An empirical model for the star formation history in dark matter haloes

Zhankui Lu,1⋆ H. J. Mo,1 Yu Lu,2 Neal Katz,1 Martin D. Weinberg,1 Frank C. van den Bosch3 and Xiaohu Yang4,5

1Department of Astronomy, University of Massachusetts, Amherst, MA 01003-9305, USA
2Kavli Institute for Particle Astrophysics and Cosmology, Stanford, CA 94309, USA
3Astronomy Department, Yale University, PO Box 208101, New Haven, CT 06520-8101, USA
4Center for Astronomy and Astrophysics, Shanghai Jiao Tong University, Shanghai 200240, China
5Key Laboratory for Research in Galaxies and Cosmology, Shanghai Astronomical Observatory, Nandan Road 80, Shanghai 200030, China

Accepted 2013 December 18. Received 2013 November 27; in original form 2013 June 10

ABSTRACT

We develop an empirical approach to infer the star formation rate in dark matter haloes from the galaxy stellar mass function (SMF) at different redshifts and the local cluster galaxy luminosity function (CGLF), which has a steeper faint end relative to the SMF of local galaxies. As satellites are typically old galaxies which have been accreted earlier, this feature can cast important constraint on the formation of low-mass galaxies at high redshift. The evolution of the SMFs suggests the star formation in high-mass haloes (>10^{12} h^{-1} M_{\odot}) has to be boosted at high redshift beyond what is expected from a simple scaling of the dynamical time. The faint end of the CGLF implies a characteristic redshift z_e ≈ 2 above which the star formation rate in low-mass haloes with masses <10^{11} h^{-1} M_{\odot} must be enhanced relative to that at lower z. This is not directly expected from the standard stellar feedback models. Also, this enhancement leads to some interesting predictions, for instance, a significant old stellar population in present-day dwarf galaxies with M_* ≤ 10^{8} h^{-2} M_{\odot} and steep slopes of high-redshift stellar mass and star formation rate functions.

Key words: methods: statistical – galaxies: formation – galaxies: haloes.

1 INTRODUCTION

With the advent of multiwaveband deep surveys from the Hubble Space Telescope (HST), galaxy stellar mass (luminosity) functions (SMF or LF) can now be determined up to a redshift of z ≈ 8 (e.g. Bradley et al. 2012). These observations have revealed a number of interesting properties about the galaxy population. The amplitude of the SMF increases as the Universe evolves from high to low redshift, while the characteristic stellar mass does not change significantly (e.g. Perez-Gonzalez et al. 2008; Marchesini et al. 2009, 2012; Santini et al. 2012). This suggests that galaxies with masses larger than the characteristic mass may have formed as early as z = 3, while galaxies of lower masses continue to grow in mass and/or in number density all the way to the present epoch. For local galaxies, large redshift surveys have now made it possible to determine the SMF of galaxies down to 10^7 M_{\odot} (e.g. Baldry et al. 2012). Such surveys have revealed that the Schechter function may not be sufficient to describe the observed LF of galaxies, especially for galaxies in massive clusters. Instead there seems to be a marked upturn at the faint end (M_∗ = 5 \log_{10} h > -17) of the cluster galaxy luminosity function (CGLF; Popesso et al. 2006; Barkhouse, Yee & Lopez-Cruz 2007; Jenkins et al. 2007; Milne et al. 2007; Banados et al. 2010; Wegner 2011). This feature has also been found for galaxies in the general field regardless of their environments, which are usually referred to as field galaxies (Blanton et al. 2005; Drory et al. 2009; Pozzetti et al. 2010; Baldry et al. 2012; Loveday et al. 2012), although the upturn appears shallower than that for cluster galaxies.

Theoretically, the standard Λ cold dark matter (ΛCDM) model assumes that galaxies form and evolve in dark matter haloes (see Mo, van den Bosch & White 2010, for an overview). However, the observed redshift evolution of the galaxy population, in which massive galaxies form earlier, appears at odds with the simplest expectations from the hierarchical nature of dark matter halo formation in the ΛCDM scenario, where small haloes form first and subsequently merge to form larger ones. Furthermore, the halo mass function predicted by the standard ΛCDM model has a characteristic shape very different from that of the galaxy SMF. Although it can also be described by a Schechter-like function, the halo mass function has a much steeper slope at the low-mass end, and an exponential cutoff at a much larger mass scale. Thus, if star formation was equally efficient in haloes of different masses, the ΛCDM model predicts too many low-mass (Klypin et al. 1999; Moore et al. 1999) and high-mass galaxies. Finally, the existence of a stronger

*E-mail: lv@astro.umass.edu
faint end upturn in galaxy clusters is not expected from simple predictions of the ΛCDM model. In fact the slope of the maximum circular velocity function of subhaloes (substructure within distinct haloes) of massive clusters is quite similar to that of distinct haloes (Klypin, Trujillo-Gomez & Primack 2011).

The apparent tension between the observations and the standard ΛCDM model indicates the complexity of the baryonic physics that regulates the galaxy evolution. The effects of the baryonic physics can be studied directly using hydrodynamic simulations or semi-analytic models (hereafter SAMs). Hydrodynamic simulations find two basic modes by which baryonic matter is accreted into galaxies, cold mode and hot mode, depending on the mass of the host halo (Keres et al. 2005, 2009). Other processes, especially those concerning star formation and feedback, which determine how efficiently cold gas turns into stars, are still beyond the capability of current computational resources to model and uncertain subgrid prescriptions are used to model them. In a SAM (e.g. Kauffmann, White & Guiderdoni 1993; Somerville &Primack 1999; Cole et al. 2000; Kang et al. 2005; Croton et al. 2006; Guo et al. 2011; Lu et al. 2011), all the relevant processes (e.g. cooling, star formation, feedback, etc.) are modelled using parametrized prescriptions either derived from analytic models or calibrated with numerical simulations. Although SAMs have produced useful predictions about the galaxy population, many of the physical processes involved are still poorly understood at present and some uncertain assumptions have to be made about a number of model ingredients, such as the efficiency of star formation and feedback.

In recent years, much effort has been made to establish the statistical connection between galaxies and dark matter haloes via the conditional luminosity function (CLF; e.g. van den Bosch, Yang & Mo 2003; Yang, Mo & van den Bosch 2003) or the halo occupation distribution (e.g. Jing, Mo & Borner 1998; Peacock & Smith 2000; Seljak 2000; Scoccimarro et al. 2001; Berlind & Weinberg 2002). Such empirically established galaxy-dark matter halo connections describe how galaxies with different properties occupy haloes of different masses and, therefore, provide important insights into how galaxies form and evolve in dark matter haloes (Mo et al. 2010). With data obtained from deep, multiwavelength surveys, attempts have been made to establish the relation between galaxies and their dark matter haloes out to high-z using Abundance Matching (AM). This technique links galaxies of a given luminosity or stellar mass to dark matter haloes of a given mass by matching the observed abundance of the galaxies to the halo abundance obtained from the halo mass function, typically also accounting for subhaloes. This approach was first used by Mo & White (1996) and Mo, Mao & White (1999) to model the number density and clustering of Lyman-break galaxies. More recently, several studies have used this AM technique to probe the galaxy-dark matter connection out to z ≃ 5 (e.g. Vale & Ostriker 2004; Conroy, Wechsler & Kravtsov 2006; Shankar et al. 2006; Conroy & Wechsler 2009; Behroozi, Conroy & Wechsler 2010; Guo et al. 2010; Moster et al. 2010; Béthermin, Doré & Lagache 2012; Yang et al. 2012; Béthermin et al. 2013). One can infer the average star formation rate (SFR) for haloes of different masses at different redshifts from the stellar mass–halo mass relation (e.g. Behroozi, Wechsler & Conroy 2012; Moster, Naab & White 2013; Yang et al. 2013; Wang et al. 2013).

How does one learn about galaxy formation within the ΛCDM paradigm? The usual approach using cosmological hydrodynamic simulations or SAMs is to make ab initio models of galaxy formation including all the physical processes that one thinks are important. One then makes predictions from these models and compares them with observations. If the model does not compare well with the observations then one changes the model, typically either by changing the parametrizations of the previously included physical processes or by adding new physical processes. In this paper, we make a first step at a qualitatively different approach from most past work. We attempt to ask in general terms another question. What do the observations require of the galaxy formation model? In other words, we attempt to let the data speak for themselves in a way that is as independent as possible of any model assumptions. In fact, we only make three main assumptions: that we live in a Λ-dominated CDM Universe from which we extract dark matter halo merger trees, that the SFR of the central galaxies in such haloes depends only on the halo mass and the redshift, and that when a galaxy becomes a satellite its star formation is quenched exponentially and it can eventually merge with the central galaxy on a dynamical friction time-scale. Such simplifications allow us to explore a broad range of hypotheses without being restricted by our poor understanding of baryonic physics, such as cooling, star formation and feedback. We put our model on a firm statistical footing using Bayesian inference. We start with a very simple model to describe the SFR of central galaxies and only increase its complexity if the data requires it, assessed using Bayes ratios. We build up our model in a stepwise manner, increasing the complexity as we add more constraining data, instead of using a single complicated model, so that we can clearly see how each SFR model is constrained by the different observations and to see whether or not a more complex model is required.

This paper is organized as follows. The generic form of our empirical model is described in Section 2. In Section 3, the observational constraints, including the SMF at different redshifts and the CGLF, are described together with the method we use to constrain the model parameters. In Section 4, we show how we increase the complexity of our model in a series of steps. This results in a series of nested model families that can reproduce more and more of the observational constraints. Model I is able to match the SMF of local galaxies, Model II reproduces the evolution of the SMFs and Model III is the minimum model that can also explain the CGLF. In Section 5, we present the detailed predictions of Model II, and Model III for the stellar mass–halo mass relation, SFR–halo accretion rate relation, the cosmic SFR density, the SFR function, the specific star formation rate (sSFR) as a function of stellar mass and the conditional SMF. Finally, we discuss and summarize our results in Section 6.

Throughout the paper, we use a ΛCDM cosmology with Ω_{m,0} = 0.26, Ω_{Λ} = 0.74, Ω_{B} = 0.044, h = 0.71, n = 0.96 and σ_8 = 0.79. This set of parameters is consistent with the Wilkinson Microwave Anisotropy Probe 5 data (Dunkley et al. 2009; Komatsu et al. 2009).

