

Galaxy formation spanning cosmic history

Andrew J. Benson^{1★} and Richard Bower²

¹Mail Code 350-17, California Institute of Technology, Pasadena, CA 91125, USA

²Institute for Computational Cosmology, University of Durham, Durham

Accepted 2010 February 24. Received 2010 January 25; in original form 2009 October 19

ABSTRACT

Over the past several decades, galaxy formation theory has met with significant successes. In order to test current theories thoroughly we require predictions for as yet unprobed regimes. To this end, we describe a new implementation of the GALFORM semi-analytic model of galaxy formation. Our motivation is the success of the model described by Bower et al. in explaining many aspects of galaxy formation. Despite this success, the Bower et al. model fails to match some observational constraints, and certain aspects of its physical implementation are not as realistic as we would like. The model described in this work includes substantially updated physics, taking into account developments in our understanding over the past decade, and removes certain limiting assumptions made by these (and most other) semi-analytic models. This allows it to be exploited reliably in high-redshift and low-mass regimes. Furthermore, we have performed an exhaustive search of model parameter space to find a particular set of model parameters which produce results in good agreement with a wide range of observational data (luminosity functions, galaxy sizes and dynamics, clustering, colours, metal content) over a wide range of redshifts. This model represents a solid basis on which to perform calculations of galaxy formation in as yet unprobed regimes.

Key words: galaxies: evolution – galaxies: formation – galaxies: general – galaxies: high-redshift – intergalactic medium.

1 INTRODUCTION

Understanding the physics of galaxy formation has been an active field of study ever since it was demonstrated that galaxies are stellar systems external to our own Milky Way. Modern galaxy formation theory grew out of early studies of cosmology and structure formation and is set within the cold dark matter cosmological model and therefore proceeds via a fundamentally hierarchical paradigm. Observational evidence and theoretical expectations indicate that galaxy formation is an ongoing process which has been occurring over the vast majority of the Universe's history. The goal of galaxy formation theory then is to describe how underlying physical principles give rise to the complicated set of phenomena which galaxies encompass.

Approaches to modelling the complex and non-linear processes of galaxy formation fall into two broad categories: direct hydrodynamical simulation and semi-analytic modelling. The division is of a somewhat fuzzy nature: semi-analytic models frequently make use of N -body simulation merger trees and calibrations from simulations, while simulations themselves are forced to include semi-analytical prescriptions for sub-resolution physics. The direct

simulation approach has the advantage of, in principle, providing precise solutions (in the limit of large number of particles and assuming that numerical artefacts are kept under control), but require substantial investments of computing resources and are, at present (and for the foreseeable future), more fundamentally limited by our incomplete understanding of the various sub-resolution physical processes incorporated into them. The semi-analytical approach is less precise, but allows for rapid exploration of a wide range of galaxy properties for large, statistically useful samples. A primary goal of the semi-analytic approach is to develop insights into the process of galaxy formation that are comprehensible in terms of fundamental physical processes or emergent phenomena.¹

The problem is therefore one of complexity: can we extract the underlying mechanisms that drive different aspects of galaxy formation and evolution from the numerous and complicated physical mechanisms at work. The key here is then 'understanding'. One can easily comprehend how a $1/r^2$ force works and can, by

¹A good example of an emergent phenomenon here is dynamical friction. Gravity (in the non-relativistic limit) is described entirely by $1/r^2$ forces and at this level makes no mention of frictional effects. The phenomenon of dynamical friction emerges from the interaction of large numbers of gravitating particles.

*E-mail: abenson@its.caltech.edu

extrapolation, understand how this force applies to the billions of particles of dark matter in an N -body simulation. However, it is not directly obvious (at least not to these authors) how a $1/r^2$ force leads to the formation of complex filamentary structures and collapsed virialized objects. Instead, we have developed simplified analytic models (e.g. the Zel'dovich approximation, spherical top-hat collapse models, etc.) which explain these phenomena in terms more accessible to the human intellect. It seems that this is what we must strive for in galaxy formation theory – a set of analytic models that we can comprehend and which allow us to understand the physics and a complementary set of precision numerical tools to allow us to determine the quantitative outcomes of that physics (in order to make precision tests of our understanding). As such, it is our opinion that no set of numerical simulations of galaxy formation, no matter how precise, will directly result in understanding. Instead, analytic methods, perhaps of an approximate nature, must always be developed (and, of course, checked against those numerical simulations) to allow us to understand galaxy formation.

Modern semi-analytic models of galaxy formation began with White & Frenk (1991), drawing on earlier work by Rees & Ostriker (1977) and White & Rees (1978). Since then, numerous studies (Kauffmann, White & Guiderdoni 1993; Cole et al. 1994; Baugh et al. 1998, 1999b; Somerville & Primack 1999; Cole et al. 2000; Benson et al. 2002a; Hatton et al. 2003; Monaco et al. 2007) have extended and improved this original framework. Current semi-analytic models have been used to investigate many aspects of galaxy formation including the following.

- (i) Galaxy counts (Kauffmann, Guiderdoni & White 1994; Devriendt & Guiderdoni 2000)
- (ii) Galaxy clustering (Diaferio et al. 1999; Kauffmann et al. 1999a,b; Baugh et al. 1999a; Benson et al. 2000a,b; Wechsler et al. 2001; Blaizot et al. 2006)
- (iii) Galaxy colours and metallicities (Kauffmann & Charlot 1998; Springel et al. 2001; Lanzoni et al. 2005; Font et al. 2008; Nagashima et al. 2005b)
- (iv) Sub-mm and infrared (IR) galaxies (Guiderdoni et al. 1998; Granato et al. 2000; Baugh et al. 2005; Lacey et al. 2008)
- (v) Abundance and properties of Local Group galaxies (Benson et al. 2002b; Somerville 2002)
- (vi) The reionization of the Universe (Devriendt et al. 1998; Benson et al. 2001; Somerville & Livio 2003; Benson et al. 2006)
- (vii) The heating of galactic discs (Benson et al. 2004)
- (viii) The properties of Lyman-break galaxies (Governato et al. 1998; Blaizot et al. 2003, 2004)
- (ix) Supermassive black hole formation and active galactic nuclei (AGN) feedback (Kauffmann & Haehnelt 2000; Croton et al. 2006; Bower et al. 2006; Malbon et al. 2007; Somerville et al. 2008b; Fontanot et al. 2009a)
- (x) Damped Lyman α systems (Maller et al. 2001, 2003)
- (xi) The X-ray properties of galaxy clusters (Bower et al. 2001; Bower, McCarthy & Benson 2008)
- (xii) Chemical enrichment of the intracluster medium (ICM) and intergalactic medium (IGM) (De Lucia, Kauffmann & White 2004; Nagashima et al. 2005a)
- (xiii) The formation histories and morphological evolution of galaxies (Kauffmann 1996; De Lucia et al. 2006; Fontanot et al. 2007; Somerville et al. 2008a).

The goal of this approach is to provide a coherent framework within which the complex process of galaxy formation can be studied. Recognizing that our understanding of galaxy formation is far from complete these models should not be thought of as attempting to pro-

vide a ‘final theory’ of galaxy formation (although that, of course, remains the ultimate goal), but instead to provide a means by which new ideas and insights may be tested and by which quantitative and observationally comparable predictions may be extracted in order to test current theories.

In order for these goals to be met we must endeavour to improve the accuracy and precision of such models and to include all of the physics thought to be relevant to galaxy formation. The complementary approach of direct numerical (N -body and/or hydrodynamic) simulation has the advantage that it provides high precision, but is significantly limited by computing power, resulting in the need for inclusion of semi-analytic recipes in such simulations. In any case, while a simulation of the entire Universe with infinite resolution would be impressive, the goal of the physicist is to understand nature through relatively simple arguments.²

The most recent incarnation of the GALFORM model was described by Bower et al. (2006). The major innovation of that work was the inclusion of feedback from AGN which allowed it to produce a very good match to the observed local luminosity functions of galaxies. In particular, the Bower et al. (2006) model was designed to explain the phenomenon of ‘down sizing’. While the Bower et al. (2006) model turned out to also give a good match to several other data sets – including stellar mass functions at higher redshifts, the luminosity function at $z = 3$ (Marchesini & van Dokkum 2007), the abundance of $5 < z < 6$ galaxies (McLure et al. 2009), overall colour bimodality (Bower et al. 2006), morphology (Parry, Eke & Frenk 2009), the global star formation rate and the black hole mass versus bulge mass relation (Bower et al. 2006) – it fails in several other areas, such as the mass–metallicity relation for galaxies, the sizes of galactic discs (González et al. 2009), the small-scale clustering amplitude (Kim et al. 2009), the normalization and environmental dependence of galaxy colours (Font et al. 2008) and the X-ray properties of groups and clusters (Bower et al., in preparation). Additionally, while the implementation of physics in semi-analytic models must always involve approximations, there are several aspects of the Bower et al. (2006) model which call out for improvement and updating. Chief amongst these is the cooling model – crucial to the implementation of AGN feedback – which retained assumptions about dark matter halo ‘formation’ events which make implementing feedback physics difficult. Our motivation for this work is therefore to attempt to rectify these shortcomings of the Bower et al. (2006) model by updating the physics of GALFORM, removing unnecessary assumptions and approximations, and adding in new physics that is thought to be important for galaxy formation but which has previously been neglected in GALFORM. In addition, we will systematically explore the available model parameter space to locate a model which best agrees with a wide range of observational constraints.

In this current work, we describe the advances made in the GALFORM semi-analytic model over the past 9 years. Our goal is to present a comprehensive model for galaxy formation that best agrees with current experimental constraints. In future papers we will utilize this model to explore and explain features of the galaxy population through cosmic history.

²For example, while it is clear from N -body simulations that the action of $1/r^2$ gravitational forces in a cold dark matter (CDM) universe leads to dark matter haloes with approximately Navarro–Frenk–White (NFW) density profiles, there is a clear drive to provide simple, analytic models to demonstrate that we understand the underlying physics of these profiles (Taylor & Navarro 2001; Barnes et al. 2007a,b).

The remainder of this paper is structured as follows. In Section 2, we describe the details of our revised GALFORM model. In Section 3, we describe how we select a suitable set of model parameters. In Section 4, we present some basic results from our model, while in Section 5 we explore the effects of certain physical processes on the properties of model galaxies. Finally, in Section 6 we discuss their implications and in Section 7 we give our conclusions. Readers less interested in the technicalities of semi-analytic models and how they are constrained may wish to skip Sections 2 and 3 and most of Section 4, and jump directly to Section 4.12 where we present two interesting predictions from our model and Section 5 in which we explore the effects of varying key physical processes.

2 MODEL

In this section, we provide a detailed description of our model.

2.1 Starting point

The starting point for this discussion is Cole et al. (2000) and we will refer to that work for details which have not changed in the current implementation. We choose Cole et al. (2000) as a starting point for the technical description of our model as it represents the last point at which the details of the GALFORM model were presented as a coherent whole in a single document. As noted in Section 1, however, the scientific predecessor of this work is Bower et al. (2006). That paper, and several others, introduced many improvements relative to Cole et al. (2000), many of which are described in more detail here. A brief chronology of the development of GALFORM from Cole et al. (2000) to the present is as follows.

- (i) Cole et al. (2000): previous full description of the GALFORM model.
- (ii) Granato et al. (2000): detailed dust modelling utilizing GRASIL (see Section 2.14.1).
- (iii) Benson et al. (2001): treatment of reionization and the evolution of the IGM (see Section 2.10).
- (iv) Bower et al. (2001): treatment of heating and ejection of hot material from haloes due to energy input (see Section 2.13).
- (v) Benson et al. (2002b): back reaction of reionization and photoionizing background on galaxy formation (see Section 2.10) and detailed treatment of satellite galaxy dynamics (a somewhat different approach to this is described in Sections 2.8 and 2.9).
- (vi) Benson et al. (2003): effects of thermal conduction on cluster cooling rates and ‘superwind’ feedback from supernovae (SNe; described in further detail by Baugh et al. 2005).
- (vii) Benson et al. (2004): heating of galactic discs by orbiting dark matter haloes.
- (viii) Nagashima et al. (2005a): detailed chemical enrichment models (incorporating delays and tracking of individual elements; see Section 2.11).
- (ix) Bower et al. (2006): feedback from AGN (see Section 2.13).
- (x) Malbon et al. (2007): black hole growth (see Section 2.13) as applied to Baugh et al. (2005) – see Fanidakis et al. (in preparation) for a similar (and more advanced) treatment of black holes in the Bower et al. (2006) model.
- (xi) Stringer & Benson (2007): radially resolved structure of galactic discs.
- (xii) Font et al. (2008): ram-pressure stripping of cold gas from galactic discs (see Section 2.9).

2.2 Executive summary

Having developed these treatments of various physical processes one by one, our intention is to integrate them into a single baseline model. In addition to the accumulation of many of these improvements (many of which have not previously been utilized simultaneously), the two major modifications to the GALFORM model introduced in this work are as follows.

- (i) The removal of discrete ‘formation’ events for dark matter haloes (which previously occurred each time a halo doubled in mass and caused calculations of cooling and merging times to be reset). This has facilitated a major change in the GALFORM cooling model which previously made fundamental reference to these formation events.
- (ii) The inclusion of arbitrarily deep levels of subhaloes within subhaloes and, as a consequence, the possibility of mergers between satellite galaxies.

