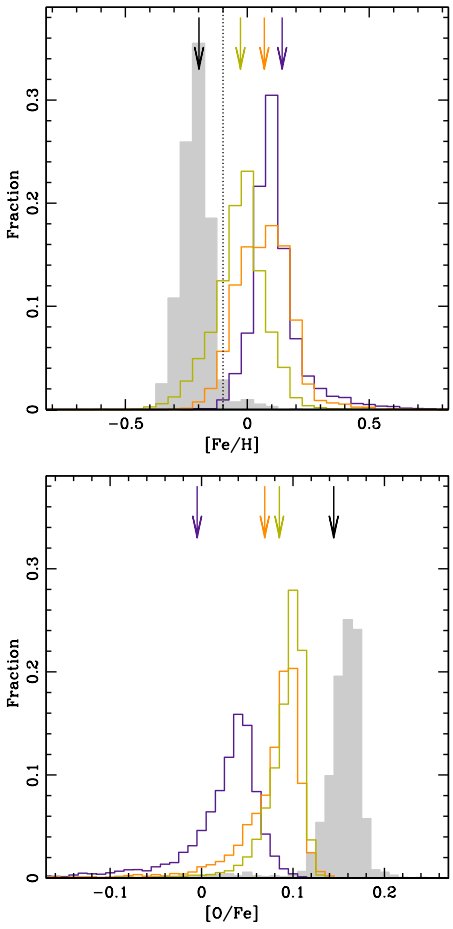

Figure A2. As in Fig. A1, but this time showing the average [Fe/H] (top panel) and [O/Fe] (bottom panel) of stars in the disc of Milky Way-like galaxies.

Universe. Their Milky Way-like galaxies are selected from the Millennium, but using criteria slightly different from those we have used above. In Figs A3 and A4, we show the same distributions considered above, but this time using the same criteria adopted in Yates et al. (2013). Specifically, we select only central galaxies in haloes with mass between $\sim 3 \times 10^{11}\,M_{\odot}$ and $\sim 3 \times 10^{12}\,M_{\odot}$, SFR measured below $z = 0.25$ between $1\,M_{\odot}yr^{-1}$ and $10\,M_{\odot}yr^{-1}$, and stellar bulge-to-total ratio smaller than 0.5. When considering the FIRE and zDEP models, these criteria bring the cold gas mass and SFR (by construction) of our Milky Way-like galaxies much closer to the observational estimates. This applies also for the average [Fe/H] of stars in the Milky Way discs. Therefore, our favourite models are able to reproduce, simultaneously, the metal content of Milky Way like galaxies and the overall normalization of the mass–metallicity relation of local galaxies (as for the model discussed in Yates et al. 2013), as well as the GSMF in the local Universe and its evolution up to $z \sim 3$.

APPENDIX B: RESOLUTION STUDY

In this section, we discuss to what extent our model predictions vary when increasing the resolution of the underlying dark matter simulation. To this aim, we take advantage of the MillenniumII simulation (Boylan-Kolchin et al. 2009), that adopts the same cosmology as the Millennium simulation, but corresponds to a smaller box (100 Mpc h^{-1} against 500 Mpc h^{-1}), five times better spatial

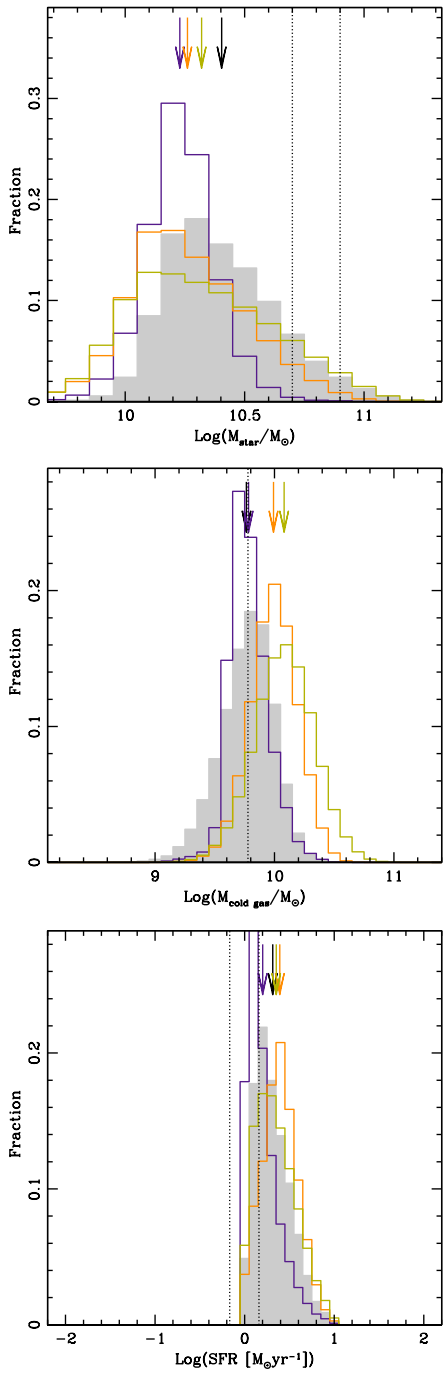

Figure A3. As in Fig. A1, but using the Yates-like criteria described in the text.