2 THE EMPIRICAL MODEL

2.1 Dark halo merger histories

Our empirical model is built upon dark halo merger histories generated using the algorithm developed by Parkinson, Cole & Helly (2008). The algorithm is based on the Extended Press-Schechter formalism and it is tuned to agree with the conditional mass functions of merger trees (Cole et al. 2008) constructed from the Millennium Simulation (MS: Springel et al. 2005). As shown by Jiang & van den Bosch (2013), this algorithm is in good agreement with simulations in terms of many other properties, such as mass assembly history, merger rate and unevolved subhalo mass function. For this work, the merger trees span a redshift range 0 ≤ z ≤ 15 with 100
2.2 Star formation in central galaxies

We assume that the SFR of the central galaxy of a halo at a given redshift \( z \) is completely determined by the virial mass of the host halo, \( M_{\text{vir}}(z) \), and \( z \). Then the SFR can be written as

\[
\text{SFR} \equiv \dot{M}_* = \dot{M}_*(M_{\text{vir}}(z), z).
\]

The above equation describes an average among haloes of a given mass at a given \( z \). It ignores any variations in the SFR that owe to variations in the formation histories of individual haloes of a given mass and any large-scale environmental effects. Note though that the \( M_t[M_{\text{vir}}(z)] \) can still result in haloes of the same \( M_{\text{vir}}(z) \) having different \( M_* \) simply because of scatter in the halo assembly histories.

Guided by the observational demand that the star formation efficiency must be suppressed in both low- and high-mass haloes (Yang et al. 2003), we assume the following form for the dependence of the SFR on redshift and halo mass:

\[
\dot{M}_* = \mathcal{E} \frac{f_{\text{b}} M_{\text{vir}}}{\tau_0} (1 + z)^\alpha (X + 1)^\beta \left( \frac{X + \mathcal{R}}{X + 1} \right)^\gamma,
\]

where \( \mathcal{E} \) is an overall efficiency; \( f_{\text{b}} \) is the cosmic baryonic mass fraction; \( \tau_0 \) is a dynamic time-scale of the haloes at the present day, set to be \( \tau_0 \equiv 1/(10H_0) \) and \( \kappa \) is fixed to be 3/2 so that \( \tau_0/(1 + z)^{3/2} \) is roughly the dynamical time-scale at redshift \( z \). We define the quantity \( X \) to be \( X \equiv M_{\text{vir}}/M_* \), where \( M_* \) is a characteristic mass and \( \mathcal{R} \) is a positive number that is smaller than 1. Hence, in our model the SFR of a galaxy depends on its dark matter halo mass through a piecewise power law, with \( \alpha, \beta, \gamma \) being the three power indices in the three different mass ranges separated by the two characteristic masses \( M_c \) and \( \mathcal{R} M_c \).

\[
M_* \propto \begin{cases} 
M_{\text{vir}}^{\alpha} & \text{if } M_{\text{vir}} \gg M_c, \\
\mathcal{R}^{-\beta} M_{\text{vir}}^{\beta} & \text{if } M_c \geq M_{\text{vir}} > \mathcal{R} M_c, \\
M_{\text{vir}}^{\gamma} & \text{if } M_{\text{vir}} \ll \mathcal{R} M_c.
\end{cases}
\]

The simplest model is the one where all the model parameters, \( \mathcal{E}, \alpha, \beta, \gamma, M_c \) and \( \mathcal{R} \) are redshift-independent. In what follows, we will make more parameters redshift-dependent whenever the observational data demands it.

2.3 Star formation in satellite galaxies

For satellites, the SFR has to be modelled differently. As a dark matter halo gets accreted by a larger one, it becomes a subhalo and experiences environmental effects such as tidal stripping, galaxy harassment (Moore et al. 1996) and tidal disruption. The satellite galaxy associated with the subhalo may also be affected as it orbits in the host halo. For example, the diffuse gas initially in the subhalo and the cold gas disc may get stripped by the ram pressure or tidal forces of the host halo. Consequently, the star formation in the satellite can be suppressed or even quenched. Indeed, observations clearly show that satellite galaxies have larger quenched fractions than centrals of the same stellar mass (e.g. Balogh, Navarro & Morris 2000; van den Bosch et al. 2008; Wetzel, Tinker & Conroy 2012).

We use a simple \( \tau \) model to describe the suppression of star formation in a satellite after it is accreted:

\[
M_{*,\text{sat}}(t) = M_{*,\text{sat}}(t_0) \exp \left( -\frac{t - t_0}{\tau_{\text{sat}}} \right),
\]

where \( t_0 \) is the time when the satellite is accreted into its host, \( M_{*,\text{sat}}(t_0) \) is the SFR of the satellite galaxy at \( t = t_0 \), and \( \tau_{\text{sat}} \) is a timescale characterizing the decline of the star formation. We adopt the following model for the characteristic time

\[
\tau_{\text{sat}} = \tau_{\text{rt},0} \exp \left( -\frac{M_*}{M_{*,\text{sat}}} \right),
\]

where \( \tau_{\text{rt},0} \) is the exponential decay time for a galaxy with a stellar mass of \( M_{*,\text{sat}} \), with both of these being free parameters in our model.

The choice of \( \tau_{\text{rt},0} \) is motivated by the fact that massive galaxies tend to be more quenched in star formation than low-mass galaxies, independent of environment (Peng et al. 2010; Wetzel et al. 2013). For central galaxies, this trend is naturally reproduced by assuming that in massive haloes the star formation efficiency decreases with halo mass (as long as \( \alpha < 0 \)). For satellites, however, such a halo-mass dependence of star formation efficiency does not work, because the host subhalo mass is expected to decrease with time owing to stripping.

2.4 Merging and stripping of satellite galaxies

A halo with mass \( M_{\text{sat}} \) accreted by a larger (primary) halo with mass \( M_{\text{pry}} \) will gradually sink towards the centre of the primary halo owing to dynamical friction. Using N-body simulations, Boylan-Kolchin, Ma & Quataert (2008) found that the dynamical friction time-scale, \( \tau_{\text{merger}} \), depends on the mass ratio between the satellite halo and the host halo, as well as on the orbital parameters of the satellite halo \( \eta \) at the time of accretion:

\[
\tau_{\text{merger}} = 0.216 \left( \frac{M_{\text{pry}}/M_{\text{sat}}}{\ln(1 + M_{\text{pry}}/M_{\text{sat}})} \right)^{1.3} \exp((1.9\eta)\tau_{\text{dyn}}),
\]

where \( \tau_{\text{dyn}} = r_{\text{vir}}/V_{\text{vir}} \) is the dynamical time of the halo, with \( r_{\text{vir}} \) and \( V_{\text{vir}} \) being the virial radius and virial velocity of the halo, respectively; \( \eta \) is the orbital circularity, which is the ratio between the orbital angular momentum and the orbital angular momentum of a circular orbit with the same energy. Following Zentner et al. (2005), we assume \( \eta \) to have the following distribution,

\[
P(\eta) \propto \eta^{1.2}(1 - \eta)^{1.2}.
\]

For each satellite galaxy at the time of accretion, we draw a value of \( \eta \) from this distribution and use it in equation (6) to estimate a dynamical friction time-scale.

We assume that a satellite galaxy merges with the central galaxy of the primary halo in a time \( \tau_{\text{merger}} \) after accretion. We further assume that only a fraction \( f_{\text{TS}} \) of the stellar mass of the satellite is added on to the central galaxy and the rest of its stellar mass becomes halo stars. In our model, we treat \( f_{\text{TS}} \) as a free parameter.

2.5 Spectral synthesis and metal enrichment

Given a star formation model, we use the procedure described above, together with a halo merger tree, to predict the star formation history of a given galaxy. To convert that star formation history to a stellar mass, which takes into account mass-loss owing to stellar evolution, and to calculate luminosities in different bands, we adopt a Chabrier (2003) initial mass function (IMF) and the stellar population synthesis model of Bruzual & Charlot (2003). We correct the
3 OBSERVATIONAL CONSTRAINTS AND LIKELIHOOD FUNCTIONS

In this paper, we use ‘field galaxies’ to refer to all galaxies independent of their environment. The data used in this paper are listed in Table 1, including the SMF of field galaxies at four different redshift bins and the z-band CGLF.

Table 1. Observed constraints. Column 2 lists the redshift ranges of the SMFs and the mean redshift of the z-band CGLF. Column 3 lists the error model we use in the likelihood function: linear means normal distribution and log means log-normal distribution. Column 4 lists the sources of the data.

| Redshift | Error model | Reference |
|----------|-------------|-----------|
| SMF | 0, 0.06 | linear | Baldry et al. (2012) |
| SMF | 1, 1.3 | log | Perez-Gonzalez et al. (2008) |
| SMF | 2, 3 | log | Marchesini et al. (2009) |
| SMF | 3.19, 4.73 | log | Stark et al. (2009) |
| CGLF (z-band) | 0.1 (mean) | linear | Popesso et al. (2006) |

3.1 The stellar mass functions of field galaxies

The SMF of the local Universe (z \approx 0) is from Baldry et al. (2012). The galaxies are selected from the Sloan Digital Sky Survey (SDSS) DR6 down to a magnitude limit of m_\text{r} \approx 19.8 with a sky coverage of 143 deg^2. Redshift measurements for galaxies fainter than the SDSS redshift survey limit are made with the Galaxy And Mass Assembly (GAMA) survey. Stellar masses for individual galaxies are estimated from their ugriz photometry using a spectral synthesis model with the assumption of a Chabrier (2003) IMF. Owing to surface brightness incompleteness, the measured number density for galaxies with masses below 10^8 h^{-2} M_\odot can only be considered as lower limits. In our analysis we, therefore, only use data points above 10^8 h^{-2} M_\odot. All models presented here predict mass functions that lie above the data points of Baldry et al. (2012) for M_\star < 10^8 h^{-2} M_\odot, as required.

In addition to local galaxies, we also use the SMFs of galaxies in the following three redshift bins: 1.0 < z < 1.3 from Perez-Gonzalez et al. (2008); 2.0 < z < 3.0 from Marchesini et al. (2009) and 3.19 < z < 4.73 from Stark et al. (2009). Stellar masses in Perez-Gonzalez et al. (2008) and Stark et al. (2009) were estimated using broad-band photometry assuming a Salpeter IMF. Following Stark et al. (2007), we convert these masses into those corresponding to a Chabrier IMF by dividing them by a factor of 1.4. The stellar masses in Marchesini et al. (2009) were derived using a pseudo-Kroupa (2001) IMF, but it was shown that adopting a Chabrier (2003) IMF does not have a significant effect on their SMFs so we use their SMF directly without any corrections.