Aspects of the model that are essentially unchanged from Cole et al. (2000) are listed in Section 2.3. Before launching into the detailed discussion of the model, Section 2.4 provides a quick overview of what has changed between Cole et al. (2000) and the current implementation. In addition to changes to the physics of the model, the GALFORM code has been extensively optimized and made OpenMP parallel to permit rapid calculation of self-consistent galaxy/IGM evolution (see Section 2.10).

2.3 Unchanged aspects

Below we list aspects of the current implementation of GALFORM that are unchanged relative to that published in Cole et al. (2000).

- (i) *Virial overdensities*: virial overdensities of dark matter haloes are computed as described by Cole et al. (2000), i.e. using the spherical top-hat collapse model for the appropriate cosmology and redshift. Given the mass and virial overdensity of each halo the corresponding virial radii and velocities are easily computed.
- (ii) *Star formation rate*: the star formation rate in disc galaxies is given by

$$\dot{\phi} = M_{\text{cold}}/\tau_{\star} \text{ where } \tau_{\star} = \epsilon_{\star}^{-1} \tau_{\text{disc}} (V_{\text{disc}}/200 \text{ km s}^{-1})^{\alpha_{\star}}, \quad (1)$$

where M_{cold} is the mass of cold gas in the disc, $\tau_{\text{disc}} = r_{\text{disc}}/V_{\text{disc}}$ is the dynamical time of the disc at the half-mass radius r_{disc} and V_{disc} is the circular velocity of the disc at that radius. The two parameters ϵ_{\star} and α_{\star} control the normalization of the star formation rate and its scaling with galaxy circular velocity, respectively.

(iii) *Mergers/morphological transformation*: the classification of merger events as minor or major follows the logic of Cole et al. (2000; section 4.3.2). However, the rules which determine when a burst of star formation occurs are altered to become:

(a) Major merger?

$$(1) \quad \text{Requires } M_{\text{sat}}/M_{\text{cen}} > f_{\text{burst}}.$$

(b) Minor merger?

$$(2) \quad \text{Requires } \begin{cases} M_{\text{cen(bulge)}}/M_{\text{cen}} < B/T_{\text{burst}} \\ \text{and} \\ M_{\text{cen(cold)}}/M_{\text{cen}} \geq f_{\text{gas,burst}}. \end{cases}$$

where M_{cen} and M_{sat} are the baryonic masses of the central and satellite galaxies involved in the merger, respectively; $M_{\text{cen(bulge)}}$ is the mass of the bulge component in the central galaxy and $f_{\text{burst}}, f_{\text{gas,burst}}$

and B/T_{burst} are parameters of the model. The parameter B/T_{burst} is intended to inhibit minor merger-triggered bursts in systems that are primarily spheroid dominated (since we may expect that in such systems the minor merger cannot trigger the same instabilities as it would in a disc-dominated system and therefore be unable to drive inflows of gas to the central regions to fuel a burst). We would expect that the value of this parameter should be of the order of unity (i.e. the system should be spheroid dominated in order that the burst triggering be inhibited).

(iv) *Spheroid sizes*: the sizes of spheroids formed through mergers are computed using the approach described by Cole et al. (2000; section 4.4.2).

(v) *Calculation of luminosities*: the luminosities and magnitudes of galaxy are computed from their known stellar populations as described by Cole et al. (2000; section 5.1). (However note that the treatment of dust extinction has changed; see Section 2.14.1.)

2.4 Overview of changes

We list below the changes in the current implementation of GALFORM relative to that published in Cole et al. (2000). These are divided into ‘minor changes’, which are typically simple updates of fitting formulae, and ‘major changes’, which are significant additions to or modifications of the physics and structure of the model.

2.4.1 Minor changes

(i) *Dark matter halo mass function* (see Section 2.5.1): Cole et al. (2000) use the Press & Schechter (1974) mass function for dark matter haloes. In this work, we use the more recent determination of Reed et al. (2007) which is calibrated against N -body simulations over a wide range of masses and redshifts.

(ii) *Dark matter merger trees* (see Section 2.5.2): Cole et al. (2000) use a binary split algorithm utilizing halo merger rates inferred from the extended Press–Schechter formalism (Lacey & Cole 1993). We use an empirical modification of this algorithm proposed by Parkinson, Cole & Helly (2008), which provides a much more accurate match to progenitor halo mass functions as measured in N -body simulations.

(iii) *Density profile of dark matter haloes* (see Section 2.5.4): Cole et al. (2000) employed NFW (Navarro, Frenk & White 1997) density profiles. We instead use Einasto density profiles (Einasto 1965) consistent with recent findings (Navarro et al. 2004; Merritt et al. 2005; Prada et al. 2006).

(iv) *Density and angular momentum of halo gas* (see Section 2.6.3): Cole et al. (2000) adopted a cored isothermal profile for the hot gas in dark matter haloes and furthermore assumed a solid body rotation, normalizing the rotation speed to the total angular momentum of the gas (which was assumed to have the same average specific angular momentum as the dark matter). We choose to adopt the density and angular momentum distributions measured in hydrodynamical simulations by Sharma & Steinmetz (2005).

(v) *Dynamical friction time-scales* (see Section 2.8.5): Cole et al. (2000) estimated dynamical friction time-scales using the expression derived by Lacey & Cole (1993) for isothermal dark matter haloes and the distribution of orbital parameters found by Tormen (1997). In this work, we adopt the fitting formula of Jiang et al. (2008) to compute dynamical friction time-scales and the orbital parameter distribution of Benson (2005).

(vi) *Disc stability*: We retain the same test of disc stability as did Cole et al. (2000) and similarly assume that unstable discs undergo

bursts of star formation resulting in the formation of a spheroid.³ One slight difference is that we assume that the instability occurs at the largest radius for which the disc is deemed to be unstable rather than at the rotational support radius as Cole et al. (2000) assumed. This prevents galaxies with very low angular momenta from contracting to extremely small sizes (and thereby becoming very highly self-gravitating and unstable) before the stability criterion is tested. Additionally, we allow for different stability thresholds for gaseous and stellar discs. We employ the stability criterion of Efstathiou, Lake & Negroponte (1982) such that discs require

$$\frac{V_d}{(GM_d/R_s)^{1/2}} > \epsilon_d \quad (2)$$

to be stable, where V_d is the disc rotation speed at the half-mass radius, M_d is the disc mass and R_s is the disc radial scalelength. Efstathiou et al. (1982) found a value of $\epsilon_{d,*} = 1.1$ was applicable for purely stellar discs. Christodoulou, Shlosman & Tohline (1995) demonstrate that an equivalent result for gaseous discs gives $\epsilon_{d,\text{gas}} = 0.9$. We choose to make $\epsilon_{d,\text{gas}}$ a free parameter of the model and enforce $\epsilon_{d,*} = \epsilon_{d,\text{gas}} + 0.2$. For discs containing a mixture of stars and gas we linearly interpolate between $\epsilon_{d,*}$ and $\epsilon_{d,\text{gas}}$ using the gas fraction as the interpolating variable. As has been recently pointed out by Athanassoula (2008), this treatment of the process of disc destabilization, similar to that in other semi-analytic models, is dramatically oversimplified. As Athanassoula (2008) also describes, a more realistic model would need both a much more careful assessment of the disc stability and a consideration of the process of bar formation. This currently remains beyond the ability of our model to address, although it should clearly be a priority area in which semi-analytic models should strive to improve. In GALFORM we can consider an alternative disc instability treatment in which during an instability event only just enough mass is transferred from the disc to the spheroid component to re-stabilize the disc. While this does not explore the full range of uncertainties arising from the treatment of this process, it gives at least some idea of how significant they may be. We find that the net result of switching to the alternative treatment of instabilities is to slightly increase the number of bulgeless galaxies at all luminosities, with a corresponding decrease in the numbers of intermediate and pure spheroid galaxies. The changes, however, do not alter the qualitative trends of morphological mix with luminosity nor global properties of galaxies such as sizes and luminosity functions at $z = 0$. At higher redshifts (e.g. $z \geq 5$), the change is more significant, with a reduction in star formation rate by a factor of 2–3 resulting from the lowered frequency of bursts of star formation. This change could be offset by adjustments in other parameters, but demonstrates the need for a refined understanding and modelling of the disc instability process in semi-analytic models.

(vii) *Sizes of galaxies* (see Section 2.7): sizes of discs and spheroids are determined as described by Cole et al. (2000), although the equilibrium is solved for in the potential corresponding to an Einasto density profiles (used throughout this work) rather than the NFW profiles assumed by Cole et al. (2000) and adiabatic contraction is computed using the methods of Gnedin et al. (2004) rather than that of Blumenthal et al. (1986).

While we class the above as minor changes, the effects of some of these changes can be significant in the sense that reverting to the previous implementation would change some model predictions by

³While the implementation of this physical process is unchanged, Cole et al. (2000) actually ignored this process in their fiducial model, while we include it in our work.

an amount comparable to or greater than the uncertainties in the relevant observational data. However, none of these modifications leads to fundamental changes in the behaviour of the model and their effects could all be counteracted by small adjustments in model parameters. This is why we classify them as ‘minor’ and do not explore their consequences in any greater detail.

2.4.2 Major changes

(i) *Spins of dark matter halo* (see Section 2.5.5): in Cole et al. (2000) spins of dark matter haloes were assigned randomly by drawing from the distribution of Cole & Lacey (1996). In this work, we implement an updated version of the approach described by Vitvitska et al. (2002) to produce spins correlated with the merging history of the halo and consistent with the distribution measured by Bett et al. (2007).

(ii) *Removal of discrete formation events* (see Section 2.5.3): the discrete ‘formation’ events (associated with mass doublings) in merger trees which Cole et al. (2000) utilized to reset cooling and merging calculations are no longer utilized. Instead, cooling, merging and other processes related to the merger tree evolve smoothly as the tree grows.

(iii) *Cooling model* (see Section 2.6): the cooling model has been revised to remove the dependence on halo formation events, allow for gradual recooling of gas ejected by feedback and accounts for cooling due to molecular hydrogen and Compton cooling and for heating from a photon background.

(iv) *Ram-pressure and tidal stripping* (see Section 2.9): ram-pressure and tidal stripping of both hot halo gas and stars and interstellar medium (ISM) gas in galaxies are now accounted for.

(v) *IGM interaction* (see Section 2.10): galaxy formation is solved simultaneously with the evolution of the IGM in a self-consistent way: emission from galaxies and AGN ionize and heat the IGM which in turn suppresses the formation of future generations of galaxies.

(vi) *Full hierarchy of subhaloes* (see Section 2.8): all levels of the substructure hierarchy (i.e. subhaloes, sub-subhaloes, sub-sub-subhaloes . . .) are included in calculations of merging. This allows for satellite–satellite mergers.

(vii) *Non-instantaneous recycling* (see Section 2.11): the instantaneous recycling approximation for mass loss, chemical enrichment and feedback has been dropped and the full time and metallicity-dependencies included. All models presented in this work utilize fully non-instantaneous recycling, metal production and SNe feedback.

2.5 Dark matter haloes

2.5.1 Mass function

We assume that the masses of dark matter haloes at any given redshift are distributed according to the mass function found by Reed et al. (2007). Specifically, the mass function is given by

$$\begin{aligned} \frac{dn}{d \ln M_v} &= \sqrt{\frac{2}{\pi}} \frac{\Omega_0 \rho_{\text{crit}}}{M_v} \left| \frac{d \ln \sigma}{d \ln M} \right| \\ &\times [1 + 1.047(\omega^{-2p}) + 0.6G_1 + 0.4G_2] A' \omega \\ &\times \exp\left(-\frac{1}{2} \omega^2 - 0.0325 \frac{\omega^{2p}}{(n_{\text{eff}} + 3)^2}\right), \end{aligned} \quad (3)$$

where $dn/d \ln M_v$ is the number of haloes with virial mass M_v per unit volume per unit logarithmic interval in M_v , $\sigma(M)$ is the

fractional mass root-variance in the linear density field in top-hat spheres containing, on average, mass M , $\delta_c(z)$ is the critical overdensity for spherical top-hat collapse at redshift z (Eke, Cole & Frenk 1996),

$$n_{\text{eff}} = -6 \frac{d \ln \sigma}{d \ln M} - 3, \quad (4)$$

$$\omega = \sqrt{ca} \frac{\delta_c(z)}{\sigma}, \quad (5)$$

$$G_1 = \exp\left(-\frac{1}{2} \left[\frac{(\log \omega - 0.788)}{0.6}\right]^2\right), \quad (6)$$

$$G_2 = \exp\left(-\frac{1}{2} \left[\frac{(\log \omega - 1.138)}{0.2}\right]^2\right), \quad (7)$$

$A' = 0.310$, $ca = 0.764$ and $p = 0.3$ as in equations (11) and (12) of Reed et al. (2007).⁴ The mass variance, $\sigma^2(M)$, is computed using the cold dark matter transfer function of Eisenstein & Hu (1999) together with a scale-free primordial power spectrum of slope n_s and normalization σ_8 .

When constructing samples of dark matter haloes we compute the number of haloes, N_{halo} , expected in some volume V of the Universe within a logarithmic mass interval, $\Delta \ln M_v$, according to this mass function, requiring that the number of haloes in the interval never exceeds N_{max} and is never less than N_{min} to ensure a fair sample. We then generate halo masses at random using a Sobol’ sequence (Sobol’ 1967) drawn from a distribution which produces, on average, N_{halo} haloes in each interval. This ensures a quasi-random, fair sampling of haloes of all masses with no quantization of halo mass and with sub-Poissonian fluctuations in the number of haloes in any mass interval.