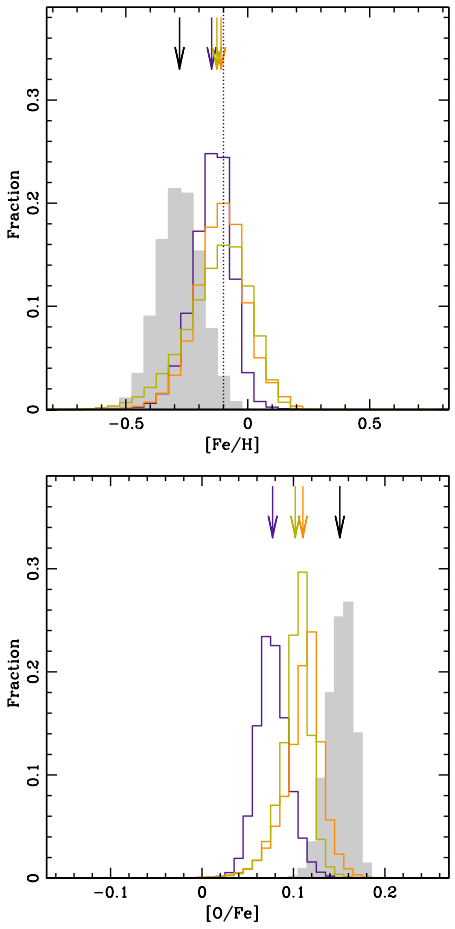

Figure A4. As in Fig. A2, but using the Yates-like criteria described in the text.

resolution (the Plummer equivalent softening of the Millennium II is 1.0 kpc h^{-1}) and 125 times better mass resolution (the particle mass of the Millennium II simulation is $6.9 \times 10^6\,M_\odot h^{-1}$).

Fig. B1 shows the evolution of the galaxy stellar mass function for the Fiducial, zDEP, and FIRE models based on the Millennium (thin solid lines) and the MillenniumII simulation (thick solid lines), and compares model predictions to observational measurements (as discussed in Fig. 1). For galaxies with masses between 10^9 and $10^{11}\,M_\odot$ in the FIRE and zDEP models, we find small differences (<0.2 dex) between the different resolution runs at high redshift $z > 0$. At redshift $z = 0$, some discrepancies are visible at the very massive end for all models, and at the low-mass end ($<10^{10}\,M_\odot$) for the fiducial model by maximum 0.2 dex. The former can be at least in part due to the small number statistics for very massive galaxies in the smaller box of the Millennium II. In addition, infalling (satellite) galaxies tend to be more gas rich in the higher resolution simulation, leading to more gas rich mergers on to central galaxies and thus, to more star formation.

Fig. B2 illustrates the evolution of the baryon conversion efficiencies as a function of halo mass for the fiducial, the zDEP, and the FIRE models, based on merger trees from the Millennium (thin solid lines) and the MillenniumII (thick solid lines) simulation. A 1000 particle, and thus resolved, halo in the Millennium simulation corresponds to a mass of $\sim 1 \times 10^{12}\,M_\odot h^{-1}$, and $\sim 9 \times 10^9\,M_\odot h^{-1}$ for the MillenniumII simulation.

We find that baryon conversion efficiencies only weakly converge with resolution, even if for a halo mass range of 10^{11}–$10^{12}\,M_\odot$ the differences are small. Conversion efficiencies of massive haloes in the higher resolution simulation are larger, which is a reflection of the larger number densities of massive galaxies in the Millennium-II run compared to the Millennium simulation as shown in Fig. B1.

At $z = 0$ and $z = 5$, the FIRE and zDEP predictions, based on the better resolved merger trees, are still in fairly good agreement with abundance matching predictions (grey shaded areas, Moster et al. 2013). At higher redshifts, abundance matching predictions are not available for halo masses below $10^{11}\,M_\odot$. When extrapolating their fits, baryon conversion efficiencies of the high-resolution FIRE and zDEP models, which are slightly reduced compared to the lower resolution models, are in even better agreement with the abundance matching trends.