To make predictions for the field galaxy SMF, we use 5000 dark matter haloes drawn from a power-law distribution \mathcal{f}(M_{\text{vir}}) \propto M_{\text{vir}}^a \text{d}M_{\text{vir}} in the mass range \log M_{\text{vir}} \text{d}M_{\text{vir}} < 5 \times 10^8 h^{-1} M_\odot < 5 \times 10^{14} h^{-1} M_\odot. Including haloes outside this mass range does not change our results significantly. The fact that our merger trees do not resolve progenitor haloes with M_{\text{vir}}(z) < 2 \times 10^9 h^{-1} M_\odot implies that we implicitly assume that star formation is suppressed in haloes with masses below this ‘threshold’. In other words, our star formation model (equation 2) should be augmented with M_\star = 0 for M_{\text{vir}}(z) < 2 \times 10^9 h^{-1} M_\odot. When we calculate the stellar mass/LF of galaxies, the predicted number of galaxies associated with a halo of mass M_{\text{vir}} is assigned a weight \omega = f_{\text{SMF}}(M_{\text{vir}})/f(M_{\text{vir}}) where f_{\text{SMF}}(M_{\text{vir}}) is the halo mass function from Sheth, Mo & Tormen (2001) for our adopted cosmology.

3.2 The composite luminosity function of cluster galaxies

As an additional constraint, we also use the z-band composite LF of cluster galaxies obtained by Popesso et al. (2006). The z-band luminosities are less affected by dust extinction than their bluer counterparts, making the comparison between our model and the data less affected by dust corrections. Furthermore, for a stellar population older than 100 Myr the predicted z-band luminosity depends only weakly on metallicity, making our results less sensitive to the chemical evolution model described above (see Fig. 1). Finally, the use of the composite LF, a weighted average over a number of individual clusters, reduces statistical uncertainties arising from variances in the formation history of clusters.

The composite LF of Popesso et al. (2006) is based on 69 clusters selected from ROSAT X-ray data using cross-identifications with galaxies in the SDSS. Once the LF of individual clusters are known, the composite LF can be evaluated using

$$N_{ij} = \frac{N_{ij}}{m_j} \sum_i N_{ij} \frac{N_{ij}}{N_{ij}} ,$$

where $N_{ij}$ is the number of galaxies in bin $i$ of the composite LF and $m_j$ is the mass of the $j$th cluster.
where $N_{ij}$ is the number of galaxies in the $j$th bin of the composite LF, $N_{i0}$ is the number of galaxies in the $i$th cluster, $N_{i0}$ is the normalization factor for the $i$th cluster, which is the total number of the member galaxies brighter than the magnitude limit $m_{\text{lim}}$ at the cluster redshift, $m_{i}$ is the number of clusters contributing to the $j$th bin and

$$N_{ij} = \sum N_{i0}.$$  \hfill (10)

We follow the same method to calculate the composite LF in our model predictions.

To use the CGLF as a constraint, it is necessary to know the masses of those clusters accurately, since a systematic error in cluster mass can lead to an error in the amplitude of the predicted CGLF. Unfortunately, the mass estimates for the clusters in the Popesso et al. (2006) sample are uncertain, and so it is dangerous to use the overall amplitude of the CGLF to constrain the model. To bypass this problem, we treat the ratio between the real mass of a cluster $M_{\text{real}}$ and the measured value $M_{\text{obs}}$,

$$e_{M} \equiv \frac{M_{\text{real}}}{M_{\text{obs}}},$$ \hfill (11)

as a free parameter, and we renormalize the LF of an individual cluster by $e_{M}$ (i.e. we marginalize over potential systematic biases in the inferred cluster masses).

To make predictions for the CGLF, we generate 23 merger trees for haloes with a present-day mass distribution similar to that observed. The mass resolution adopted for these merger trees is also $2 \times 10^{10} \, h^{-1} M_{\odot}$. We confirmed that this number of haloes is sufficiently large so that the variance between the different realizations is smaller than the uncertainties in the observational data.

### 3.3 The likelihood function and sampling algorithm

The likelihood function describes the probability of the data given the model and its parameters. A rigorous likelihood function includes all the processes in the data acquisition. For the SMFs, we study in this paper, the data acquisition process includes deriving a stellar mass from multiband photometry using a stellar population synthesis model, correcting for incompleteness, weighting each galaxy sample according to its corresponding survey volume, and so on. As a result, the uncertainties in individual stellar mass bins are not independent. As is demonstrated in Lu et al. (2011), the covariance for binned data may change the posterior distribution substantially. Unfortunately, the covariance matrix in the data used here is not available and we have to assume that the SMF in different bins is independent and that the likelihood is approximated by a Gaussian function. As shown in Appendix A, with these assumptions the likelihood function can be written as

$$\ln L(\Phi_{\text{obs}}|\theta) = C - \frac{1}{2} \sum_{i} \frac{[\Phi_{i,\text{obs}} - \Phi_{i,\text{mod}}(\theta)]^2}{\sigma_{i,\text{obs}}^2},$$ \hfill (12)

where $\Phi_{i}$ and $\sigma_{i}$ are either defined in linear space or logarithmic space, depending on the observational data (see Table 1), and $C$ is an unimportant factor.

To efficiently sample the high dimensional parameter space, we make use of the MULTINEST method developed by Feroz, Hobson & Bridges (2009), which implements the nested sampling algorithm of Skilling (2006). We have compared this method with the Markov chain Monte Carlo implemented in Lu et al. (2011), Tempered Differential Evolution. Both methods are designed to deal with probability distributions with multiple modes and strong degeneracies between model parameters. For the problem in this paper, we found the two give identical results, but that the number of likelihood evaluations required by MULTINEST is smaller by more than a factor of 10. A more detailed description of the method is given in Appendix B.

### 4 MODELS OF STAR FORMATION IN DARK MATTER HALOES

We explore a series of nested model families, based on the generic parametrization given by equation (2). At each step, we increase the complexity of the model and check whether the fit to the observational constraints is acceptable and whether any improvement to the fit is sufficient to justify the increased complexity by comparing the posterior predictions with the constraints. In addition, we also make use of the Bayes factor to avoid developing an overcomplicated model (Table 2). In Bayesian statistics, the Bayes factor is

$$K = \frac{p(M_{\text{mod}}|D)}{p(M_{\text{obs}}|D)},$$ \hfill (13)

where $p(M|D)$ is the probability of model $M$ given data $D$. It is obtained by marginalizing over all the model parameters of the posterior distribution. The advantage of the Bayes factor is that it automatically includes a penalty for too much complexity in the model. The values of $\ln [p(M|D)]$ for all the models discussed in the text are listed Table 2.

The model families explored are summarized in Table 3, and the prior ranges of the model parameters are also given in the table. For some parameters, the prior ranges are motivated by other studies. For instance, $\alpha$ is only allowed to be negative because of the strong quenching of star formation found in massive galaxies. Some prior choices are made to avoid unphysical modes. A prior range that is too wide sometimes contains unphysical modes with high probability, making the physical modes negligible. Such priors have to be excluded. In general, the prior covers a pretty large range of possibilities. The median and 95 per cent interval of the posterior model parameters are also given in the table.

### 4.1 The fiducial model family (Model I)

We start with the simplest case in which all the parameters in equation (2) are assumed to be time independent, so that the SFR is determined completely by the halo mass and by the dynamical...
Table 3. A list of the model parameters of the four main model families considered. The prior ranges, the medians and the 95 per cent credible intervals of the model parameters are listed. \( M_e \) is in units of \( 10^{10} h^{-1} M_\odot \), and \( M_{c,e} \) is in units of \( 10^{10} h^{-2} M_\odot \).

| Parameter | Model I, SMF(z = 0) | Model II, SMF | Model III, SMF+CGLF | Model IV, SMF+CGLF |
|-----------|---------------------|---------------|----------------------|---------------------|
|           | median prior 95 per cent range | median prior 95 per cent range | median prior 95 per cent range | median prior 95 per cent range |
| \( \alpha \) | -1.6 \([-5, 0] \) | -3.6 \([-5, 0] \) | -3.1 \([-5, 0] \) | -4.2 \([-5, 0] \) |
| \( \beta \) | 3.5 \([0, 5] \) | 1.8 \([0, 5] \) | 2.6 \([0, 5] \) | 2.9 \([0, 5] \) |
| \( \gamma \) | 0.92 \([-1, 3] \) | 1.9 \([-1, 3] \) | 2.6 \([-1, 3] \) | 2.5 \([-1, 3] \) |
| \( M_e \) | 1.4 \([0, 4] \) | 1.9 \([0, 4] \) | 1.8 \([0, 4] \) | 1.6 \([0, 4] \) |
| \( R \) | -0.83 \([-2, 0] \) | -0.96 \([-2, 0] \) | -1.1 \([-2, 0] \) | -0.91 \([-2, 0] \) |
| \( E \) | -0.21 \([-2, 1] \) | -0.27 \([-2, 1] \) | -0.32 \([-2, 1] \) | 0.36 \([-2, 1] \) |
| \( H_0 \tau_{a,0} \) | -0.71 \([-2, 0] \) | -1.1 \([-2, 0] \) | -1.37 \([-2, 0] \) | -0.86 \([-2, 0] \) |
| \( M_{c,e} \) | -1.1 \([-2, 0] \) | -1.4 \([-2, 0] \) | -1.4 \([-2, 0] \) | -1.5 \([-2, 0] \) |
| \( \tau_{a} \) | 0.65 \([0, 1] \) | 0.099 \([0, 1] \) | 0.13 \([0, 1] \) | 0.11 \([0, 1] \) |
| \( \log(10)(M_e) \) | 1.4 \([0, 4] \) | 1.9 \([0, 4] \) | 1.8 \([0, 4] \) | 1.6 \([0, 4] \) |

Time-scale of the halo at the time in question. We call this Model I. First, we use only the SMF at \( z \approx 0 \) Baldry et al. (2012) to constrain the model parameters. Fig. 2 shows that Model I fits the constraining data well (see the first panel). For reference, the constrained model parameters are listed in Table 3. We accomplish this fit using only nine parameters compared to the 13 (more physically motivated) parameters of the SAM used to fit similar data in Lu et al. (2011). However, as one can see from the other three panels of Fig. 2, the posterior of the model predicts too few massive galaxies at high-\( z \) compared to the observational data. Next, we use the SMFs at all four different redshifts, as listed in Table 1, to constrain the model parameters. The resulting fits to the data are shown in Fig. 3. This model family still fails to reproduce the SMFs at the bright end: it overestimates the number density of bright galaxies at \( z = 0 \) but underestimates the number density at \( z = 2.5 \) and \( z = 4 \). This suggests that massive galaxies form too late, and that the SFR at high-\( z \) in massive haloes needs to be boosted.