2.5.2 Merger trees

Dark matter halo merger trees, which describe the hierarchical growth of structure in a cold dark matter universe, form the backbone of our model within which the process of galaxy formation proceeds. Merger trees are either constructed through a variant of the extended Press–Schechter Monte Carlo methodology or extracted from N -body simulations.

When constructing trees using Monte Carlo methods, we employ the merger tree algorithm described by Parkinson et al. (2008) which is itself an empirical modification of that described by Cole et al. (2000). We adopt the parameters $(G_0, \gamma_1, \gamma_2) = (0.57, 0.38, -0.01)$ that Parkinson et al. (2008) found provided the best fit⁵ to the statistics of halo progenitor masses measured from the Millennium Simulation. We typically use a mass resolution (i.e. the lowest mass halo which we trace in our trees) of $5 \times 10^9 h^{-1} M_\odot$, which is sufficient to achieve resolved galaxy properties for all of the calculations considered in this work. An exception is when we consider Local Group satellites (see

⁴With minor corrections to the published version (Reed, private communication).

⁵Benson (2008) found an alternative set of parameters which provided a better match to the evolution of the overall halo mass function but performed slightly less well (although still quite well) for the progenitor halo mass functions. We have chosen to use the parameters of Parkinson et al. (2008) as for the properties of galaxies we wish to get the progenitor masses as correct as possible.

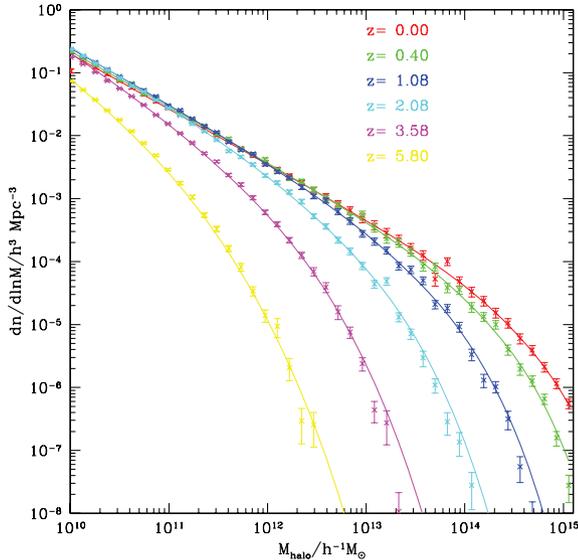


Figure 1. The dark matter halo mass function is shown at a number of different redshifts. Solid lines indicate the mass function expected from equation (3) while points with error bars indicate the mass function constructed using merger trees from our model. The trees in question were initiated at $z = 0$ and grown back to higher redshifts using the methods of Parkinson et al. (2008).

Section 4.10), for which we instead use a mass resolution of $10^7 h^{-1} M_{\odot}$. Fig. 1 shows the resulting dark matter halo mass functions at several different redshifts and demonstrates that they are in good agreement with that expected from equation (3).

2.5.3 (Lack of) halo formation events

Cole et al. (2000) identified certain haloes in each dark matter merger tree as being newly formed. ‘Formation’ in this case corresponded to the point where a halo had doubled in mass since the previous formation event. The characteristic circular velocity and spin of haloes were held fixed in between formation events, and the time available for hot gas in a halo to cool was measured from the most recent formation event (such that the cooling radius was reduced to zero at each formation event). Additionally, any gas ejected by feedback was only allowed to begin recooling after a formation event, and any satellite haloes that had not yet merged with the central galaxy of their host halo were assumed to have their orbits randomized by the formation event and consequently their merger time-scales were reset.

While computationally useful, these formation events lack any solid physical basis. As such, we have excised them from our current implementation of GALFORM. Halo properties (virial velocity and spin) now change at each time-step in response to mass accretion. Additionally, the cooling and merging calculations no longer make use of formation events (see Sections 2.6 and 2.8, respectively).

2.5.4 Density profiles

Recent N -body studies (Navarro et al. 2004; Merritt et al. 2005; Prada et al. 2006) indicate that the density profiles of dark matter haloes in CDM cosmologies are better described by the Einasto profile (Einasto 1965) than the NFW profile (Navarro et al. 1997).

As such, we use the Einasto density profile,

$$\rho(r) = \rho_{-2} \exp\left(-\frac{2}{\alpha} \left[\left(\frac{r}{r_{-2}}\right)^{\alpha} - 1\right]\right), \quad (8)$$

where r_{-2} is a characteristic radius at which the logarithmic slope of the density profile equals -2 and α is a parameter which controls how rapidly the logarithmic slope varies with radius. To fix the value of α we adopt the fitting formula of Gao et al. (2008), truncated so that α never exceeds 0.3,

$$\alpha = \begin{cases} 0.155 + 0.0095\nu^2 & \text{if } \nu < 3.907 \\ 0.3 & \text{if } \nu \geq 3.907, \end{cases} \quad (9)$$

where $\nu = \delta_c(a)/\sigma(M)$ which is a good match to haloes in the Millennium Simulation.⁶ The value of r_{-2} for each halo is determined from the known virial radius, r_v , and the concentration, $c_{-2} \equiv r_v/r_{-2}$. Concentrations are computed using the method of Navarro et al. (1997) but with the best-fitting parameters found by Gao et al. (2008).

Various integrals over the density and mass distribution are needed to compute the enclosed mass, angular momentum, velocity dispersion, gravitational energy and so on of the Einasto profile. Some of these may be expressed analytically in terms of incomplete gamma functions (Cardone, Piedipalumbo & Tortora 2005). Expressions for the mass and gravitational potential are provided by Cardone et al. (2005). One other integral, the angular momentum of material interior to some radius, can also be found analytically:

$$\begin{aligned} J(r) &= \pi^2 V_{\text{rot}} \int_0^r r'^{(3+\alpha_{\text{rot}})} \rho(r') dr' \\ &= \pi^2 V_{\text{rot}} \rho_{-2} r_{-2}^{4+\alpha_{\text{rot}}} \frac{e^{2/\alpha}}{\alpha} \left(\frac{\alpha}{2}\right)^{4+\alpha_{\text{rot}}} \\ &\quad \times \Gamma\left(\frac{4+\alpha_{\text{rot}}}{\alpha}, \frac{2(r/r_{-2})^{\alpha}}{\alpha}\right), \end{aligned} \quad (10)$$

where the specific angular momentum at radius r is assumed to be $r V_{\text{rot}}(r/r_v)^{\alpha_{\text{rot}}}$ and Γ is the lower incomplete gamma function. Other integrals (e.g. gravitational energy) are computed numerically as needed.

2.5.5 Angular momentum

As first suggested by Hoyle (1949), and developed further by Doroshkevich (1970), Peebles (1969) and White (1984), the angular momenta of dark matter haloes arises from tidal torques from surrounding large-scale structure and is usually characterized by the dimensionless spin parameter,

$$\lambda \equiv \frac{J_v |E_v|^{1/2}}{GM_v^{5/2}}, \quad (11)$$

where J_v is the angular momentum of the halo and E_v its energy (gravitational plus kinetic). The distribution of λ has been measured numerous times from N -body simulations (Barnes & Efstathiou 1987; Efstathiou et al. 1988; Warren et al. 1992; Cole & Lacey 1996; Lemson & Kauffmann 1999) and found to be reasonably well approximated by a lognormal distribution. More recent estimates by

⁶Gao et al. (2008) were not able to probe the behaviour of α in the very high ν regime. Extrapolating their formula to $\nu > 4$ is not justified and we instead choose to truncate it at a maximum of $\alpha = 0.3$.

Bett et al. (2007) using the Millennium Simulation show a somewhat different form for this distribution:

$$P(\lambda) \propto \left(\frac{\lambda}{\lambda_0}\right)^3 \exp\left[-\alpha_\lambda \left(\frac{\lambda}{\lambda_0}\right)^{3/\alpha_\lambda}\right], \quad (12)$$

where $\alpha_\lambda = 2.509$ and $\lambda_0 = 0.04326$ are parameters.

Cole et al. (2000) assigned spins to dark matter haloes by drawing them at random from the distribution of Cole & Lacey (1996). This approach has the disadvantage that spin is not influenced by the merging history of a given dark matter halo and, furthermore, spin can vary dramatically from one time-step to the next even if a halo experiences no (or only very minor) merging. This was not a problem for Cole et al. (2000), who made use of the spin of each newly formed halo, ignoring any variation between formation events.⁷ However, in our case, such behaviour would be problematic. We therefore revisit an idea first suggested by Vitvitska et al. (2002; see also Maller, Dekel & Somerville 2002). They followed the contribution to the angular momentum of each halo from its progenitor haloes (which carry angular momentum in both their internal spin and orbit). Note that the angular momentum still arises via tidal torques (which are responsible for the orbital angular momenta of merging haloes).

Haloes in the merger tree which have no progenitors are assigned a spin by drawing at random from the distribution of Bett et al. (2007). For haloes with progenitors, we proceed as follows.

(i) Compute the internal angular momenta of all progenitor haloes using their previously assigned spin and equation (11).

(ii) Select orbital parameters (specifically the orbital angular momentum) for each merging pair of progenitors by drawing at random from the distribution found by Benson (2005).

(iii) Sum the internal and orbital angular momenta of all progenitors assuming no correlations between the directions of these vectors.⁸

(iv) Determine the spin parameter of the new halo from this summed angular momentum and equation (11).

Benson (2005) report orbital velocities for merging haloes and give expressions for the angular momenta of those orbits assuming point mass haloes. While this will be a reasonable approximation for high mass ratio mergers it will fail for mergers of comparable mass haloes. In addition, halo mergers may not necessarily conserve angular momentum in the sense that some material, plausibly with the highest specific angular momentum, may be thrown out during the merging event leaving the final halo with a lower angular momentum. To empirically account for these two factors we divide the orbital angular momentum by a factor of $f_2 \equiv 1 + M_2/M_1$ (where $M_2 < M_1$ are the masses of the dark matter haloes). We find that this empirical factor leads to good agreement with the measured N -body spin distribution, but could be justified more rigorously by measuring the angular momentum (accounting for finite size effects) of the progenitor and remnant haloes in N -body mergers.

To test the validity of this approach we generated 51 625 Monte Carlo realizations of merger trees drawn from a halo mass function consistent with that of the Millennium Simulation and with a range

⁷As it seems reasonable to assume that the spins of a halo at two successive formation events, i.e. separated by a factor of 2 in halo mass, would be only weakly correlated.

⁸Additionally, we are assuming that mass accretion below the resolution of the merger tree contributes the same mean specific angular momentum as accretion above the resolution.

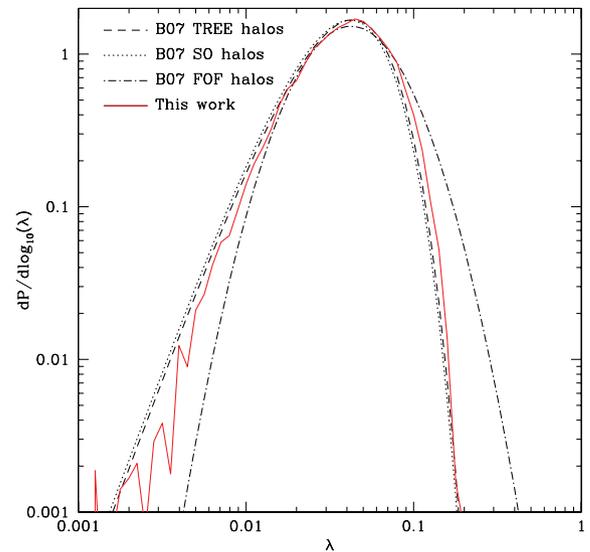


Figure 2. The distribution of dark matter halo spin parameters. Black lines show measurements of this distribution from the Millennium N -body simulation (Bett et al. 2007), for three different group finding algorithms. Bett et al. (2007) note that the ‘TREE’ haloes give the most accurate determination of the spin distribution. The red line shows the results of the Monte Carlo model described in this work, using 51 625 Monte Carlo realizations of merger trees spanning a range of masses identical to that used by Bett et al. (2007).

of masses consistent with that for which Bett et al. (2007) were able to measure spin parameters and applied the above procedure. Fig. 2 shows the results of this test. We find remarkably good agreement between the distribution of spin measured by Bett et al. (2007) and the results of our Monte Carlo model. It should be noted that our assumption of no correlation between the various angular momenta vectors of progenitor haloes is not correct. However, Benson (2005) shows that any such correlations are weak. Therefore, given the success of a model with no correlations, we choose to ignore them.

Our results are in good agreement with previous attempts to model the halo spin distribution in this way. Maller et al. (2002) found good agreement with N -body results using the same principles, although they found that introducing some correlation between the directions of spin and orbital angular momenta improved their fit. Vitvitska et al. (2002) also found generally good agreement with N -body simulations using orbital parameters of haloes drawn from an N -body simulation. Both of these earlier calculations relied on much less well calibrated orbital parameter distributions for merging haloes and the simulations to which they compared their results had significantly poorer statistics than the Millennium Simulation. Our results confirm that this approach to calculating halo spins from a merger history still works extremely well even when confronted with the latest high-precision measures of the spin distribution.