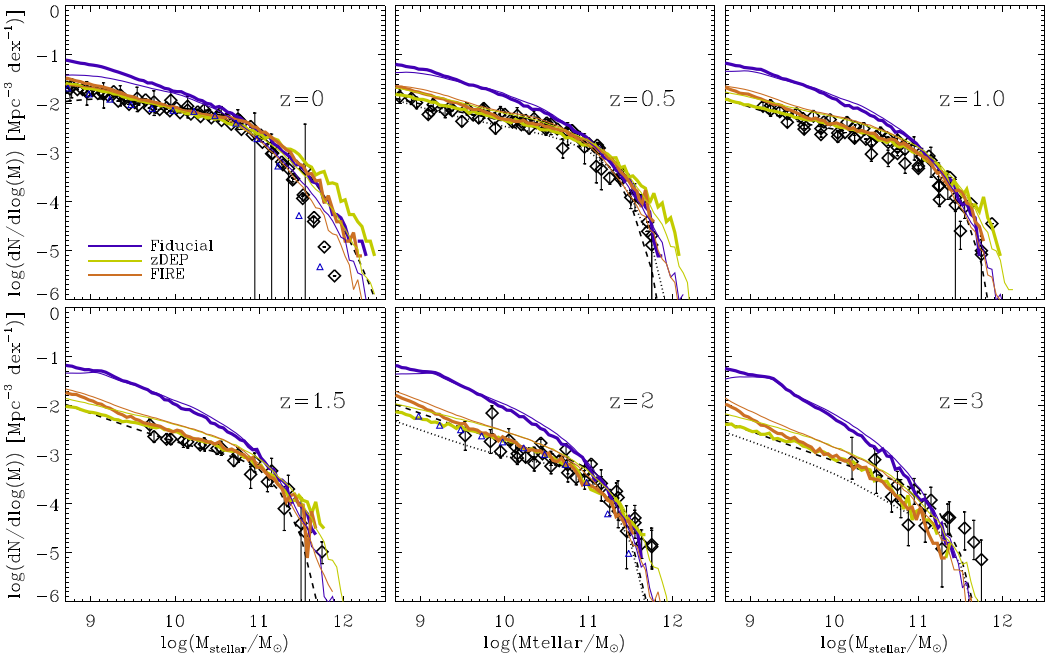

Figure B1. Evolution of the GSMF for the fiducial, zDEP and FIRE models based on the Millennium (thin solid lines), and on the Millennium-II trees (thick solid lines), and compared to observational measurements (black symbols and black lines).

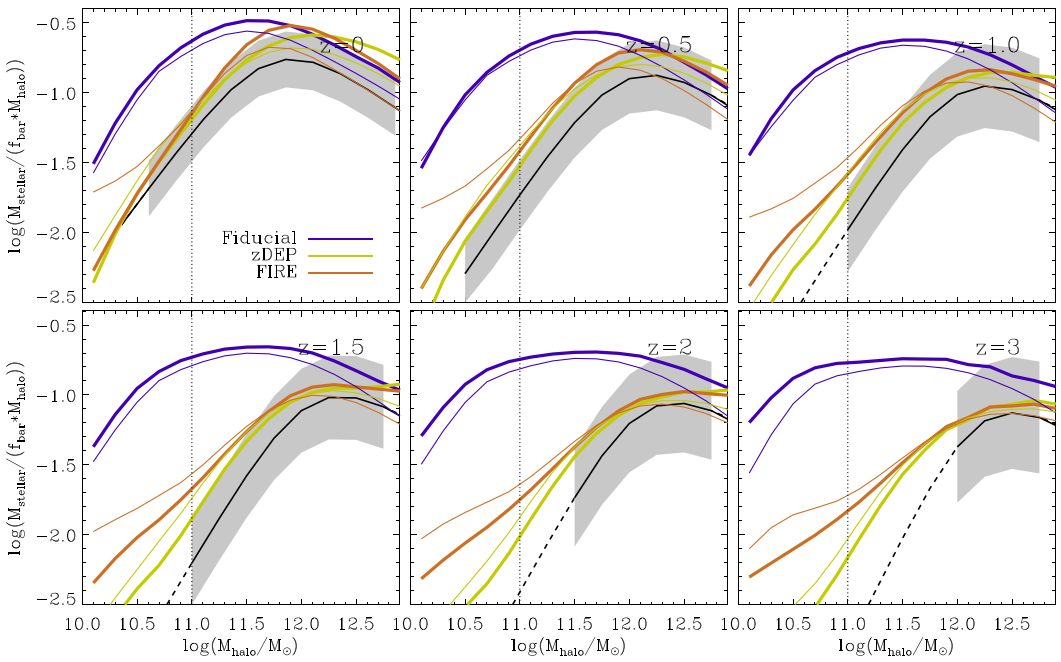

Figure B2. Evolution of the average baryon conversion efficiencies for the fiducial, zDEP and FIRE models based on the Millennium (thin solid lines) and on the Millennium-II trees (thick solid lines), compared to predictions from subhalo abundance matching methods (black lines with grey shaded areas). The black dotted line indicates the mass limit, where model predictions start to get more strongly affected by resolution.

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