4.2 Fixing the bright-end problem (Model II)

As an attempt to fix the bright-end problem identified above, we consider a second model family (Model II) that allows \( \alpha \) to be redshift dependent. We assume that the redshift-dependence be given by the following power law,

\[
\alpha = \alpha_0 (1 + z) \gamma',
\]
where both $\alpha_0$ and $\alpha'$ are introduced as new free parameters. The SFR in massive haloes with $M_{\text{vir}} > M_c$ will be enhanced at high-$z$ if $\alpha'$ is a negative number.

Once again we use the four SMFs listed in Table 1 as observational constraints. Fig. 4 compares the posterior predictions with the constraining data. This model family matches the observational data over the entire redshift range from $z = 0$ to $z = 4$. The natural log of the Bayes ratio between Model II over Model I (constrained by the same data sets) is $-31.2 \pm 120 = 88.8$, making the odds of preferring Model II over Model I given the data a whopping $e^{88.8}$ to one (Table 2). At $z = 0$, the contribution of satellites to the total mass function is lower than that of the central galaxies, with a satellite to central ratio of about 1/2 at the faint end, decreasing to about 1/3 for higher stellar masses (see the top-left panel). These results are in good agreement with those obtained by Yang et al. (2008) based on galaxy groups selected from the SDSS (see also Mandelbaum et al. 2006; Cacciato et al. 2013). The credible intervals for the parameters are listed in Table 3. $\alpha'$ lies between $-1.09$ and $-0.54$, which means significant evolution in the SFR of massive central galaxies. The constrained $f_{\text{TS}}$ lies below 0.3, indicating that less than 30 per cent of the stars in disrupted satellites will be accreted by the centrals, with the rest of them remaining as halo stars (or intracluster stars in the case of massive haloes). This is consistent with early studies. Yang et al. (2012) find that in haloes as massive as $10^{14} h^{-1} M_\odot$, the total mass in disrupted satellites exceeds the mass of centrals, giving rise to intracuster stars. Also Purcell, Bullock & Zentner (2007), Conroy, Wechsler & Kravtsov (2007) and Watson, Berlind & Zentner (2012) suggest that the majority of the stars in the satellites will become halo stars when the host subhaloes are disrupted.

Fig. 5 shows the posterior prediction of Model II for the SFR as a function of halo mass at different redshifts (left-hand panel) and as a function of redshift for different halo mass bins (right-hand panel). The SFR peaks at around $10^{12} h^{-1} M_\odot$ at each redshift. In haloes with low masses, the SFR increases rapidly with halo mass, roughly as $\text{SFR} \propto M_{\text{vir}}^{5.5}$, quite independent of redshift. The SFR decreases with halo mass for haloes above $10^{12} h^{-1} M_\odot$. The decrease is more pronounced at low redshifts, becomes weaker as one goes to higher redshifts, and is quite weak by $z = 4$. It should be noted, however, that at $z \approx 4$ the SFR beyond $10^{13} h^{-1} M_\odot$ can only be considered
as an extrapolation because the number density of such high-mass haloes is extremely small at such high redshifts. For haloes with masses $\sim 10^{12} h^{-1} M_{\odot}$, the SFR increases with redshift roughly as $(1 + z)^{2.3}$, which is roughly proportional to the halo mass accretion rate as a function of redshift (Genel et al. 2008; Neistein & Dekel 2008; McBride, Fakhouri & Ma 2009). This increase with redshift is faster for more massive haloes, while for $M_{\text{vir}} < 10^{11} h^{-1} M_{\odot}$, $\text{SFR} \propto (1 + z)^{1.5}$.

In the following, we examine how Model II matches the $z$-band CGLF. The posterior prediction is shown in the left-hand panel of Fig. 6. The normalization of the prediction is adjusted to match the observations at $M_z = -5 \log_{10}(h) \approx -20$ to account for a potential systematic bias in the cluster masses. The model prediction is consistent with the observational data at the bright end but it underpredicts the number of dwarf galaxies with $M_z = -5 \log_{10}(h) > -17$. In particular, the predicted faint end of the LF is roughly a power law, in contrast with the observational data that shows a significant upturn. To test whether the model family Model II can accommodate the observed LF of cluster galaxies, we carry out a new inference, this time including the $z$-band LF as an additional data constraint. The predicted CGLF still fails to match the observed one at the faint end, as shown in the right-hand panel of Fig. 6. Thus, it is unlikely to find a model in the parameter space of Model II to simultaneously match the SMFs and CGLF.

4.3 Fixing the faint-end problem of the cluster galaxy luminosity function (Model III)

There are at least two possibilities that could lead to the differences between the SMF of field galaxies and the CGLF. Since the over-abundance of dwarf galaxies is more prominent in galaxy clusters, some environmental effects specific to high-density environments may cause the strong upturn. However, many processes such as tidal stripping and tidal disruption are destructive, which would suppress the number of satellite galaxies rather than enhance it. It is possible that relatively massive galaxies could have experienced significant mass-loss, moving them to lower masses. However, since more massive galaxies are less abundant, it is difficult to make such a scenario work in detail. Thus, unless environmental effects in clusters operate in a way very different from what is generally believed, it is
difficult to explain the overabundance of dwarf galaxies in clusters with such effects.

Another possibility is that the environmental dependence of the LF may be the result of some time-dependent processes. Indeed, galaxies in a cluster are expected to form early in their progenitor haloes. Thus, if star formation in dark matter haloes were different at high redshift when the majority of cluster galaxies formed, then the LF of cluster galaxies and field galaxies could show different behaviours owing to their systematically different formation times. One concrete example is the pre-heating model proposed in Mo & Mao (2002), where the intergalactic medium (IGM) is assumed to be pre-heated at some high redshift, so that the star formation in dark matter haloes proceeds differently before and after the pre-heating epoch. Such pre-heating may owe to the formation of pancakes, as envisaged in Mo et al. (2005), or to an episode of starbursts and active galactic nucleus (AGN) activity (Mo & Mao 2002), or owe to heating by high-energy gamma rays generated by blazars, as envisaged in Chang, Broderick & Pfrommer (2011). In all these cases, the pre-heating is expected to occur around $z \approx 2$ and the pre-heated entropy of the IGM is a few times $10^7$ KeV cm$^2$. In what follows, we consider a generic model family (Model III) inspired by the physical processes discussed above. We allow $\gamma$, which controls star formation in low-mass haloes, to be time dependent in such a way that it changes from $\gamma_a$ at high-$z$ to $\gamma_c$ at low-$z$, with a transition redshift $z_c$. Specifically, we assume that

$$
\gamma = \begin{cases} 
\frac{\gamma_a - \gamma_b}{z_{c} + 1} + \gamma_b & \text{if } z < z_c \\
\gamma_a & \text{otherwise} 
\end{cases}
$$

(15)

Note that Model II is a special case of Model III, with $\gamma' = 0$. If $z_c = 0$ and $\gamma_b = 0$ then $\gamma$ is a simple power law of $(1 + z)$ with an index of $\gamma'$. Before we add this to Model II, thereby creating Model III, we want to explore the possibility that adding this behaviour to Model I and not allowing the massive end slope $\alpha$ to depend on redshift might also provide a viable fit to the SMFs at the four redshifts. We refer to this as Model IIb. We perform an inference with Model IIb using just the SMFs as data constraints. As one can see from Table 2, the SMFs prefer Model II over Model IIb by a probability of $e^{4.3}$ to one, but they still prefer Model IIb over Model I.

Figs 7 and 8 compare the posterior predictions with the constraining data, which are the SMFs and the CGLF. We see that Model III can accommodate both observational data sets. In particular, the upturn in the faint end of the CGLF is well reproduced (see Fig. 8). The two data sets prefer Model III over Model II by a probability of $e^{8.4}$ to one (Table 2). Comparing Fig. 7 with Fig. 4 shows that the SMFs predicted by Model III are steeper than those predicted by Model II, particularly at high redshift. The more recent results obtained by Santini et al. (2012) from recent Wide Field Camera 3 (WFC3) data, which are also plotted in Fig. 7 for comparison, are consistent with the model predictions. Also, at the low-mass end ($M_* \approx 10^8 h^{-1} M_\odot$) the satellite fraction predicted by Model III is higher than that predicted by Model II, and eventually overtakes the fraction of central galaxies of similar masses. This could be checked by studying galaxy groups in deeper future surveys.

The posterior model parameters obtained for Model III are listed in Table 3. As one can see, except for $\gamma$, the values of all the other parameters obtained from Model III are quite similar to those obtained from Model II. For the new model parameters introduced in Model III, we have $\gamma' = -4.5$, $z_c = 2.2$, $\gamma_a = 2.5$ and $\gamma_b = 0.89$. The significant difference of $\gamma'$ from zero implies that a redshift-dependent $\gamma$ is preferred by the data, and the fact that $\gamma_a$ is much larger than $\gamma_b$ indicates that the SFR increases with halo mass much faster at low redshift ($z < z_c$) than at $z > z_c$. The value of $z_c$ is constrained to $2.2 \pm 0.5$ and, interestingly, is very close to the value expected from the pre-heating scenarios mentioned above. We will come back to discuss further the implications of these results. For both Model II and Model III, the parameters that control the star formation in satellites are not well constrained. This indicates that the observational constraints used in this work are

Figure 7. The SMFs predicted by the posterior of Model III constrained by both the SMFs and the observed $z$-band CGLF. The yellow band encompasses the 95 per cent credible intervals of the posterior distribution, and the red solid lines are the medians. The black data points with error bars are the observational constraints. For comparison, we also plot, as coloured points, the results of Santini et al. (2012) obtained from recent WFC3 data. For $z \approx 0$ galaxies (upper-left panel), the contributions of satellites and centrals are plotted separately.

Figure 8. The CGLF predicted by the posterior of Model III constrained by both the SMFs and the observed $z$-band CGLF. The yellow band encompasses the 95 per cent credible interval of the posterior distribution, and the red solid line is the median. The black data points with error bars are the observational constraints.
not particularly sensitive to the star formation after infall of the galaxies. As suggested by many other works, other observations, such as the clustering (Yang et al. 2012; Watson & Conroy 2013) or the quenched fraction of satellites (Wetzel et al. 2013) could lead to a better understanding of the star formation history of satellite galaxies.

Fig. 9 shows the model prediction of Model III for the SFR as a function of halo mass and redshift. For haloes more massive than $10^{11} M_{\odot}$, the results are almost identical to those given by Model II. For less massive haloes, however, Model III predicts a clear transition at $z_\gamma \approx 2$ from a phase of elevated star formation at higher $z$ to a phase of reduced star formation at lower $z$, as can be seen in the right-hand panel of Fig. 9. This behaviour owes to the adding of the CGLF data instead of the use of the more extended Model III.