2.6 Cooling model

The cooling model described by Cole et al. (2000) determines the mass of gas able to cool in any time-step by following the propagation of the cooling radius in a notional hot gas density profile⁹

⁹We refer to this as a ‘notional’ profile since it is taken to represent the profile before any cooling can occur. Once some cooling occurs presumably the actual profile adjusts in some way to respond to this and so will no longer look like the notional profile, even outside of the cooling radius.

which is fixed when a halo is flagged as ‘forming’ and is only updated when the halo undergoes another formation event. The mass of gas able to cool in any given time-step is equal to the mass of gas in this notional profile between the cooling radius at the present step and that at the previous step. The cooling time is assumed to be the time since the formation event of the halo. Any gas which is reheated into or accreted by the halo is ignored until the next formation event, at which point it is added to the hot gas profile of the newly formed halo. The notional profile is constructed using the properties (e.g. scale radius, virial temperature, etc.) of the halo at the formation event and retains a fixed metallicity throughout, corresponding to the metallicity of the hot gas in the halo at the formation event.

In this work we implement a new cooling model. We do away with the arbitrary ‘formation’ events and instead use a continuously updating estimate of cooling time and halo properties. For the purposes of this calculation we define the following quantities:

- (i) M_{hot} : the current mass of hot (i.e. as yet uncooled) gas remaining in the notional profile.
- (ii) M_{cooled} : the mass of gas which has cooled out of the notional profile into the galaxy phase.
- (iii) M_{reheated} : the mass of gas which has been reheated (by SNe feedback) but has yet to be reincorporated back into the hot gas component.
- (iv) M_{ejected} : the mass of gas which has been ejected beyond the virial radius of this halo, but which may later reaccrete into other, more massive haloes.

The notional profile always contains a mass $M_{\text{total}} = M_{\text{hot}} + M_{\text{cooled}} + M_{\text{reheated}}$. The properties (density normalization, core radius) are reset, as described in Section 2.6.3, at each time-step. The previous infall radius (i.e. the radius within which gas was allowed to infall and accrete on to the galaxy) is computed by finding the radius which encloses a mass $M_{\text{cooled}} + M_{\text{reheated}}$ (i.e. the mass previously removed from the hot component) in the current notional profile.

We aim to compute a time available for cooling for the halo, t_{avail} , from which we can compute a cooling radius in the usual way (i.e. by finding the radius in the notional profile at which $t_{\text{cool}} = t_{\text{avail}}$). In Cole et al. (2000), the time available for cooling is simply set to the time since the last formation event of the halo.

At any time, the rate of cooling per particle is just $\Lambda(T, \mathbf{Z}, n_{\text{H}}, F_{\nu})n_{\text{H}}$ where $\Lambda(T, \mathbf{Z}, n_{\text{H}}, F_{\nu})$ is the cooling function, and n_{H} the number density of hydrogen, \mathbf{Z} a vector of metallicity (such that the i th component of \mathbf{Z} is the abundance by mass of the i th element) and F_{ν} the spectrum of background radiation. The total cooling luminosity is then found by multiplying by the number of particles, N , in some volume V that we want to consider. If we take this volume to be the entire halo then $N \equiv M_{\text{total}}/\mu m_{\text{H}}$. If we integrate this luminosity over time, we find the total energy lost through cooling. The total thermal energy in our volume V is just $3Nk_{\text{B}}T/2$. The gas will have completely cooled once the energy lost via cooling equals the original thermal energy, i.e.

$$3Nk_{\text{B}}T_{\nu}/2 = \int_0^t \Lambda(t')n_{\text{H}}N dt', \quad (13)$$

where for brevity we write $\Lambda(t) \equiv \Lambda[T_{\nu}(t), \mathbf{Z}(t), n_{\text{H}}(t), F_{\nu}(t)]$. We can write this as

$$t_{\text{cool}} = t_{\text{avail}}, \quad (14)$$

where

$$t_{\text{cool}}(t) = \frac{3k_{\text{B}}T_{\nu}(t)}{2\Lambda(t)n_{\text{H}}} \quad (15)$$

is the usual cooling time and

$$t_{\text{avail}} = \frac{\int_0^t \Lambda(t')n_{\text{H}}(t')N dt'}{\Lambda(t)n_{\text{H}}(t)N} \quad (16)$$

is the time available for cooling. We can re-write this as

$$t_{\text{avail}} = \frac{\int_0^t [T_{\nu}(t')N/t_{\text{cool}}(t')] dt'}{[T_{\nu}(t)N/t_{\text{cool}}(t)]}. \quad (17)$$

In the case of a static halo, where T_{ν} , \mathbf{Z} , F_{ν} and N are independent of time, t_{avail} reduces to the time since the halo came into existence as we might expect. For a non-static halo the above makes more physical sense. For example, consider a halo which is below the 10^4K cooling threshold from time $t = 0$ to time $t = t_4$, and then moves above that threshold (with fixed properties after this time). Since $t_{\text{cool}} = \infty$ [i.e. $\Lambda(t) = 0$] before t_4 in this case we find that $t_{\text{avail}} = t - t_4$ as expected. Note that since the number of particles, N , appears in both the numerator and denominator of equation (17) we can, in practice, replace N by M_{total} without changing the resulting time.

The cooling time in the above must be computed for a specific value of the density. We choose to use the cooling time at the mean density of the notional profile at each time-step. This implicitly assumes that the density of each mass element of gas in the notional profile has the same time dependence as the mean density of the profile, i.e. that the profile evolves in a self-similar way and that $\Lambda(t)$ is independent of n_{H} (which will only be true in the collisional ionization limit). This may not be true in general, but serves as an approximation allowing us to describe the cooling of the entire halo with just a single integral.¹⁰

Having computed the time available for cooling we solve for the cooling radius in the notional profile at which $t_{\text{cool}}(r_{\text{cool}}) = t_{\text{avail}}$ (as described in Section 2.6.4). We also estimate the largest radius from which gas has had sufficient time to freefall to the halo centre (as described in Section 2.6.5). The current infall radius is taken to be the smaller of the cooling and freefall radii. Any mass between the current infall radius and that at the previous time-step is allowed to infall on to the galaxy during the current time-step – that is, it is transferred from M_{hot} to M_{cooled} .

One refinement which must be introduced is to limit the integral

$$\mathcal{E} = \frac{3}{2}k_{\text{B}} \int_0^t [T_{\nu}(t')N/t_{\text{cool}}(t')] dt', \quad (18)$$

so that the total radiated energy cannot exceed the total thermal energy of the halo. This limit is given by

$$\mathcal{E}_{\text{max}} = \frac{3}{2}k_{\text{B}}T_{\nu}(t)N \frac{\bar{\rho}_{\text{total}}}{\rho_{\text{total}}(r_{\nu})}, \quad (19)$$

where $\bar{\rho}_{\text{total}}$ is the mean density of the notional profile and $\rho_{\text{total}}(r_{\nu})$ is the density of the notional profile at the virial radius. For the entire halo (out to the virial radius) to cool takes longer than for gas at the mean density of the halo to cool, by a factor of $\bar{\rho}_{\text{total}}/\rho_{\text{total}}(r_{\nu})$. This is the origin of the ratio of densities in equation (19).

We must then consider two additional effects: accretion (Section 2.6.1) and reheating (Section 2.6.2). The cooling model is then fully specified once we specify the distribution of gas in the notional profile (Section 2.6.3), determine a cooling radius (Section 2.6.4) and freefall radius (Section 2.6.5), and consider how to compute the angular momentum of the infalling gas (Section 2.6.6).

¹⁰A more elaborate model could compute a separate integral for each shell of gas, following the evolution of its density as a function of time as the profile evolves due to continued infall and cooling.

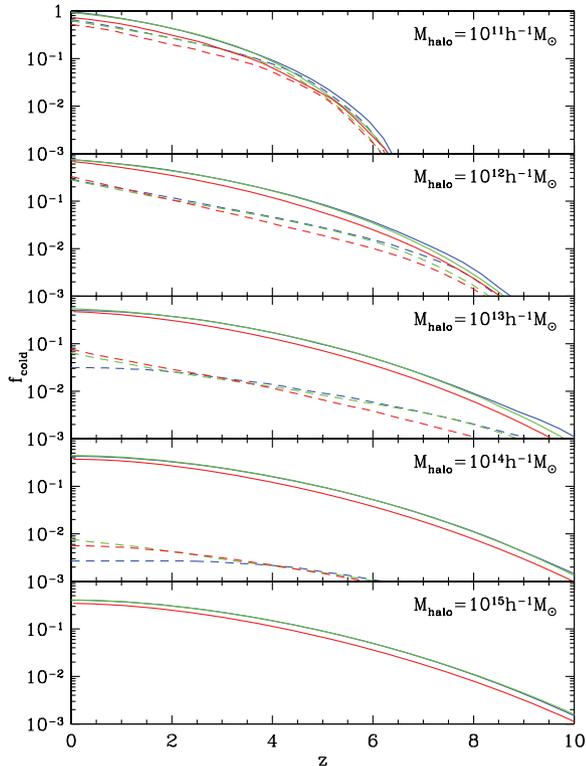


Figure 3. The mean cooled gas fractions in the merger trees of haloes with masses 10^{11} , 10^{12} , 10^{13} , 10^{14} and $10^{15} h^{-1} M_{\odot}$ at $z = 0$ are shown by coloured lines. Green lines show results from the cooling model described in this work while red lines indicate the model of Cole et al. (2000) and blue lines the cooling model of Bower et al. (2006). Solid lines show the total cooled fraction in all branches of the merger trees, while dashed lines show the cooled fraction in the main branch of the trees. For the purposes of this figure, no star or black hole formation was included in these calculations, so consequently there is no reheating of gas, expulsion of gas from the halo or metal enrichment. Additionally, no galaxy merging was allowed. As such, the differences between models arise purely from their different implementations of cooling.

2.6.1 Accretion

When a halo accretes another halo, we merge their notional gas profiles. Since the integral, $\mathcal{E} = \int (NT_v/t_{\text{cool}}) dt$, that we are computing is the total energy lost we simply add \mathcal{E} from the accreted halo to that of the halo it accretes into. This gives the total energy lost from the combined notional profile. However, we must consider the fact that only a fraction $M_{\text{hot}}/M_{\text{total}}$ of the gas from the accreted halo is added to the hot gas reservoir of the combined halo (the mass M_{cooled} from the accreted halo becomes the satellite galaxy while the mass M_{reheated} is added to the reheated reservoir of the new halo to await reincorporation into the hot component; see Section 2.6.2). We simply multiply the integral \mathcal{E} of the accreted halo by this fraction before adding it to the new halo.

Fig. 3 compares the mean cooled gas fractions in haloes of different masses computed using the cooling model described here (green lines) and two previous cooling models used in GALFORM: that of Cole et al. (2000; red lines) and that of Bower et al. (2006; blue lines). The only significant difference between the cooling implementations of Cole et al. (2000) and Bower et al. (2006) is that Bower et al. (2006) allow reheated gas to gradually return to the hot component (and so be available for re-cooling) at each time-step (in

the same manner as in the present work), while Cole et al. (2000) simply accumulated this reheated gas and returned it all to the hot component only at the next halo formation event (i.e. after a halo mass doubling). No star or black hole formation was included in these calculations, so consequently there is no reheating of gas, expulsion of gas from the halo or metal enrichment. Additionally, no galaxy merging was allowed. The thick lines show the total cooled fraction in all branches of the merger trees, while the thin lines show the cooled fraction in the main branch of the trees.¹¹

The cooling model utilized by Bower et al. (2006) was similar to that of Cole et al. (2000) except that it allowed accreted and reheated gas to rejoin the hot gas reservoir in a continuous manner rather than only at each halo formation event. Additionally, it used the current properties of the halo (e.g. virial temperature) to compute cooling rates rather than the properties of the halo at the previous formation event. As such, the Bower et al. (2006) model contains many features of the current cooling model, but retains the fundamental division of the merger tree into discrete branches as in the Cole et al. (2000) model.

We find that, in general, the cooling model described here predicts a total cooled fraction very close to that predicted by the cooling model of Bower et al. (2006), the exception being at very early times in low-mass haloes where it gives a slightly lower value. The difference of course is that the new model does not contain artificial resets in the cooling calculation which, although they make little difference to this statistic, have a strong influence on, for example, calculations of the angular momentum of cooling gas. Both of these models predict somewhat more total cooled mass than the Cole et al. (2000) model. This is due entirely to the allowance of accreted gas to begin cooling immediately.

If we consider the cooled fraction in the main branch of each tree (i.e. the mass in what will become the central galaxy in the final halo) we see rather different behaviour. At early times, the new model tracks the Bower et al. (2006) model. At late times, however, the Bower et al. (2006) model shows a much lower cooling rate while the new model tracks the cooled fraction in the Cole et al. (2000) model quite closely. This occurs in massive haloes where in the Bower et al. (2006) model the use of the current halo properties to determine cooling rates results in ever increasing cooling times as the virial temperature of the halo increases and the halo density (and hence hot gas density) decline. The Cole et al. (2000) model is less susceptible to this as it computes halo properties based on the halo at formation. The new cooling model produces results comparable to the Cole et al. (2000) model since, while it utilizes the present properties of the halo just as does the Bower et al. (2006) model, it retains a memory of the early properties of the halo.