When using only the four SMFs as constraints, the less restrictive model III is only preferred over Model II by a factor $e^{0.6} \approx 1.8$, and the posterior for the SFR is similar to that of Model II shown in Fig. 5. However, when including data constraints from the cluster LF, the Bayes factor increases to $e^{26.4}$.

4.4 Are more general model families necessary?

As shown above, Model I successfully matches the observed $z \approx 0$ SMF, Model II successfully matches the observed galaxy SMFs over the redshift range from $z = 0$ to $z = 4$, and Model III successfully matches not only the observed galaxy SMFs but also the $z$-band CGLF at $z \approx 0$. Perhaps the observational data could be fit even better with a more complex model? Is the resulting SFR as a function of halo mass and redshift unique or are there other models that could match the observational data equally well but predict a SFR$(M_{\text{vir}}, z)$ that is very different from that predicted by the previous models? To investigate these questions, we performed three more inferences. First, using only the $z \approx 0$ SMF as data constraint for Model II, Table 2 shows that this data only marginally prefers the more complex Model II over Model I and hence Model I is sufficient to describe the $z \approx 0$ SMF. Next, we performed an inference with Model III using the four observed galaxy SMFs as data constraints only. Again, Table 2 shows that these data only modestly prefer the more complex Model III over Model II, making Model II sufficient to describe the SMFs over the range $z = 0$ to 4.

Finally, to test the robustness of the Model III predictions using both the observed galaxy SMFs and also the CGLF, we consider another model family, Model IV, which allows even more model parameters to be redshift dependent. Specifically, we write

$$\gamma = \begin{cases} \gamma_a \left( \frac{z+1}{z_\gamma+1} \right)^{\gamma_a} & \text{if } z < z_c \\ (\gamma_a - \gamma_b) \left( \frac{z+1}{z_\gamma+1} \right)^{\gamma_a} + \gamma_b & \text{otherwise} \end{cases}$$

(16)

$$M_c = M_{c,0}(z+1)^{\mu_c};$$

(17)

$$R = R_0(z+1)^{\rho_c};$$

(18)

$$\kappa = \frac{3}{2} + \kappa'',$$

(19)

with $\gamma', \mu, \rho, \kappa'$ introduced as four new free parameters. Model IV reduces to Model III if these four parameters are set to be zero.

In Figs 10 and 11, we show the posterior predictions of Model IV for the galaxy SMFs and CGLF, respectively, compared with the corresponding constraining data. As expected, the larger parameter number Model IV fits the constraining data better. In terms of the Bayes Factor, the ratio between Models IV and III is $e^{4.9} \approx 134$, which represents a marginally significant improvement. An improvement in the fit of the $z \approx 0$ SMF near the knee is evident, but no other significant improvements are noticeable. Furthermore, the SFR as a function of halo mass and redshift predicted by Model IV is qualitatively similar to that predicted by Model III, as shown in Fig. 12. The only significant difference is that the star formation in massive haloes with $M_{\text{vir}} > 10^{12.5} h^{-1} M_{\odot}$ predicted by Model IV increases faster with increasing redshift than that predicted by Model III. Thus, the SFR as a function of halo mass and redshift predicted by Model III does not seem to owe to the particular parameterizations adopted but rather reflects requirements of the observational data. Unfortunately, such tests can never be exhaustive; there is always the possibility that some other model could match the data constraints and yet give a different SFR$(M_{\text{vir}}, z)$. In this sense, all our conclusions and predictions are restricted to the model families that we actually explore.
just the SMFs at different redshifts, and Model III, constrained using both the SMFs and the $z$-band CGLF. We feel that presenting Model II’s predictions could still be informative because: (1) it is possible that the stronger faint-end upturn of the CGLF turns out to be spurious, and (2) Model II is similar to the results of other past work (Behroozi et al. 2012; Yang et al. 2012; Behroozi, Wechsler & Conroy 2013; Moster et al. 2013; Mutch, Croton & Poole 2013) that also do not use the CGLF as a constraint.

5.1 The star formation history and the stellar mass assembly of galaxies

Let us first look at the star formation and assembly histories of present-day galaxies. Fig. 13 shows the predictions of Model II (dashed lines) and Model III (solid lines). Each of the curves shown is the average over a large number of galaxies with the same halo mass at $z = 0$, as indicated in the lower-right panel. For comparison, the average final stellar mass of the central galaxies for each halo mass is also indicated in the lower-right panel. The top-left panel shows the star formation history, which takes into account the contribution of all the progenitors. The two models yield similar results for galaxies with present stellar masses $M_*$ $> 10^{10} h^{-1} M_\odot$ (halo masses $M_{\rm vir}$ $> 10^{12} h^{-1} M_\odot$). The SFR shows a peak that shifts from $z \approx 4$ for massive haloes ($M_{\rm vir}$ $\sim 10^{13} h^{-1} M_\odot$) to $z \approx 1$ for Milky Way mass haloes ($M_{\rm vir}$ $\sim 10^{12} h^{-1} M_\odot$). It declines exponentially after the peak, and the rate of decline strongly correlates with halo mass (or galaxy mass), from a very fast decline for massive haloes to an almost constant SFR for haloes with $M_{\rm vir}$ $\sim 10^{12} h^{-1} M_\odot$. For galaxies in haloes with $M_{\rm vir}$ $< 10^{12} h^{-1} M_\odot$, the star formation histories predicted by the two models have different characteristic shapes, although the predicted final stellar masses are similar. The star formation histories of dwarf galaxies predicted by Model II show a rapid increase before becoming almost flat. For Model III, on the other hand, the average star formation history in low-mass haloes is bimodal: it starts with a high value at $z > z_c \approx 2$, declines with time to a minimum at $z = z_c$, and then increases to an extended period of an almost constant star formation. Hence, a ‘smoking gun’ difference between models II and III is that the latter predicts a much larger fraction of old stars in (central) dwarf galaxies than model II.

Indeed, observations of nearby dwarf galaxies indicate that such an old stellar population is ubiquitous. With the use of deep HST imaging, individual stars of nearby dwarf galaxies can be resolved, and the colour–magnitude diagram (CMD) can be constructed to obtain the detailed star formation histories of these galaxies. Using this technique, Weisz et al. (2011) investigated 60 dwarf galaxies within a distance of 4 Mpc, many of which are field galaxies located outside the Local Group. These galaxies cover a wide range of morphological types, including dEs, dIrrs and dSpirals. They found that, on average, the dwarf galaxies formed 60 per cent of their stars by $z \approx 2$ and 70 per cent by $z \approx 1$, regardless of morphological type. In the top-right panel of Fig. 13, we compare our model predictions with the CMD-inferred star formation histories (SFHs) obtained by Weisz et al. (2011). While the predictions of Model III are in qualitative agreement with the data, Model II predicts an age distribution of stars that is clearly too skewed towards relatively young stars.

Finally, the bottom panel shows the assembly history: the total stellar mass contained in the main progenitor branch of a halo as a function of time, normalized by the final stellar mass. Note that the assembly history includes stellar mass growth both by in situ star formation and by mergers with satellite galaxies. Although the most
Empirical model of star formation history

Figure 12. The SFR as a function of halo mass (left-hand panel) and redshift (right-hand panel) predicted by Model IV. The solid lines are the medians of the posterior prediction and the bands are the 95 per cent credible intervals.

Figure 13. Left-hand panels: the star formation history (upper panel) and assembly history (lower panel) of present-day central galaxies in haloes of different masses. The mass of the host halo and the stellar mass of the galaxy (both at the present time) are indicated in the lower-right panel, both in units of $h^{-1} M_\odot$. The lines are the medians and the shaded areas encompass the 95 per cent ranges of the posterior predictions. Top-right panel: the predicted star formation history of dwarf central galaxies in haloes with masses $\approx 3 \times 10^{10} h^{-1} M_\odot$ compared with the measurements of Weisz et al. (2011) for 60 dwarf galaxies of different types. In all panels, the predictions of Model II are shown by dashed lines while the corresponding predictions of Model III are shown by solid lines with the same colour.

A massive model galaxy formed most of its stars as early as $z \approx 4$, it is assembled much later; about half of its final mass is added at $z < 2$ (see the red curves). On average, Milky Way mass galaxies experienced a dramatic increase in their masses after $z \approx 2$; less than 10 per cent of their present-day stellar masses were assembled by $z = 2$. Thus, to identify the progenitors of present-day Milky Way mass galaxies at $z \approx 2$, one has to look at galaxies with stellar masses $\approx 10^9 M_\odot$. Note that the 95 per cent range of the posterior...
model prediction is quite broad for both massive and low-mass galaxies, suggesting that better observational data are required to provide more stringent constraints on the model.

5.2 The stellar mass–halo mass relation of central galaxies

The left-hand panels in Fig. 14 show the stellar mass to halo mass ratio as a function of halo mass at different redshifts as predicted by Model II. The dispersion owing to parameter variance and a variance in the merger histories of the host haloes are shown separately. These results are similar to those obtained from earlier investigations (e.g. Conroy & Wechsler 2009; Behroozi et al. 2012; Leauthaud et al. 2012; Yang et al. 2012; Moster et al. 2013). The ratio shows a broad peak around $10^{12} h^{-1} M_\odot$ with a gradual shift towards larger halo masses at high redshifts. The right-hand panels show the prediction of Model III. Compared with the predictions of Model II, we see a different evolution in haloes with masses below $2 \times 10^{11} h^{-1} M_\odot$, where the stellar mass to halo mass ratio is independent of halo mass at high redshifts, owing to an almost constant star formation efficiency.

5.3 Star formation rate and halo mass accretion rate

One way to understand the star formation history in a halo is to examine how the SFR correlates with the mass accretion rate of the host halo. To do this, we follow Behroozi et al. (2013) and define a star formation efficiency factor, which is the SFR at a given redshift, $M_\star(z)$, divided by the mean mass accretion rate at the same redshift, $\dot{M}_{\text{vir}}(z)$, times the universal baryon fraction $f_\text{b}$ (we refer to $f_\text{b} M_{\text{vir}}(z)$ as the baryonic accretion rate):

$$\epsilon_{\text{SFR}}(z) = \frac{M_\star(z)}{f_\text{b} \dot{M}_{\text{vir}}(z)}. \quad (20)$$

The left two panels of Fig. 15 show the star formation efficiency as a function of $M_\text{vir}$, and the right two panels show the same star formation efficiency as a function of $M_\star$. The predictions of Model II (upper two panels) and Model III (lower two panels) are almost identical for host haloes with masses above $10^{12} h^{-1} M_\odot$ (stellar masses above $10^{10} h^{-1} M_\odot$). For low-mass haloes, the two models behave quite differently.