2.6.2 Reheating

When gas is reheated (via feedback; Section 2.12) we assume that it is heated to the virial temperature of the current halo (i.e. the host halo for satellite galaxies) and is placed into a reservoir M_{reheated} . Mass is moved from this reservoir back into the hot gas reservoir on a time-scale of the order of the halo dynamical time, τ_{dyn} . Specifically,

¹¹ We define the main branch of the merger tree as the set of progenitor haloes found by starting from the final halo and repeatedly stepping back to the most massive progenitor of the current halo at each time-step. It should be noted that the definition is not unique, and can depend on the time resolution of the merger tree. It can also result in situations where the main branch does not correspond to the most massive progenitor halo at a given time-step.

mass is returned to the hot phase at a rate

$$\dot{M}_{\text{hot}} = \alpha_{\text{reheat}} \frac{M_{\text{reheated}}}{\tau_{\text{dyn}}} \quad (20)$$

during each time-step. This effectively undoes the cooling energy loss which caused this gas to cool previously. The energy integral \mathcal{E} is therefore modified by subtracting from it an amount $\Delta N_{\text{reheated}} T_v$ where $\Delta N_{\text{reheated}}$ is the number of particles reheated.

Similarly, the notional profile is allowed to ‘forget’ about any cooled gas on a time-scale of the order of the dynamical time (i.e. we assume that the notional profile adjusts to the loss of this gas). This is implemented by removing mass from the cooled reservoir at a rate

$$\dot{M}_{\text{cooled}} = -\alpha_{\text{remove}} \frac{M_{\text{cooled}}}{\tau_{\text{dyn}}}. \quad (21)$$

2.6.3 Hot gas distribution

The hot gas is assumed to be distributed in a notional profile with a run of density consistent with that found in hydrodynamical simulations (Sharma & Steinmetz 2005; Stringer & Benson 2007). Sharma & Steinmetz (2005) performed non-radiative cosmological spectral energy distribution (SPH) simulations and studied the properties of the hot gas in dark matter haloes. These simulations are therefore well suited to our purposes since they relate to the notional profile which is defined to be that in the absence of any cooling. The gas density profiles found by Sharma & Steinmetz (2005) are well described by the expression:

$$\rho(r) \propto \frac{1}{(r + r_{\text{core}})^3}, \quad (22)$$

where r_{core} is a characteristic core radius for the profile. We choose to set $r_{\text{core}} = a_{\text{core}} r_v$ where a_{core} is a parameter whose value is the same for all haloes at all redshifts. The simulations suggest that $a_{\text{core}} \approx 0.05$ (Stringer & Benson 2007), but we will treat a_{core} as a free parameter to be constrained by observational data. The density profile is normalized such that

$$\int_0^{r_v} \rho(r) 4\pi r^2 dr = M_{\text{total}}, \quad (23)$$

and the hot gas is assumed to be isothermal at the virial temperature

$$T_v = \frac{1}{2} \frac{\mu m_{\text{H}}}{k} V_v^2 \quad (24)$$

with a metallicity equal to $Z = M_{Z,\text{hot}}/M_{\text{hot}}$. Initially, $M_{Z,\text{hot}} = 0$ but can become non-zero due to metal production and outflows as a result of star formation and feedback.

2.6.4 Cooling radius

Given the time available for cooling from equation (17) the cooling radius is found by solving

$$t_{\text{avail}} = \frac{\frac{3}{2}(n_{\text{tot}}/n_{\text{H}})k_{\text{B}}T_v}{n_{\text{H}}(r_{\text{cool}})\Lambda(t)}, \quad (25)$$

where n_{tot} is the total number density of the atoms in the gas. Due to the dependence of $\Lambda(t)$ on density when a photoionizing background is present (see Section 5.1) this equation must be solved numerically.

2.6.5 Freefall radius

To compute the mass of gas which can actually reach the centre of a halo potential well at any given time we require that not only has the gas had time to cool but also that it has had time to freefall to the centre of the halo starting from zero velocity at its initial radius. To estimate the maximum radius from which cold gas could have reached the halo centre through freefall we proceed as follows. We compute a time available for freefall in the halo, $t_{\text{avail,ff}}$, using equation (17), but limit the integral \mathcal{E} (defined in equation 18) such that the time available cannot exceed the freefall time at the virial radius. We then solve the freefall equation

$$\int_0^{r_{\text{ff}}} \frac{dr'}{\sqrt{2[\Phi(r') - \Phi(r_{\text{ff}})]}} = t_{\text{avail,ff}}, \quad (26)$$

where $\Phi(r)$ is the gravitational potential of the halo, for the radius r_{ff} at which the freefall time equals the time available. Only gas within the minimum of the cooling and freefall radii at each time-step is allowed to reach the centre of the halo and become part of the forming galaxy.

2.6.6 Angular momentum

The angular momentum of gas in the notional halo is tracked using a similar approach as for the mass. We define the following quantities:

J_{hot} : the total angular momentum in the M_{hot} reservoir of the notional profile;

J_{cooled} : the total angular momentum in the M_{cooled} reservoir of the notional profile;

J_{reheated} : the total angular momentum in the M_{reheated} reservoir of the notional profile;

j_{new} : the specific angular momentum which newly accreted material must have in order to produce the correct change in angular momentum for this halo.¹²

J_{cooled} and J_{reheated} are initialized to zero at the start of the calculation. J_{hot} is initialized by assuming that any material accreted below the resolution of the merger tree arrives with the mean specific angular momentum of the halo. Angular momentum is then tracked using the following method:

(i) At the start of a time-step, all three angular momentum reservoirs from the most massive progenitor halo are added to those of the current halo.

(ii) We assume that the specific angular momentum of the gas halo is distributed according to the results of Sharma & Steinmetz (2005) such that the differential distribution of specific angular momentum, j , is given by

$$\frac{1}{M} \frac{dM}{dj} = \frac{1}{j_{\text{d}}^{\alpha_j} \Gamma(\alpha_j)} j^{\alpha_j-1} e^{-j/j_{\text{d}}}, \quad (27)$$

where Γ is the gamma function, M is the total mass of gas, $j_{\text{d}} = j_{\text{tot}}/\alpha$ and j_{tot} is the mean specific angular momentum of the gas. The parameter α_j is chosen to be 0.89, consistent with the median value found by Sharma & Steinmetz (2005) in simulated haloes.

¹²The angular momentum of a halo differs from that of its main progenitor due to an increase in mass, change in virial radius and change in spin parameter. j_{new} is computed by finding the difference in the angular momentum of a halo and its main progenitor and dividing by their mass difference. Note that this quantity can therefore be negative.

The fraction of mass with specific angular momentum less than j is then given by

$$f(< j) = \gamma \left(\alpha_j, \frac{j}{j_d} \right), \quad (28)$$

where γ is the incomplete gamma function. Once the mass of gas cooling in any given time-step is known the above allows the angular momentum of that gas to be found. This amount is added to the J_{cooled} reservoir.

(iii) If $J_{\text{reheated}} > 0$ then an angular momentum

$$\Delta J_{\text{hot}} = \begin{cases} J_{\text{reheated}} \alpha_{\text{reheat}} \Delta t / \tau_{\text{dyn}} & \text{if } \alpha_{\text{reheat}} \Delta t < \tau_{\text{dyn}} \\ J_{\text{reheated}} & \text{if } \alpha_{\text{reheat}} \Delta t \geq \tau_{\text{dyn}} \end{cases} \quad (29)$$

is transferred back to the hot phase, consistent with the fraction of mass returned to the hot phase (see Section 2.6.2).

(iv) When a halo becomes a satellite of a larger halo, J_{hot} of the larger halo is increased by an amount, $j_{\text{new}} M_{\text{hot,sat}}$. This accounts for the orbital angular momentum of the gas in the satellite halo assuming that, on average, satellites have specific angular momentum of j_{new} . We do the same for J_{reheated} , assuming that the M_{reheated} reservoir of the satellite arrives with the same specific angular momentum.

(v) When gas is ejected from a galaxy disc to join the reheated reservoir, it is ejected with the mean specific angular momentum of the disc. Gas ejected during a starburst is also assumed to be ejected with the mean pseudo-specific angular momentum¹³ of the bulge.

Because j_{new} can be negative on occasion it is possible that $J_{\text{hot}} < 0$ can occur. This, in turn, can lead to a galaxy disc with a negative angular momentum. We do not consider this to be a fundamental problem due to the vector nature of angular momentum. When computing disc sizes we simply consider the magnitude of the disc angular momentum, ignoring the sign.

2.6.7 Cooling/heating rates of hot gas in haloes

The cooling model described above requires knowledge of the cooling function, $\Lambda(T, Z, n_{\text{H}}, F_{\nu})$. Given a gas metallicity and density and the spectrum of the ionizing background we can compute cooling and heating rates for gas in dark matter haloes. Calculations were performed with version 08.00 of CLOUDY, last described by Ferland et al. (1998). In practice, we compute cooling/heating rates as a function of temperature, density and metallicity using the self-consistently computed photon background (Section 2.10) after each time-step. The rates are computed on a grid which is then interpolated on to find the cooling/heating rate for any given halo.

Chemical abundances are assumed to behave as follows.

(i) $Z = 0$: ‘zero’ metallicity corresponding to the ‘primordial’ abundance ratios as used by CLOUDY version 08.00 (see the *Hazy* documentation of CLOUDY for details).

(ii) $[\text{Fe}/\text{H}] < -1$: ‘primordial’ abundance ratios from Sutherland & Dopita (1993).

(iii) $[\text{Fe}/\text{H}] \geq 1$: solar abundance ratios as used by CLOUDY version 08.00 (see the *Hazy* documentation of CLOUDY for details).

However, since our model can track the abundances of individual elements we know the abundances in each cooling halo. In principle, we could recompute a cooling/heating rate for each halo using its specific abundances as input into CLOUDY. This is computationally

impractical however. Instead, we follow the approach of Martínez-Serrano et al. (2008) who perform a principal components analysis (PCA) to find the optimal linear combination of abundances which minimizes the variance between cooling/heating rates computed using that linear combination as a parameter and a full calculation using all abundances. The best linear combination turns out to be a function of temperature. We therefore track this linear combination of abundances at 10 different temperatures for all of the gas in our models and use it instead of metallicity when computing cooling/heating rates.

Compton cooling: Cole et al. (2000) allowed hot halo gas to cool via two-body collisional radiative processes. However, as we go to higher redshifts the effect of Compton cooling must be considered. The Compton cooling time-scale is given by (Peebles 1968)

$$\tau_{\text{Compton}} = \frac{3m_e c(1 + 1/x_e)}{8\sigma_{\text{T}} a T_{\text{CMB}}^4 (1 - T_{\text{CMB}}/T_e)}, \quad (30)$$

where $x_e = n_e/n_t$, n_e is the electron number density, n_t is the number density of all atoms and ions, T_{CMB} is the cosmic microwave background (CMB) temperature and T_e is the electron temperature of the gas.

The electron fraction, x_e , is determined from photoionization equilibrium computed using CLOUDY (see above).

Molecular hydrogen cooling: the molecular hydrogen cooling time-scale is found by first estimating the abundance, $f_{\text{H}_2,c}$, of molecular hydrogen that would be present if there is no background of H_2 -dissociating radiation from stars. For gas with hydrogen number density n_{H} and temperature T_{V} the fraction is (Tegmark et al. 1997)

$$f_{\text{H}_2,c} = 3.5 \times 10^{-4} T_3^{1.52} [1 + (7.4 \times 10^8 (1+z)^{2.13} \times \exp\{-3173/(1+z)\} / n_{\text{H}1})]^{-1}, \quad (31)$$

where T_3 is the temperature T_{V} in units of 1000 K and $n_{\text{H}1}$ is the hydrogen density in units of cm^{-3} . Using this initial abundance, we calculate the final H_2 abundance, still in the absence of a photodissociating background, as

$$f_{\text{H}_2} = f_{\text{H}_2,c} \exp\left(\frac{-T_{\text{V}}}{51\,920\text{K}}\right), \quad (32)$$

where the exponential cut-off is included to account for collisional dissociation of H_2 , as in Benson et al. (2006).

Finally, the cooling time-scale due to molecular hydrogen was computed using (Galli & Palla 1998)

$$\tau_{\text{H}_2} = 6.56419^{-33} T_e f_{\text{H}_2}^{-1} n_{\text{H}1}^{-1} \Lambda_{\text{H}_2}^{-1}, \quad (33)$$

where

$$\Lambda_{\text{H}_2} = \frac{\Lambda_{\text{LTE}}}{1 + n^{\text{cr}}/n_{\text{H}}}, \quad (34)$$

where

$$\frac{n^{\text{cr}}}{n_{\text{H}}} = \frac{\Lambda_{\text{H}_2}(\text{LTE})}{\Lambda_{\text{H}_2}[n_{\text{H}} \rightarrow 0]}, \quad (35)$$

and

$$\log_{10} \Lambda_{\text{H}_2}[n_{\text{H}} \rightarrow 0] = -103 + 97.59 \ln(T) - 48.05 \ln(T)^2 + 10.8 \ln(T)^3 - 0.9032 \ln(T)^4 \quad (36)$$

is the cooling function in the low-density limit (independent of hydrogen density) and we have used the fit given by Galli & Palla (1998),

$$\Lambda_{\text{LTE}} = \Lambda_r + \Lambda_v \quad (37)$$

¹³As defined by Cole et al. (2000; their equation C11) and equal to the product of the bulge half-mass radius and the circular velocity at that radius.

is the cooling function in local thermodynamic equilibrium and

$$\Lambda_r = \frac{1}{n_{\text{H}_1}} \left\{ \frac{9.5 \times 10^{-22} T_3^{3.76}}{1 + 0.12 T_3^{2.1}} \exp\left(-\left[\frac{0.13}{T_3}\right]^3\right) + 3 \times 10^{-24} \exp\left(-\frac{0.51}{T_3}\right) \right\} \text{erg cm}^3 \text{s}^{-1}, \quad (38)$$

$$\Lambda_v = \frac{1}{n_{\text{H}_1}} \left\{ 6.7 \times 10^{-19} \exp\left(-\frac{5.86}{T_3}\right) + 1.6 \times 10^{-18} \exp\left(-\frac{11.7}{T_3}\right) \right\} \text{erg cm}^3 \text{s}^{-1} \quad (39)$$

are the cooling functions for rotational and vibrational transitions in H_2 (Hollenbach & McKee 1979).