Let us first look at the results for Model II. The star formation efficiency at $z \approx 4$ is strongly peaked at $M_\text{vir} \approx 10^{12} h^{-1} M_\odot$, with a peak value $\epsilon_{\text{SFR}} \approx 1/2$. The position and height of the peak depend mildly on redshift; at $z \approx 0$ it shifts to $M_\text{vir} \approx 4 \times 10^{11} h^{-1} M_\odot$, with a peak value $\epsilon_{\text{SFR}} \approx 0.8$. $\epsilon_{\text{SFR}}$ increases (decreases) with halo mass as a steep power law at the low (high) mass end. These results are similar to those obtained by Bouche et al. (2010), Behroozi et al. (2013) and Tacchella, Trenti & Carollo (2013). In the toy model proposed by Bouche et al. (2010), the SFR in haloes with masses between $3 \times 10^{11} h^{-1} M_\odot$ and $2 \times 10^{12} h^{-1} M_\odot$ follows the baryonic accretion rate, and is completely quenched in haloes outside this range. Those studies also imply that the star formation efficiency shows no dependence on redshift. Our results, on the other hand, reveal a slow but steady increase of $\epsilon_{\text{SFR}}$ near the peak with decreasing redshift, which owes to a decrease in the halo assembly rate, $\dot{M}_{\text{vir}}$, of haloes with masses $< 10^{12} h^{-1} M_\odot$ at low-$z$. While in haloes that are much more massive than $10^{12} h^{-1} M_\odot$, the star formation efficiency increases with redshift. Such evolution with redshift agrees with the results from Yang et al. 2013; Béthermin et al. 2013.

Model III predicts a different behaviour for haloes with masses $< 10^{11} h^{-1} M_\odot$ than Model II or the simple models proposed by Bouche et al. (2010); Behroozi et al. (2013); Yang et al. (2013); Béthermin et al. (2013), which implies a simple but strong quenching of star formation. Beyond $z_e$, at $z \approx 4$, the SFR is roughly about
1/10 of the baryonic accretion rate, and is independent of the host halo mass. The star formation is quenched abruptly at \( z \approx z_c \) if the halo is much smaller than \( 10^{11} \, h^{-1} \, M_\odot \) and makes a mild recovery at \( z \approx 0 \).

### 5.4 The cosmic star formation history

We compare the predicted star formation rate density (SFRD) of the universe with observations. Results are shown for Model II (upper panels) and Model III (lower panels) in Fig. 16. The total SFRD is decomposed into contributions by haloes of different masses (left panels) and into contributions by galaxies of different stellar masses (right panels). The observations shown in the figure are compiled by Hopkins & Beacom (2006) and Bouwens et al. (2012). The data points shown here are produced using a Chabrier IMF. At \( z \leq 2 \), the observational results from different sources are consistent with each other, and so are the predictions of both Model II and Model III. All of them show a fast decline, by an order of magnitude, towards low-\( z \) beginning at \( z = 2 \). In both Model II and Model III, the predicted total SFRD in this redshift range owes mainly to star formation in haloes with masses between \( 3 \times 10^{11} \, h^{-1} \, M_\odot \) and \( 3 \times 10^{12} \, h^{-1} \, M_\odot \), about Milky Way mass. As one can see from Figs 5 and 9, the SFR in such haloes is proportional to \((1 + z)^{2.3}\). However, at \( z > 3 \) the two models behave differently. Model II predicts a rapid decline in the total SFRD towards high-\( z \), as the abundance of \( 10^{12} \, h^{-1} \, M_\odot \) haloes decreases towards higher redshift while in dwarf haloes, which are abundant at high-\( z \), the star formation is strongly suppressed. The SFRD at \( z > 3 \) predicted by Model III is substantially higher, mainly because the SFR in low-mass haloes \((<3 \times 10^{11} \, h^{-1} \, M_\odot)\) or dwarf galaxies \((<10^{9} \, h^{-2} \, M_\odot)\) is boosted at \( z > z_c \) in this model. This difference provides an observational test to distinguish these two models. Unfortunately, current observational results of the SFRD are still quite uncertain. A simple comparison between observations and our model predictions shows that the data of Hopkins & Beacom (2006) favours Model III but that of Bouwens et al. (2012) favours Model II. The difference in the data owes to the uncertainty in dust corrections (e.g. Reddy & Steidel 2009). Also, corrections for sample incompleteness contain large uncertainties. One derives the high-\( z \) SFRD by integrating the observed UV LF. Unfortunately, the faint-end behaviour of the high-\( z \) LF is usually poorly constrained, and the exact value of the assumed faint-end slope can affect the derived SFRD significantly. It may still be that the dust correction used by Bouwens et al. (2012) is correct, but that one underestimates the total cosmic SFR at high-\( z \) because one misses a large number of low-mass galaxies.

### 5.5 The star formation rate function

In Fig. 17, we show the predicted SFR functions for galaxies at different redshifts. The characteristic SFR clearly decreases with decreasing redshift, by a factor of almost 100 from \( z = 4 \) to \( z = 0 \). Assuming that the faint part of the distribution can be fit by a power law, the power-law index predicted by Model III changes significantly from roughly \(-2.0\) at \( z = 4 \) to roughly \(-1.2\) at \( z = 0 \). For Model II the change in the faint-end slope is much more moderate, from roughly \(-1.5\) at \( z = 4 \) to \(-1.2\) at \( z = 0 \). For comparison, we also show the SFR function derived by Smit et al. (2012) from the UV LF of galaxies. We see that Model II significantly underpredicts the number density of galaxies at the lower redshifts.

Figure 16. The cosmic star formation rate density (SFRD) as a function of redshift predicted by Model II (upper panels) and Model III (lower panels). The solid lines and the bands are the medians and 95 per cent credible intervals, respectively. The total SFRD is shown as the grey band. The data points are taken from Hopkins & Beacom (2006) (black dots) and Bouwens et al. (2012) (the shaded area and the data points with error bars). In the left-hand panel, we plot the contributions by haloes of different masses separately, while in the right-hand panel, we plot the contributions by galaxies of different stellar masses.

Figure 17. The SFR functions at different redshifts predicted by Model II (upper panels) and Model III (lower panels). The solid lines are the medians and the bands encompass the 95 per cent credible intervals. The data points are the observational results from Smit et al. (2012).
low-SFR end, while Model III matches the data much better. Note that the observed SFR functions at the high-SFR ends are lower than the predictions of both Model II and Model III. One possible reason for this discrepancy is that the highest SFR galaxies are dusty and could be missed in UV observations.

5.6 The specific star formation rate

In Fig. 18, we show the specific star formation rate (sSFR; defined to be the SFR divided by the stellar mass at different redshifts) as a function of stellar mass at different redshifts. Here again we compare between the predictions of Model II (left-hand panels) and Model III (right-hand panels). For a given stellar mass, the sSFR increases with redshift. On the other hand, for a given redshift, the sSFR declines rapidly with galaxy mass as the mass goes beyond a critical mass, which increases from \( \approx 10^{10} h^{-2} M_\odot \) at \( z = 0 \) to \( \approx 5 \times 10^{10} h^{-2} M_\odot \) at \( z = 4 \). For galaxies between \( 10^9 h^{-2} M_\odot \) and the critical mass, the sSFR is almost independent of stellar mass, in qualitative agreement with the observations (e.g. Daddi et al. 2007; Noeske et al. 2007a). Model II and Model III differ in their predictions for low-mass galaxies with stellar masses \(< 10^9 h^{-2} M_\odot \). For Model II, the weak correlation between sSFR and stellar mass extends all the way down to such galaxies. For Model III, however, the correlation is much more complicated: the sSFR and stellar mass show no significant correlation at \( z = 0 \), show a strong positive correlation between \( z = 1 \) and \( z = 2 \), and show a weak correlation at higher redshifts. For low-mass galaxies at high-\( z \), Model III predicts a lower sSFR compared to Model II, because these galaxies in Model III form their stars earlier than galaxies of the same mass in Model II. In Model III, the sSFR in low-mass galaxies drops dramatically at the critical redshift \( z \approx 2 \) after a significant amount of stars have already formed at higher \( z \), making the sSFR in dwarf galaxies much lower than that in more massive galaxies at the same epoch.

At \( z \approx 0 \), the sSFR in dwarf galaxies catches up with those in more massive galaxies, because of the growth of their haloes and because of the strong mass dependence of the SFR at the low-mass end at low-\( z \).

Observations indicate that there is a negative correlation between sSFR and stellar mass for dwarf star-forming galaxies (Noeske et al. 2007a), a trend that appears contrary to the predictions of Model III shown in the right-hand panel. However, the existence of such a correlation in the observational data is still uncertain owing to sample incompleteness. Indeed, empirical models based on such correlations (Noeske et al. 2007b; Leitner 2012) suggest that low-mass galaxies form most of their stars in the past few billion years, in apparent contradiction with the star formation histories inferred directly from the CMDs of stars (e.g. Weisz et al. 2011), which seem to support Model III (see Section 5.1). More data on the detailed SFHs of isolated dwarf galaxies is required to better discriminate between these different models. Note that spectral energy distribution (SED) modelling is used to infer SFRs and stellar masses. If these SED models do not include the ‘bimodal’ SFHs, such as those predicted by Model III, significant systematic errors in the inferred \( M_\star \) and \( M_\ast \) of dwarf galaxies can occur. It is, therefore, important to check the magnitudes of such systematic effects.

5.7 The conditional stellar mass functions

We also make predictions for the Conditional Stellar Mass Function (hereafter CSMF), which is the SMF of galaxies hosted by haloes of a given mass. We show them in Fig. 19 together with the observational results of Yang et al. (2008). Owing to the detection limits, the CSMF below \( 10^9 h^{-2} M_\odot \) is either noisy or unavailable. The predictions of both Model II and Model III above the detection limit are consistent with the observational data. (Note that a quantitative comparison requires the prediction be convolved with the effects of the group finder (Reddick et al. 2013), which is beyond the scope of this article.)

Figure 18. The specific star formation rate (sSFR) versus stellar mass at different redshifts in Model II (left-hand panels) and Model III (right-hand panels). The upper panels show the medians of the posterior predictions as well as their 96 per cent ranges (bands). The lower panels show the predictions of the best-fitting model parameters, and the variance among individual haloes owing to their different merger histories (error bars).