The model also allows for an estimate of the rate of molecular hydrogen formation on dust grains using the approach of Cazaux & Spaans (2004). In this case, we have to modify equation (13) of Tegmark et al. (1997), which gives the rate of change of the H_2 fraction, to account for the dust grain growth path. The molecular hydrogen fraction growth rate becomes

$$\dot{f} = k_d f(1 - x - 2f) + k_m n(1 - x - 2f)x, \quad (40)$$

where f is the fraction of H_2 by number, x is the ionization fraction of H which has total number density n ,

$$k_d = 3.025 \times 10^{-17} \frac{\xi_d}{0.01} S_{\text{H}}(T) \sqrt{\frac{T_g}{100\text{K}}} \text{cm}^3 \text{s}^{-1} \quad (41)$$

is the dust formation rate coefficient (Cazaux & Spaans 2004; equation 4) and k_m is the effective rate coefficient for H_2 formation (Tegmark et al. 1997; equation 13). We adopt the expression given by Cazaux & Spaans (2004; equation 3) for the H sticking coefficient, $S_{\text{H}}(T)$ and $\xi_d = 0.53Z$ for the dust-to-gas mass ratio as suggested by Cazaux & Spaans (2004) and which results in $\xi_d \approx 0.01$ for solar metallicity. This equation must be solved simultaneously with the recombination equation governing the ionized fraction x . The solution, assuming $x(t) = x_0/(1 + x_0 n k_1 t)$ and $1 - x - 2f \approx 1$ as do Tegmark et al. (1997), is

$$f(t) = f_0 \frac{k_m}{k_1} \exp\left[\frac{\tau_r + t}{\tau_d}\right] \left\{ \text{E}_i\left(\frac{\tau_r}{\tau_d}\right) - \text{E}_i\left(\frac{\tau_r + t}{\tau_d}\right) \right\} \quad (42)$$

where $\tau_r = 1/x_0/n_{\text{H}}/k_1$, $\tau_d = 1/n_{\text{H}}/k_d$, k_1 is the hydrogen recombination coefficient and E_i is the exponential integral.

2.7 Sizes and adiabatic contraction

The angular momentum content of galactic components is tracked within our model, allowing us to compute sizes for discs and bulges. We follow the same basic methodology as Cole et al. (2000) – simultaneously solving for the equilibrium radii of discs and bulges under the influence of the gravity of the dark matter halo and their own self-gravity and including the effects of adiabatic contraction – but treat adiabatic contraction using updated methods.

For the bulge component with pseudo-specific angular momentum j_b the half-mass radius, r_b , must satisfy

$$j_b^2 = G[M_{\text{h}}(r_b) + M_{\text{d}}(r_b) + M_{\text{b}}(r_b)]r_b, \quad (43)$$

where $M_{\text{h}}(r)$, $M_{\text{d}}(r)$ and $M_{\text{b}}(r)$ are the masses of dark matter, disc and bulge within radius r , respectively, and which we can write as

$$c_b = [M_{\text{h}}(r_b) + M_{\text{d}}(r_b) + M_{\text{b}}(r_b)]r_b, \quad (44)$$

where $c_b = j_b^2/G$. In the original Blumenthal et al. (1986) treatment of adiabatic contraction the right-hand side of equation (44) is an

adiabatically conserved quantity allowing us to write

$$c_b = M_{\text{h}}^0(r_{\text{b},0})r_{\text{b},0}, \quad (45)$$

where M_{h}^0 is the unperturbed dark matter mass profile and $r_{\text{b},0}$ the original radius in that profile. This allows us to trivially solve for $r_{\text{b},0}$ and $M_{\text{h}}^0(r_{\text{b},0})$ and so, assuming no shell crossing, $M_{\text{h}}(r_b) = f_{\text{h}}M_{\text{h}}^0(r_{\text{b},0})$, where f_{h} is the fraction of mass that remains distributed like the halo. Given a disc mass and radius this allows us to solve for r_b .

In the Gnedin et al. (2004) treatment of adiabatic contraction however, $M(r)r$ is no longer a conserved quantity. Instead, $M(\bar{r})r$ is the conserved quantity where $(\bar{r})/r_{\text{h}} = A_{\text{ac}}(r/r_{\text{h}})^{w_{\text{ac}}}$. In this case, we write

$$r_b = \langle \bar{r}'_b \rangle = A_{\text{ac}} r_{\text{h}} (r'_b/r_{\text{h}})^{w_{\text{ac}}}. \quad (46)$$

Equation (45) then becomes

$$c'_b = [M_{\text{h}}(\langle \bar{r}'_b \rangle) + M_{\text{d}}(\langle \bar{r}'_b \rangle) + M_{\text{b}}(\langle \bar{r}'_b \rangle)]r'_b, \quad (47)$$

where

$$c'_b = \frac{c_b}{A_{\text{ac}}} \left(\frac{r'_b}{r_{\text{h}}} \right)^{1-w_{\text{ac}}}. \quad (48)$$

The right-hand side of equation (47) is now an adiabatically conserved quantity and we can write

$$c'_b = M_{\text{h}}^0(\langle \bar{r}'_{\text{b},0} \rangle)r_{\text{b},0}. \quad (49)$$

If we know c'_b this expression allows us to solve for $r_{\text{b},0}$ and $M_{\text{h}}^0(\langle \bar{r}'_{\text{b},0} \rangle)$ which in turns gives $M_{\text{h}}(r_b) = f_{\text{h}}M_{\text{h}}^0(\langle \bar{r}'_{\text{b},0} \rangle)$. Of course, to find c'_b we need to know r_b . This equation must therefore be solved iteratively. In practice, for a galaxy containing a disc and bulge, the coupled disc and bulge equations must be solved iteratively in any case, so this does not significantly increase computational demand.

The disc is handled similarly. We have

$$\frac{j_{\text{d}}^2}{k_{\text{d}}^2} = G \left[M_{\text{h}}(r_{\text{d}}) + \frac{k_{\text{h}}}{2} M_{\text{d}} + M_{\text{b}}(r_{\text{d}}) \right] r_{\text{d}}, \quad (50)$$

where k_{h} gives the contribution to the rotation curve in the mid-plane and k_{d} relates the total angular momentum of the disc to the specific angular momentum at the half-mass radius (Cole et al. 2000). This becomes

$$c'_{\text{d},2} = M_{\text{h}}^0(\langle \bar{r}'_{\text{d},0} \rangle)r_{\text{d},0}, \quad (51)$$

where

$$c'_{\text{d},2} = \frac{c_{\text{d},2}}{A_{\text{ac}}} \left(\frac{r'_b}{r_{\text{h}}} \right)^{1-w_{\text{ac}}} \quad (52)$$

and

$$c_{\text{d},2} = \frac{j_{\text{d}}^2}{Gk_{\text{d}}^2} - \left(\frac{k_{\text{h}}}{2} - \frac{1}{2} \right) r_{\text{d}} M_{\text{d}}. \quad (53)$$

This system of equations must be solved simultaneously to find the radii of disc and bulge in a given galaxy. Once these are determined, the rotation curve and dark matter density as a function of radius are trivially found from the known baryonic distribution, pre-collapse dark matter density profile and the adiabatic invariance of $M(\bar{r})r$.

2.8 Substructures and merging

N -body simulations of dark matter haloes have convincingly shown that substructure persists within dark matter haloes for cosmological time-scales (Moore et al. 1999). Moreover, recent ultra-high-resolution simulations (Kuhlen et al. 2008; Springel et al. 2008;

Stadel et al. 2009) demonstrate that multiple levels of substructure (e.g. sub-sub-haloes) can exist. This ‘substructure hierarchy’ is often neglected in semi-analytic models when merging is being considered. For example, Cole et al. (2000) and all other semi-analytic models to date¹⁴ consider only one level of substructure – a substructure in a group halo which merges into a cluster immediately becomes a substructure of the cluster for the purposes of merging calculations. This is unrealistic and may:

- (i) neglect mergers between galaxies in substructures which Angulo et al. (2009) have recently shown to be important for lower mass subhaloes.
- (ii) bias the estimation of merging time-scales for haloes (and their galaxies).

Angulo et al. (2009) examine rates of subhalo–subhalo mergers in the Millennium Simulation and find that for subhaloes with masses below 0.1 per cent the mass of the main halo mergers with other subhaloes become equally likely as a merger with the central galaxy of the halo. They also find that subhalo–subhalo mergers tend to occur between subhaloes that were physically associated before falling into the larger potential. This suggests that a treatment of subhalo–subhalo mergers must consider the interactions between subhaloes and not simply consider random encounters as was done, for example, by Somerville & Primack (1999).

We therefore implement a method to handle an arbitrarily deep hierarchy of substructure. We refer to isolated haloes as S^0 substructures (i.e. not substructures at all); substructures of S^0 haloes are called S^1 substructures and substructures of S^n haloes are called S^{n+1} substructures. When a halo forms it is an S^0 substructure, and when it first becomes a satellite it becomes an S^1 substructure.

For S^n substructures with $n \geq 2$ we check at the end of each time-step whether the substructure has been tidally stripped out of its S^{n-1} host. If it has, it is promoted to being an S^{n-1} substructure in the S^{n-2} substructure which hosts its S^{n-1} host.

2.8.1 Orbital parameters

When a halo first becomes an S^1 subhalo it is assigned orbital parameters drawn from the distribution of Benson (2005) which was measured from N -body simulations. This distribution gives the radial and tangential velocity components of the orbit. For later convenience, we compute from these velocities the radius of a circular orbit with the same energy as the actual orbit, $r_C(E)$, and the circularity (the angular momentum of the actual orbit in units of the angular momentum of that circular orbit), ϵ . These are computed using the gravitational potential of the host halo.

2.8.2 Adiabatic evolution of host potential

As a subhalo orbits inside of a host halo the gravitational potential of that host halo will evolve due to continued cosmological infall. To model how this evolution affects the orbital parameters of each

subhalo we assume that it can be well described as an adiabatic process.¹⁵ As such, the azimuthal and radial actions of the orbits,

$$J_a = \frac{1}{2\pi} \int_0^{2\pi} r^2 \dot{\phi} d\phi, \quad (54)$$

and

$$J_r = \frac{1}{\pi} \int_{r_{\min}}^{r_{\max}} \dot{r} dr, \quad (55)$$

should be conserved (assuming a spherically symmetric potential). Therefore, at each time-step, we compute J_a and J_r for each satellite from the known orbital parameters in the current host halo potential. We assume these quantities are the same in the new host halo potential and convert them back into new orbital parameters $r_C(E)$ and ϵ .

2.8.3 Tidal stripping of dark matter substructures

Given orbital parameters $r_C(E)$ and ϵ we can compute the apocentric and pericentric distances of the orbit of each subhalo. At the end of each time-step, for each subhalo we find the pericentric distance and compute the tidal field of its host halo at that point:

$$\mathcal{D}_t = \frac{d}{dr_h} \left[-\frac{GM_h(r_h)}{r_h^2} \right] + \omega^2, \quad (56)$$

where ω is the orbital frequency of the subhalo, and find the radius, r_s , in the subhalo at which this equals

$$\mathcal{D}_s = \frac{GM_s(r_s)}{r_s^3}. \quad (57)$$

This gives the tidal radius, r_s , in the subhalo.

2.8.4 Promotion through the hierarchy

After computing tidal radii, for each $S^{\geq 2}$ subhalo we compute the apocentric distance of its orbit and ask if this exceeds the tidal radius of its host. If it does, the subhalo is assumed to be tidally stripped from its host halo and promoted to an orbit in the host of its host: $S^n \rightarrow S^{n-1}$. To compute orbital parameters of the satellite in this new halo we determine its radius and velocity at the point where it crosses the tidal radius of its old host. These are added vectorially (assuming random orientations) to the position and velocity of its old host at pericentre in the new host. From this new position and velocity values of $r_C(E)$ and ϵ are computed.

This approach can handle an arbitrarily deep hierarchy of substructure. In practice, the actual depth of the hierarchy will depend on both the mass resolution of the merger trees used and the efficiency of tidal forces to promote substructures through the hierarchy. Given the resolution of the trees used in our calculations we find that most substructures belong to the S^1 and S^2 levels. However, the deepest substructure level that we have found at $z = 0$ is S^7 .

¹⁴Taylor & Babul (2004), who describe a model of the orbital dynamics of subhaloes, do account for the orbital grouping of subhaloes arriving as part of a pre-existing bound system (i.e. when a halo becomes a subhalo its own subhaloes are given similar orbits in the new host). However, as noted by Taylor & Babul (2005), they do not include the self-gravity of subhaloes and so sub-subhaloes do not remain gravitationally bound to their subhalo. As such, sub-subhaloes will gradually disperse and cannot merge with each other via dynamical friction.

¹⁵Haloes are expected to grow on the Hubble time, while the characteristic orbital time is shorter than this by a factor of $\sqrt{\Delta}$ where Δ is the overdensity of dark matter haloes. This expected validity of the adiabatic approximation has been confirmed in N -body simulations by Book et al. (in preparation).

2.8.5 Dynamical friction

We adopt the fitting formula found by Jiang et al. (2008) to estimate merging time-scales for dark matter substructures (and, consequently, the galaxies that they contain). The multiple levels of substructure hierarchy in our model allow for the possibility of satellite–satellite mergers. We intend to compare results from our model with N -body measures of this process in a future work.

When a halo first becomes a satellite, we set a dimensionless merger clock, $x_{\text{DF}} = 0$. On each subsequent time-step, x_{DF} is incremented by an amount $\Delta t/\tau_{\text{DF}}$ where τ_{DF} is the dynamical friction time-scale for the satellite in the current host halo according to the expression of Jiang et al. (2008), including the dependence on $r_{\text{C}}(E)$. When $x_{\text{DF}} = 1$ the satellite is deemed to have merged with the central galaxy in the host halo.

When a satellite is tidally stripped out of its current orbital host and promoted to the host above it in the hierarchy the merging clock is reset so that dynamical friction calculations start anew in this new orbital host. This is something of an approximation since the dynamical friction time-scale of Jiang et al. (2008) is calibrated using satellites that enter their halo at the virial radius. As such, it does not explore as a sufficiently wide range in r_{C} as is required for our models. Furthermore, when promoted to a new orbital host, a satellite will have already lost some mass due to tidal effects. This is not accounted for when computing a new dynamical friction time-scale however, and so may cause us to underestimate merging time-scales somewhat.

Dynamical friction also affects the orbital parameters of each subhalo. To simplify matters we follow Lacey & Cole (1993) and examine the evolution of these quantities in an isothermal dark matter halo. In such a halo, and for a circular orbit, r_{C} evolves as

$$\left(\frac{r_{\text{C}}}{r_{\text{C},0}}\right)^2 = 1 - \frac{t}{\tau_{\text{DF}}}. \quad (58)$$

Therefore, after each time-step we update

$$r_{\text{C}}^2 \rightarrow r_{\text{C}}^2 - r_{\text{C},0}^2 \frac{\Delta t}{\tau_{\text{DF}}}. \quad (59)$$

The fractional change in ϵ is assumed to be given by $(\dot{\epsilon}/\epsilon)/(\dot{r}_{\text{C}}/r_{\text{C}})$ as computed for the current orbit using the expressions of Lacey & Cole (1993). This is a function of ϵ only and is plotted in Fig. 4. Note that the time-scale, τ_{DF} , used here is that from Jiang et al. (2008) and not the one from Lacey & Cole (1993).

2.9 Ram-pressure and tidal stripping

We follow Font et al. (2008) and estimate the extent to which ram pressure from the hot atmosphere of a halo may strip away the hot atmosphere of an orbiting subhalo. In addition, we also consider tidal stripping of this hot gas and both ram-pressure and tidal stripping of material from galaxies.

Ram-pressure and tidal forces are computed at the pericentre of each subhalo's orbit, which we now compute self-consistently with our orbital model (see Section 2.8). For an S^i , where $i > 1$, subhalo we compute the ram-pressure force from all haloes higher in the hierarchy and take the maximum of these to be the ram-pressure force actually felt. The tidal field (i.e. the gradient in the gravitational force across the satellite) includes the centrifugal contribution at the orbital pericentre and is given by

$$\mathcal{F} = \omega^2 - \frac{d}{dR} \frac{GM(< r)}{r^2}. \quad (60)$$

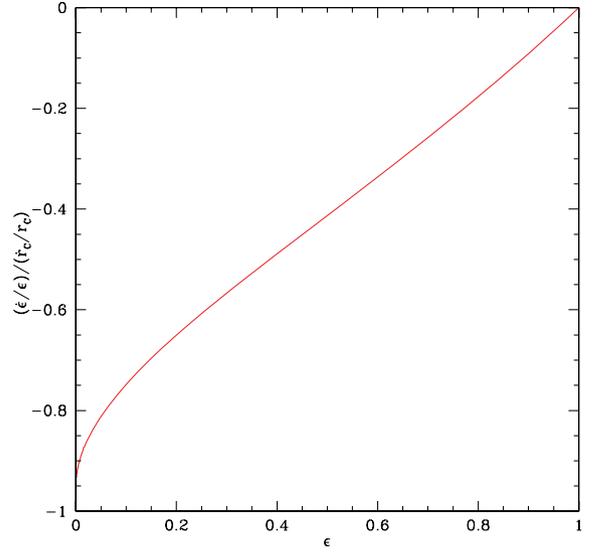


Figure 4. The ratio $(\dot{\epsilon}/\epsilon)/(\dot{r}_{\text{C}}/r_{\text{C}})$ for isothermal haloes. This ratio is used in solving for the evolution of orbital circularity and orbital radius under the influence of dynamical friction as described in Section 2.8.5.

The ram pressure is taken to be

$$P_{\text{ram}} = \rho_{\text{hot,host}} V_{\text{orbit}}^2 \quad (61)$$

where $\rho_{\text{hot,host}}$ is the density of hot gas in the host halo at the pericentre of the orbit and V_{orbit} is the orbital velocity of the satellite at that position.

2.9.1 Stripping of hot halo gas

We find the ram-pressure radius in the hot halo gas by solving

$$P_{\text{ram}} = \alpha_{\text{ram}} \frac{GM_{\text{sat}}(r_{\text{r}})}{r_{\text{r}}} \rho_{\text{hot,sat}}(r_{\text{r}}) \quad (62)$$

for r_{r} , where α_{ram} is a parameter that we set equal to 2 as suggested by McCarthy et al. (2008). Similarly, a tidal radius is found by solving

$$\mathcal{F} = \alpha_{\text{tidal}}^3 \frac{GM_{\text{sat}}(r_{\text{t}})}{r_{\text{t}}^3} \quad (63)$$

for r_{t} , where α_{tidal} is a parameter that we set equal to unity. Once the minimum of the ram-pressure and tidal stripping radii has been determined we follow Font et al. (2008) and compute the cooling rate of the remaining, unstripped gas by cooling only the gas within the stripping radius and assuming that stripping does not alter the mean density of gas within this radius. We implement this by giving the satellite a nominal hot gas mass $M'_{\text{hot}} = M_{\text{hot}} + M_{\text{strip}}$ (where M_{hot} is the true hot gas content of the halo) and applying the same cooling algorithm as that used for central galaxies (except limiting the maximum cooling radius to r_{strip} rather than R_{v}). This step ensures self-consistency in the treatment of the gas cooling between stripped and unstripped galaxies, and therefore that the colours of satellites are predicted correctly.

The initial stripping of re-heated gas is the same as for the hot gas, i.e. the same fraction is transferred from the re-heated gas of the satellite to the re-heated gas reservoir of the parent halo. We follow Font et al. (2008) in modelling the time-dependence of the hot gas mass in the satellite halo and refer the reader to that paper for full details. This process introduces one free parameter, ϵ_{strip} which represents the time-averaged stripping rate after the initial

pericentre. We treat ϵ_{strip} as a free parameter which we will adjust to match observational constraints.

The stripping of satellites is also affected by the growth of the halo in which the satellite is orbiting. Font et al. (2008) took this effect into account by assigning each satellite galaxy new orbital parameters and deriving a new stripping factor every time the halo doubles in mass compared to the initial stripping event. In the present work we directly follow the evolution of the pericentric radius and velocity of each satellite due to both dynamical friction and host halo mass growth. For this reason, we take a different approach from Font et al. (2008), computing a new ram-pressure radius in each time-step instead of only at every mass doubling event.

Any material stripped away from the subhalo is added to the halo which provided the greatest ram-pressure force. For tidal forces, we consider only the contribution from the current orbital host as typically if this were exceeded by the tidal force from a parent higher up in the hierarchy the subhalo would have already been tidally stripped from this orbital host and promoted to a higher level in the hierarchy.

2.9.2 Stripping of galactic gas and stars

The effective gravitational pressure that resists the ram-pressure force in the disc plane is (for an exponential disc; Abadi, Moore & Bower 1999)

$$P_{\text{grav}} = \frac{GM_d M_g}{4\pi r_d^4} x e^{-x} \left[I_0\left(\frac{x}{2}\right) K_1\left(\frac{x}{2}\right) - I_1\left(\frac{x}{2}\right) K_0\left(\frac{x}{2}\right) \right], \quad (64)$$

where $x = r/r_d$ and I_0 , I_1 , K_0 and K_1 are Bessel functions. The ram-pressure radius is found by solving for the radius at which $P_{\text{grav}} = P_{\text{ram}}$, where P_{ram} is given by equation (61). We assume that any stars in the galaxy which lie beyond the computed tidal radius and any gas which lies beyond the smaller of the tidal and ram-pressure radii are instantaneously removed. Stars become part of the diffuse light component of the halo (i.e. that which is known as intracluster light in clusters of galaxies; see Section 4.12.2), while gas is added to the reheated reservoir of the host halo. The remaining mass of each component (cold gas, disc and bulge stars) is computed and the specific angular momentum of the remaining material is computed assuming a flat rotation curve:

$$j_{\text{disc}} = j_{\text{disc}0} \times \left[\frac{\int_0^{R_*} \Sigma_*(R) R^2 dR + \int_0^{R_g} \Sigma_g(R) R^2 dR}{\int_0^\infty \Sigma(R) R^2 dR} \right] \times \left[\frac{\int_0^{R_*} \Sigma_*(R) R dR + \int_0^{R_g} \Sigma_g(R) R dR}{\int_0^\infty \Sigma(R) R dR} \right]^{-1} \quad (65)$$

$$= j_{\text{disc}0} \times \left\{ f_* \left[1 - \left(1 + x_* + \frac{x_*^2}{2} \right) e^{-x_*} \right] + f_g \left[1 - \left(1 + x_g + \frac{x_g^2}{2} \right) e^{-x_g} \right] \right\} \times \left\{ f_* [1 - (1 + x_*)e^{-x_*}] + f_g [1 - (1 + x_g)e^{-x_g}] \right\}^{-1} \quad (66)$$

for the disc (the last line assuming an exponential disc) where $R_* = r_{\text{tidal}}$, $R_g = \min(r_{\text{tidal}}, r_{\text{ram}})$, $x_* = R_*/R_d$, $x_g = R_g/R_d$, $f_* = M_*/(M_* +$

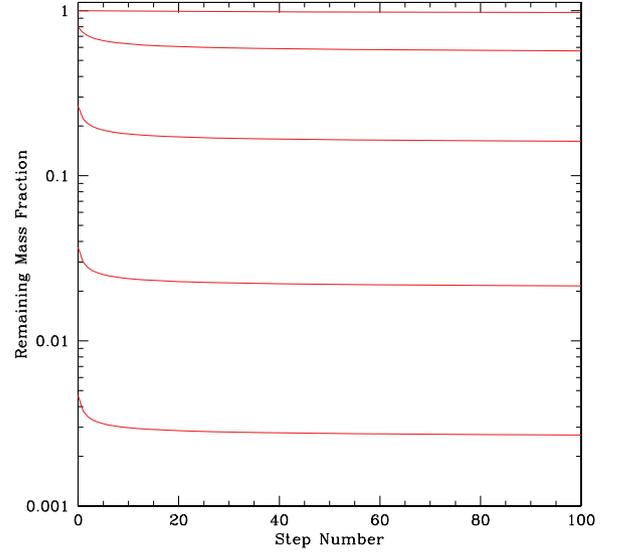


Figure 5. The remaining mass fraction in an exponential disc in a potential giving a flat rotation curve (and ignoring the disc self-gravity) subjected to tidal truncation at radius $r_t/r_{d,0} = 0.1, 0.3, 1.0, 3.0$ and 10.0 (from lower to upper lines) after a given number of steps according to our model. The remaining mass fraction quickly converges to a near constant value.

M_g) and $f_g = M_g/(M_* + M_g)$, and

$$j_{\text{sph}} = j_{\text{sph}0} \frac{\int_0^{r_{\text{tidal}}} \rho_*(R) R^3 dR / \int_0^\infty \rho_*(R) R^3 dR}{\int_0^{r_{\text{tidal}}} \rho_*(R) R^2 dR / \int_0^\infty \rho_*(R) R^2 dR} \quad (67)$$

for the bulge (and which must be evaluated numerically). Here, $j_{\text{disc}0}$ and $j_{\text{sph}0}$ are the pre-stripping specific angular momenta of disc and spheroid, respectively, $\Sigma_*(R)$ and Σ_{gas} are the surface density profiles of stars and gas in the disc prior to stripping and $\rho_*(R)$ is the stellar density profile in the spheroid prior to stripping. Since GALFORM always assumes a de Vaucouler's spheroid and an exponential disc with stars tracing gas the stripped components will readjust to these configurations with their new masses and angular momenta. This is, therefore, an approximate treatment of stripping. In particular, some material will always 'leak' back out beyond the stripping radius and so is easily stripped on the next time-step. Fig. 5 demonstrates that this is not a severe problem, with the remaining mass fraction asymptoting to a near constant value after just a few steps.

2.10 IGM interaction

Benson et al. (2002b) introduced methods to simultaneously compute the evolution of the IGM and the galaxy population in a self-consistent manner such that emission from galaxies ionized and heated the IGM which in turn lead to suppression of future galaxy formation. A major practical limitation of Benson et al.'s (2002b) method was that it required GALFORM to be run to generate an emissivity history for the Universe which was then fed into a model for the IGM evolution. The IGM evolution was used to predict the effects on galaxy formation and GALFORM run again. This loop was iterated around several times to find a converged solution. This problem was inherent in the implementation due to the fact that GALFORM was designed to evolve a single merger tree to $z = 0$ and then move on to the next one.