Figure 19. The conditional stellar mass functions for four different halo mass bins as indicated. The solid lines with bands are the predictions of Model III, while the dashed lines are the predictions of Model II. The distribution of centrals and satellites are in blue and red, respectively. The data points are from Yang, Mo & van den Bosch (2008).
of this paper.) Compared with Model II, Model III predicts more dwarf galaxies, with masses below the current detection limit, not only in massive clusters but also in low-mass groups. One expects this behaviour because in Model III dwarf galaxies in massive clusters are fossils of a relative global enhancement of star formation activity in dwarf haloes in the high-$z$ Universe. This boosted star formation at high-$z$ also leaves an imprint in present-day galaxy systems of lower halo masses, not just in rich clusters.

6 SUMMARY AND DISCUSSION

In this paper, we use the observed SMFs of galaxies in the redshift range from $z \approx 0$ to $z \approx 4$ and the LF of cluster galaxies at $z \approx 0$ to constrain the star formation histories of galaxies hosted by dark matter haloes of different masses. To this end, we parametrize the SFR as a function of halo mass and redshift using piecewise power laws. We combine this empirical model for star formation with halo merger trees to follow the evolution of the stellar masses of galaxies and to make model predictions to be compared with the data constraints. We use the MULTINEST method developed by Feroz et al. (2009) to obtain the posterior distribution of the model parameters and the marginal likelihood. A series of nested model families with increasing complexity are explored to understand how the model parameters are constrained by the different observational data sets. We use the posterior model parameter distributions to make model predictions for a number of properties of the galaxy population and compare these results with available observations. Our main results can be summarized as follows.

(i) To match the observed SMFs at different redshifts, the SFR in central galaxies residing in haloes with masses above $10^{12} h^{-1} M_\odot$ has to be boosted at high redshift relative to the increase that arises naturally from the fact that the dynamical time-scale is shorter at higher $z$.

(ii) To reproduce the faint end of the cluster and field galaxy LF ($M_\star - 5 \log_10(h) > -18$) simultaneously, we require a characteristic redshift, $z_c \approx 2$ above which the SFR in low-mass haloes with masses $< 10^{11} h^{-1} M_\odot$ must be enhanced relative to that at lower $z$. This enhancement is also supported by the fact that isolated dwarf galaxies seem to be dominated by old stellar populations (Weisz et al. 2011), and by the observed SFR functions at high redshift (Smit et al. 2012).

Our model (Model III) that successfully matches all these observations makes the following predictions.

(i) The star formation efficiency, that is the SFR divided by the baryonic mass accretion rate of the host halo, peaks in haloes with masses between $3 \times 10^{11} h^{-1} M_\odot$ and $10^{12} h^{-1} M_\odot$. In lower mass haloes, the star formation efficiency is about 1/10 at $z > z_c$, and is strongly quenched at lower $z$ and roughly scales as $M_{\text{vir}}^{\frac{1}{3}}$. While in higher mass haloes, the star formation tends to be quenched and the quenching is stronger with decreasing redshift.

(ii) The average star formation histories for the central galaxies of haloes with masses $M_{\text{vir}} > 10^{12} h^{-1} M_\odot$ are peaked, with the peak redshift shifting from $z \approx 4$ for present-day cluster haloes with $M_{\text{vir}} \sim 10^{15} h^{-1} M_\odot$ to $z \approx 1$ for present-day Milky Way mass haloes ($M_{\text{vir}} \sim 10^{12} h^{-1} M_\odot$).

(iii) The average star formation history in low-mass haloes with $M_{\text{vir}} < 10^{12} h^{-1} M_\odot$ is "bimodal": it starts with a high value at $z > z_c$, declines with time to a minimum at $z = z_c$, and then increases before it reaches an extended period of roughly constant SFR. Our model, therefore, predicts the existence of an old stellar population formed at $z > z_c$ in dwarf galaxies, consistent with the results obtained from direct observations of the stellar populations in such galaxies (Weisz et al. 2011).

(iv) Central galaxies of massive clusters formed most of their stars as early as $z \approx 4$, but on average about half of their final mass is assembled at $z < 2$. On the other hand, Milky Way mass galaxies experienced dramatic increases in their stellar masses after $z \approx 2$, and less than 10 per cent ($\sim 10^7 M_\odot$) of their present-day stellar mass was assembled by $z = 2$.

(v) The stellar mass to halo mass ratio, $M_\star/M_{\text{vir}}$, for central galaxies peaks at a halo mass of $\approx 10^{12} h^{-1} M_\odot$ with a value of $\approx 1/30$, quite independent of redshift. This is in excellent agreement with a wide range of observational constraints (see e.g. Behroozi et al. 2010; Dutton et al. 2010, and references therein).

(vi) For haloes with masses below $2 \times 10^{11} h^{-1} M_\odot$ our model predicts that $M_\star/M_{\text{vir}} \approx 1/100$ at $z = 4$ quite independent of halo mass, but this ratio decreases rapidly with decreasing halo mass at $z \approx 0$.

(vii) The low-mass end slopes of the SMF and the SFR function steepen towards high redshift. Central galaxies dominate the present-day SMF at $M_\star > 10^{9} h^{-1} M_\odot$ but satellite galaxies begin to dominate at $M_\star < 10^{9} h^{-1} M_\odot$.

(viii) Haloes with $M_{\text{vir}} \sim 10^{12} h^{-1} M_\odot$, hosting centrals with $M_\star \sim 10^{10} h^{-1} M_\odot$, dominate the SFRD of the Universe at $z < 3$ while at higher $z$ star formation in lower mass haloes takes over. Star formation in haloes more massive than $10^{12.5} h^{-1} M_\odot$ never significantly contribute to the total SFRD.

Our findings have important implications for the physical processes that regulate star formation and feedback. In general, the SFR in a halo depends on the amounts of cold gas that can be accreted into the halo centre, and on the time-scale with which the cold gas converts into stars. In current theories of galaxy formation, the amount of cold gas in a halo is determined by radiative cooling and feedback effects.

It is well known that radiative cooling introduces a characteristic halo mass, $M_{\text{cool}} \approx 6 \times 10^{11} M_\odot$, which separates cooling limited ‘hot mode’ and free-fall limited ‘cold mode’ in the accretion of cold gas into galaxies (Birnboim & Dekel 2003; Keres et al. 2005, 2009). For haloes below this characteristic mass, gas is never heated during accretion and so the amount of cold gas is limited by the free-fall time of the gas. For haloes with larger masses, on the other hand, the accreted gas first heats by accretion shocks and then cools radiatively before it sinks into the central galaxy, so that cold gas accretion by the central galaxy is limited by the radiative cooling time-scale. The characteristic mass scales we find in the star formation efficiency shown in Figs 14 and 15 are very similar to $M_{\text{cool}}$, suggesting that radiative cooling plays an important role in star formation. Furthermore, since the cooling time-scale decreases faster than the free-fall time-scale as redshift increases (see section 8.4 in Mo et al. 2010), cooling may also have played a role in the enhanced SFR in massive haloes at high-$z$ (see Section 4.2).

However, radiative cooling alone cannot explain why the star formation efficiency in lower mass haloes is suppressed. Even for massive haloes, numerical simulations have shown that the suppression in radiative cooling at low-$z$ is not sufficient to explain the observed low SFRs, and some heating sources are needed to quench the star formation in massive galaxies at low-$z$. One possible mechanism is AGN feedback. Observations show that AGN activity peaks at $z \approx 2$ and declines towards both higher and lower redshift (e.g. Hopkins et al. 2007), indicating that supermassive black holes may have already formed in massive galaxies by $z \approx 2$. 

Empirical model of star formation history 1309
Thus, the quenching of star formation in massive galaxies at $z < 2$ may owe to a combination of effective AGN feedback (e.g. in low-accretion radio mode) and inefficient radiative cooling owing to the reduced gas density. Similarly, the high SFR in high-mass haloes at high-$z$ may arise from an increased radiative cooling efficiency combined with reduced AGN feedback owing to the reduced number of supermassive black holes that have formed or the presence of cold, filamentary accretion in these massive haloes at high redshift (Keres et al. 2009). Our results for the star formation in massive galaxies are in quantitative agreement with these expectations, and a detailed comparison between these empirical results and theoretical predictions will provide important insights into the underlying physical processes.

Our results for the star formation in low-mass haloes poses a number of challenges to standard theory. First, since radiative cooling is expected to be effective at all redshifts in low-mass haloes, some feedback processes must be invoked to suppress the star formation efficiency in these haloes. A popular assumption is that galactic winds driven by supernova explosions may suppress the star formation in such haloes. In many models considered thus far, the mass loading factor, which is defined to be the mass-loss rate through winds divided by the SFR, is assumed to be some power of the circular velocity of the host halo. In contrast, our results indicate that the star formation efficiency in dwarf haloes at $z \approx 4$ is about 10 per cent of the baryonic accretion rate independent of halo mass. This suggests that the mass loading of the wind is independent of the halo mass at high redshift, so that a constant fraction of the accreted gas is converted into stars and the rest is driven out of the halo as galactic winds.

Furthermore, the existence of a transition at $z = z_\star \approx 2$ separating active from quiescent star-forming phases, which is required to explain the faint-end upturn in the CGLF, is not expected in the conventional supernova feedback model. Instead, our results lend support to a scenario in which the IGM is pre-heated at $z = z_\star$ and hence the accretion of baryons into low-mass haloes is delayed until they become sufficiently massive to allow significant accretion from the pre-heated IGM to form stars at lower redshift. The exact mechanism for pre-heating is still unclear. Possibilities that have been proposed include the formation of pancakes (Mo et al. 2005), an episode of starburst and AGN activity at $z \gtrsim 2$ (Mo & Mao 2002), and heating by high-energy gamma rays generated by blazars (Chang et al. 2011). In all these pre-heating scenarios, pre-heating is expected to be at $z \approx 2$, in excellent agreement with the value of $z_\star$ that we find. The pre-heated entropy of the IGM is about a few times $10^6$ cm$^2$, similar to what is required to match the observed LF and H I mass functions of present-day galaxies (Lu et al. 2013).

The finding that Milky Way mass galaxies experienced a significant increase in their stellar masses after $z \approx 2$ contrasts with the fact that the dark haloes of these galaxies assembled their masses at a much slower pace at $z < 2$ (Zhao et al. 2009). More specifically, the progenitors of present-day Milky Way mass galaxies at $z \approx 2$ on average have a stellar mass that is about 1/15 of the final stellar mass, while the average halo mass at the same redshift is about 1/4 of the final halo mass. Apparently, the star formation in such galaxies is delayed from and related delayed to the halo accretion. This contrasts with numerical simulations that predict that the SFR traces the accretion rate (e.g. Dave, Finlator & Oppenheimer 2012) at late times when the circular velocity of Milky Way mass haloes changes slowly with time (Zhao et al. 2003). Apparently, some process must have delayed the cold gas accretion relative to the halo mass accretion in such haloes. As discussed above, pre-heating may operate in this way. Alternatively, a large fraction of the accreted cold gas may be ejected and take a relatively long time to re-accrete on to the galaxy to feed star formation (Oppenheimer et al. 2010). Clearly, one needs a detailed analysis to see if such processes can produce the star formation and assembly histories of Milky Way type galaxies that we obtain here.

ACKNOWLEDGEMENTS

The authors acknowledge P. Popesso for providing the data of CGLFs and F. Feroz for providing the source code of MULTINEST. This work is supported by NSF AST-1109354, NSF AST-0908334 and NASA NNX10A956.

REFERENCES

Baldry I. K. et al., 2012, MNRAS, 421, 621
Balogh M. L., Navarro J. F., Morris S. L., 2000, ApJ, 540, 113
Banados E., Hung L. W., de Propris R., West M. J., 2010, ApJ, 721, 14
Barkhouse W. A., Yee H. K. C., Lopez-Cruz O., 2007, ApJ, 671, 1471
Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379
Behroozi P. S., Wechsler R. H., Conroy C., 2012, ApJ, 770, 57
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 762, 31
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Béthermin M., Doré O., Lagache G., 2012, A&A, 537, L5
Béthermin M., Wang L., Doré O., Lagache G., Sargent M., Daddi E., Cousin M., Aussel H.-C., 2013, A&A, 557, A66
Birnboim Y., Dekel A., 2003, MNRAS, 345, 349
Blanton M. R., Lupton R. H., Schlegel D. J., Strauss M. A., Brinkmann J., Fukugita M., Loveday J., 2005, ApJ, 631, 208
Bouche N. et al., 2010, ApJ, 718, 1001
Bouwens R. J. et al., 2012, ApJ, 754, 83
Boylan-Kolchin M., Ma C. P., Quataert E., 2008, MNRAS, 383, 93
Bradley L. D. et al., 2012, ApJ, 760, 108
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Cacciato M., van den Bosch F. C., More S., Mo H. J., Yang X., 2013, MNRAS, 430, 767
Chabrier G., 2003, PASP, 115, 763
Chang P., Broderick A., Pfrommer C., 2011, ApJ, 752, 23
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Cole S., Helly J. C., Frenk C. S., Parkinson H., 2008, MNRAS, 383, 546
Conroy C., Wechsler R. H., 2009, ApJ, 696, 620
Conroy C., Wechsler R. H., Kravtsov A. V., 2006, ApJ, 647, 201
Conroy C., Wechsler R. H., Kravtsov A. V., 2007, ApJ, 668, 826
Croton D. J. et al., 2006, MNRAS, 365, 11
Daddi E. et al., 2007, ApJ, 670, 156
Dave R., Finlator K., Oppenheimer B. D., 2012, MNRAS, 421, 98
Drony N. et al., 2009, ApJ, 707, 1595
Dunkley J. et al., 2009, ApJS, 180, 306D
Dutton A. A., Conroy C., van den Bosch F. C., Prada F., More S., 2010, MNRAS, 407, 2
Feroz F., Hobson M. P., Bridges M., 2009, MNRAS, 398, 1601
Gallazzi A., Charlot S., Brinchmann J., White S. D. M., Tremonti C. A., 2005, MNRAS, 362, 41
Genel S. et al., 2008, ApJ, 688, 789
Guo Q., White S., Li C., Boylan-Kolchin M., 2010, MNRAS, 404, 1111
Guo Q. et al., 2011, MNRAS, 413, 101
Hopkins P. F., Beacom J. F., 2006, ApJ, 651, 142
Hopkins P. F., Hernquist L., Cox T. J., Robertson B., Krause E., 2007, ApJ, 669, 67
Jenkins L. P., Hornschemeier A., Mobasher B., Alexander D. M., Bauer F. E., 2007, ApJ, 666, 846
Jiang Y., van den Bosch F. C., 2013, preprint (arXiv:1311.5225)
Jing Y. P., Mo H. J., Borner G., 1998, ApJ, 494, 1
Kang X., Jing Y. P., Mo H. J., Borner G., 2005, ApJ, 631, 21
Kaufmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201
Keres D., Katz N., Weinberg D. H., Dave R., 2005, MNRAS, 363, 2

1310 Z. Lu et al.

Copyright © 2014 The Authors, some rights reserved; exclusive licensee OUP, all rights reserved. For Permissions, please email: journals.permissions@oup.com
APPENDIX A: THE LIKELIHOOD FUNCTION

As discussed in the main text, the likelihood function describes the probability of the data given the model and its parameters. For the present problem it is impossible to get a rigorous likelihood function. If, for simplicity, the sampling uncertainty is assumed to be a Poisson process, then the total variance from observations can be written as

\[ \sigma^2_{\text{sys}} + \frac{(2\pi)^{\nu/2}}{\nu} \, \nu \, n_{\text{sys}} \]

which can be approximated by

\[ \sigma^2_{\text{mod}} = \sigma^2_{\text{sys}} + \frac{(2\pi)^{\nu/2}}{\nu} \, \nu \, n_{\text{sys}} \]

In this case, the likelihood function can be approximated by \( \sigma^2_{\text{mod}} = \sigma^2_{\text{sys}} + \frac{(2\pi)^{\nu/2}}{\nu} \, \nu \, n_{\text{sys}} \), where \( n_{\text{sys}} \) is the observed number counts. Replacing the Poisson process in the data by that in the model, it can be shown that the variance in the likelihood function can be approximated by \( n_{\text{obs}} \) is the number counts predicted by the model. The likelihood for each stellar mass bin is then

\[ \ln(L) = -\frac{1}{2} \left( \frac{\Phi_{\text{obs}} - \Phi_{\text{mod}}}{\sigma_{\text{mod}}} \right)^2 \]

The second term in the variance, \( \frac{(2\pi)^{\nu/2}}{\nu} \, \nu \, n_{\text{sys}} \), can be evaluated if \( n_{\text{obs}} \) is known. In theory, this term makes the likelihood deviate from a normal distribution, especially in the tails where \( \Phi_{\text{mod}} \) is far from \( \Phi_{\text{obs}} \). We perform a series of tests and study how this term affects parameter estimation and the computed value of the marginalized likelihood. We find that for the problems we study here the marginalized likelihood is hardly affected because the posteriors for our models are always dominated by the likelihood regions where \( \Phi_{\text{mod}} \) is close to \( \Phi_{\text{obs}} \). In this case, the likelihood function reduces to the form given in equation (12).

APPENDIX B: MULTINEST SAMPLING OF THE POSTERIOR DISTRIBUTION

The main goal of nested sampling is to evaluate the Bayesian evidence, and to provide samples of the posterior distribution. Briefly, the algorithm works as follows. At the beginning of the process, one randomly draws \( N \) points in parameter space from the adopted...
prior distribution. These points are called the active set. Each of the points has a likelihood value \( L_i \) \((i = 0, \ldots, N - 1)\), and associated with it is an iso-likelihood surface defined by the value of \( L_i \). The volume (modulated by the prior distribution) within the surface is \( X_i \). The point with the lowest likelihood value is denoted by \( L^{(0)} \) and the corresponding prior volume, \( X^{(0)} \), can be approximated by the total volume of the prior space. This point is removed from the active set and is added to another list called the inactive set. A new point with likelihood bigger than \( L^{(0)} \) is then drawn from the prior distribution and is added to the active set. Before going to the next iteration, it is important to know the volume (again modulated by the prior distribution) occupied by the new active set. Note that the ratios, \( t_i \equiv X_i/X^{(0)} \) \((i = 0, \ldots, N - 1)\), can be considered as an iso-likelihood surface defined by the value of \( L_i \) and the standard deviation in \( \ln (t) \) equal to \(-1/N\) and the standard deviation in \( \ln (t) \) equal to \(1/N\). The first step, therefore, ends with a new active set that occupies a total volume that is smaller by a factor of \( t^{(1)} \approx \exp (-1/N) \) than the old set, and with a new member, \( (L^{(0)}, X^{(0)}) \), in the inactive set.

By repeating the above process for the new active set produced at each subsequent step, a list of points are drawn from the posterior distribution:

\[
\{ (L^{(0)}, X^{(k)}) \} \quad \text{with} \quad X^{(k+1)} = t^{(k+1)} X^{(k)},
\]

which defines a series of nested shells in the parameter space and can be used to sample the posterior distribution. As described above, the exact value of \( t^{(k+1)} \) is unknown but can be approximated by \( \exp (-1/N) \) and the uncertainty can also be quantified with the use of equation (B1). The Bayesian evidence is simply

\[
Z = \sum_k L^{(0)} \frac{[X^{(k+1)} - X^{(k-1)}]}{2}.
\]

One iterates until the value of \( Z \) reaches a chosen accuracy, and we choose it to be 0.5 in natural logarithmic scale.

The efficiency of this algorithm depends on how efficiently the active set can be replenished at each iteration. Drawing new samples blindly from the prior leads to a lower and lower acceptance rate as the iso-likelihood surface shrinks. However, if the surface can be approximated by some regular shapes, then the active set can be efficiently replenished. The MULTINEST package developed by Feroz et al. (2009) provides such a method. At each iteration, multiple ellipsoids are used to approximate the iso-likelihood surface of the new point drawn. An ellipsoid can either overlap with others or be isolated. Too few big ellipsoids would result in a bad approximation, while too many small ellipsoids would result in too much overlap. In both cases, the acceptance rate would be low. Optimal ellipsoidal decomposition is found by minimizing

\[
F = \frac{\sum_i V(E_i)}{V(S)} \geq 1,
\]

where \( V(S) \) is the volume within the iso-likelihood surface and \( \sum_i V(E_i) \) is the total volume of all the ellipsoids. The acceptance rate is simply the inverse of \( F \).

The parameter that controls the process of posterior exploration is the size of the active set. A large active set can slow down the speed of going uphill because after each iteration the size of the volume enclosed by the iso-likelihood surface shrinks by \( \exp (-1/N) \). Conversely, to detect all the modes that are statistically significant in a high dimensional parameter space the size of the active set cannot be too small. Unfortunately, there is no good way to find the optimal active set size. For this work, we use 2000 active points for Model I, Model II and Model IIb, 5000 for Model III and 10 000 for Model IV. We arrived at these values empirically by increasing the active set size until the value of the marginal likelihood did not change appreciably.

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