To circumvent this problem, we have adapted GALFORM to allow for multiple merger trees to be evolved simultaneously: each tree is evolved for a single time-step after which the IGM evolution for

that same time-step is computed. This allows simultaneous, self-consistent evolution of the IGM and galaxies without the need for iteration.

The model we adopt for the IGM evolution is essentially identical to that of Benson et al. (2002b) and consists of a uniform IGM (with a clumping factor to account for enhanced recombination and cooling due to inhomogeneities) composed of hydrogen and helium and a photon background supplied by galaxies and AGN. The reader is therefore referred to Benson et al. (2002b) for a full discussion. Here, we will discuss only those aspects that are new or updated.

2.10.1 Emissivity

The two sources of photons in our model are quasars and galaxies. For AGN we assume that the spectral energy distribution (SED) has the following shape (Haardt & Madau 1996):

$$f_\nu(\lambda) \propto \begin{cases} \lambda^{1.5} & \text{if } \lambda < 1216 \text{ \AA}; \\ \lambda^{0.8} & \text{if } 1216 \text{ \AA} < \lambda < 2500 \text{ \AA}; \\ \lambda^{0.3} & \text{if } \lambda > 2500 \text{ \AA}, \end{cases} \quad (68)$$

where the normalization of each segment is chosen to give a continuous function and unit energy when integrated over all wavelengths. The emissivity per unit volume from AGN is then

$$\epsilon_{\text{AGN}} = f_{\text{esc,AGN}} \epsilon_\bullet \dot{\rho}_\bullet c^2 f_\nu(\lambda), \quad (69)$$

where $\epsilon_\bullet = 0.1$ is an assumed radiative efficiency for accretion on to black holes, $\dot{\rho}_\bullet$ is the rate of black hole mass growth per unit volume computed by GALFORM and $f_{\text{esc,AGN}}$ is an assumed escape fraction for AGN photons which we fix at 10^{-2} to produce a reasonable epoch of He II reionization.

The emissivity from galaxies was calculated directly by integrating the star formation rate per unit volume predicted by GALFORM over time and metallicity to give

$$\epsilon_{\text{gal}} = \int_0^{t_{\text{now}}} f_{\text{esc,gal}}(t') \dot{M}_*(t', Z) L_\nu(t_{\text{now}} - t', Z[t']) dt', \quad (70)$$

where $\dot{M}_*(t, Z)$ is the rate of star formation at metallicity Z , $L_\nu(t, Z)$ is the integrated luminosity per unit frequency and per solar mass of stars formed of a single stellar population of age t and metallicity Z and $f_{\text{esc,gal}}$ is the escape fraction of ionizing photons from the galaxy.

The fraction of ionizing photons able to escape from the disc of each galaxy is computed using the expressions derived by Benson et al. (2002a) (their equation A4) which is a generalization of the model of Dove & Shull (1994) in which OB associations with a distribution of luminosities ionize holes through the neutral hydrogen distribution through which their photons can escape.

The sum of ϵ_{AGN} and ϵ_{gal} gives the number of photons emitted from the galaxies and quasars in the model.

2.10.2 IGM ionization state

The ionization state of the IGM is computed just as in Benson et al. (2002b) except that we use effective photo-ionization cross-sections that account for the effects of secondary ionizations and are given by Shull & van Steenberg (1985; as re-expressed by Venkatesan, Giroux & Shull 2001):

$$\sigma'_H(E) = \left(1 + \phi_{\text{H I}} \frac{E - E_H}{E_H} + \phi_{\text{He I}}^* \frac{E - E_{\text{He}}}{19.95 \text{ eV}} \right) \sigma_H(E) + \left(1 + \phi_{\text{He I}} \frac{E - E_{\text{He}}}{E_{\text{He}}} \right) \sigma_{\text{He}}(E) \quad (71)$$

$$\sigma'_{\text{He}}(E) = \left(1 + \phi_{\text{He I}} \frac{E - E_{\text{He}}}{E_{\text{He}}} \right) \sigma_{\text{He}}(E) + \left(\phi_{\text{He I}} \frac{E - E_H}{24.6} \right) \sigma_H(E) \quad (72)$$

where $\sigma(E)$ is the actual cross-section (Verner & Yakovlev 1995) and

$$\phi_{\text{H I}} = 0.3908 (1 - x_e^{0.4092})^{1.7592}, \quad (73)$$

$$\phi_{\text{He I}}^* = 0.0246 (1 - x_e^{0.4049})^{1.6594}, \quad (74)$$

$$\phi_{\text{He I}} = 0.0554 (1 - x_e^{0.4614})^{1.6660}. \quad (75)$$

2.10.3 IGM thermal state

Heating of the IGM is treated as in Benson et al. (2002b) with the exception that we account for heating by secondary electrons. Photoionization heats the IGM at a rate of

$$\Sigma_{\text{photo}} = \int_0^\infty (E - E_i) c \sigma'(E) n_i n_\gamma(E) \mathcal{E} dE, \quad (76)$$

where E_i is the energy of the sampled photons which is associated with atom/ion number density n_i , c is the speed of light, σ' is the effective partial photo-ionization cross-section (accounting for secondary ionizations) for the ionization stages of H and He, $n_\gamma(E)$ is the number density of photons of energy E , E_i is the ionization potential of i and index i represents the different atoms and ions, H, H⁺, He, He⁺ and He²⁺. In the above, \mathcal{E} accounts for heating by secondary electrons and is given by (Shull & van Steenberg 1985)

$$\mathcal{E} = 0.9971 \left[1 - (1 - x_e^{0.2663})^{1.3163} \right]. \quad (77)$$

2.10.4 Suppression of baryonic infall into haloes

According to Okamoto, Gao & Theuns (2008), the mass of baryons which accrete from the IGM into a halo after reionization is given by

$$M_b = M'_b + M_{\text{acc}}, \quad (78)$$

where

$$M'_b = \sum_{\text{prog}} \exp\left(-\frac{\delta t}{t_{\text{evp}}}\right) M_b, \quad (79)$$

and where the sum is taken over the progenitor haloes of the current halo, δt is the time since the previous time-step and t_{evp} is the time-scale for gas to evaporate from the progenitor halo and is given by

$$t_{\text{evp}} = \begin{cases} R_H/c_s(\Delta_{\text{evp}}) & \text{if } T_v < T_{\text{evp}}, \\ \infty & \text{if } T_v > T_{\text{evp}}. \end{cases} \quad (80)$$

Here, T_{evp} is the temperature below which gas will be heated and evaporated from the halo. We follow Okamoto et al. (2008) and compute T_{evp} by finding the equilibrium temperature of gas at an overdensity of $\Delta_{\text{evp}} = 10^6$. The accreted mass M_{acc} is given by

$$M_{\text{acc}} = \begin{cases} \frac{\Omega_b}{\Omega_0} M_v - M'_b & \text{if } T_{\text{vir}} > T_{\text{acc}} \\ 0 & \text{if } T_{\text{vir}} < T_{\text{acc}} \end{cases} \quad (81)$$

where T_{acc} is the larger of the temperature of IGM gas adiabatically compressed to the density of accreting gas and the equilibrium

temperature, T_{eq} , at which radiative cooling balances photoheating for gas at the density expected at the virial radius. This ensures that a sensible temperature is used even when the photoionizing background is essentially zero.

The value of T_{acc} is computed at each time-step by searching for where the cooling function (see Section 2.6.7) crosses zero for the density of gas just accreting at the virial radius (for which we use one-third of the halo overdensity; Okamoto et al. 2008).

2.11 Recycling and chemical evolution

In Cole et al. (2000), the instantaneous recycling approximation for chemical enrichment was used. While this is a reasonable approximation for $z = 0$, it fails for high redshifts (where the main sequence lifetimes of the stars which do the majority of the enrichment become comparable to the age of the Universe). It also prevents predictions for abundance ratios (e.g. $[\alpha/\text{Fe}]$) from being made and ignores any metallicity dependence in the yield.

Nagashima et al. (2005a; see also Nagashima et al. 2005b, Arrighi et al. 2010) previously implemented a non-instantaneous recycling calculation in GALFORM. We implement a similar model here, following their general approach, but with some specific differences.

The fraction of material returned to the ISM by a stellar population as a function of time is given by

$$R(t) = \int_{M(t;Z)}^{\infty} [M - M_r(M; Z)] \phi(M) \frac{dM}{M}, \quad (82)$$

where $\phi(M)$ is the initial mass function (IMF) normalized to unit stellar mass, $M_r(M)$ is the remnant mass of a star of initial mass M . Here, $M(t)$ is the mass of a star with lifetime t . Similarly, the yield of element i is given by

$$p_i(t) = \int_{M(t;Z)}^{\infty} M_i(M_0; Z) \phi(M_0) \frac{dM_0}{M_0}, \quad (83)$$

where $M_i(M_0; Z)$ is the mass of metals produced by stars of initial mass M_0 . For a specified IMF we compute $R(t; Z)$ and $y_i(t; Z)$ for all times and elements of interest. This means that, unlike most previous implementations of GALFORM, the recycled fraction and yield are not free parameters of the model, but are fixed once an IMF is chosen. However, it should be noted that significant uncertainties remain in calculations of stellar yields, which may therefore influence our calculations. Note that, unlike Nagashima et al. (2005a), we include the full metallicity dependence in these functions. Stellar data are taken from Portinari, Chiosi & Bressan (1998) for low- and intermediate-mass stars and from Marigo (2001) for high-mass stars.

In GALFORM the evolution of gas and stellar masses in a galaxy is controlled by the following equations:¹⁶

$$\dot{M}_* = \frac{M_{\text{gas}}}{\tau_*} - \dot{M}_R \quad (84)$$

$$\dot{M}_{\text{gas}} = -(1 + \beta') \frac{M_{\text{gas}}}{\tau_*} + \dot{M}_R + \dot{M}_{\text{infall}}. \quad (85)$$

where

$$\tau_* = \begin{cases} \epsilon_*^{-1} \tau_{\text{disc}} \left(\frac{V_{\text{disc}}}{200 \text{ km s}^{-1}} \right)^{\alpha_*} & \text{for discs} \\ f_{\text{dyn}} \tau_{\text{bulge}} & \text{for bursts} \end{cases} \quad (86)$$

¹⁶These are identical to those given in Cole et al. (2000; their equations 4.6 and 4.8) except for the explicit inclusion of the recycling terms – Cole et al. (2000) included these using the instantaneous recycling approximation.

is the star formation time-scale, τ_{disc} is the dynamical time at the disc half-mass radius, τ_{bulge} is the dynamical time at the bulge half-mass radius, $f_{\text{dyn}} = 2$ and β' quantifies the strength of supernova feedback (see Section 2.12). In Cole et al. (2000), the instantaneous recycling approximation implies that $\dot{M}_R \propto M_{\text{gas}}/\tau_*$, and the cosmological infall term \dot{M}_{infall} is approximated as being constant over each short time-step. This permits a simple solution to these equations. In our case, we retain the assumption of constant \dot{M}_{infall} and further assume that the mass recycling rate, \dot{M}_R , can be approximated as being constant throughout the time-step.¹⁷ We therefore write

$$\dot{M}_R = \frac{M_{R,\text{past}} + M_{R,\text{now}}}{\Delta t}, \quad (87)$$

where Δt is the time-step,

$$M_{R,\text{past}} = \int_{t_0}^{t_0+\Delta t} dt'' \int_0^{t_0} dt' \dot{M}_*(t') \dot{R}(t'' - t') \quad (88)$$

is the mass of gas returned to the ISM from populations of stars formed in previous time-steps (and is trivially computed from the known star formation rate of the galaxy on past time-steps) and

$$M_{R,\text{now}} = \int_{t'}^{t_0+\Delta t} dt'' \int_{t_0}^{t_0+\Delta t} dt' \dot{M}_*(t') \dot{R}(t'' - t'), \quad (89)$$

is the mass returned to the ISM by star formation during the current time-step. With these approximations, the gas equations always have the solution

$$M_{\text{gas}}(t) = M_{\text{gas}0} \exp\left(-\frac{t}{\tau_{\text{eff}}}\right) + \dot{M}_{\text{input}} \tau_{\text{eff}} \left[1 - \exp\left(-\frac{t}{\tau_{\text{eff}}}\right) \right], \quad (90)$$

where $M_{\text{gas}0}$ is the mass of gas at time $t = 0$ (measured from the start of the time-step and

$$\begin{aligned} \dot{M}_{\text{input}} = & \dot{M}_{\text{infall}} \\ & + \left\{ \left[\frac{M_{\text{gas}0}}{\tau_{\text{eff}}} - \frac{M_{R,\text{past}}}{\Delta t} \right] I_{R1}(\Delta t, \tau_{\text{eff}}) \right. \\ & \left. + \frac{M_{R,\text{past}}}{\Delta t} I_{R0}(\Delta t) \right\} \\ & \times \{ (1 + \beta) + [I_{R1}(\Delta t, \tau_{\text{eff}}) - I_{R0}(\Delta t)] / \Delta t \}^{-1} \end{aligned} \quad (91)$$

where

$$I_{R0}(t) = \int_0^t R(t - t') dt', \quad (92)$$

$$I_{R1}(t, \tau) = \int_0^t \exp(-t'/\tau) R(t - t') dt'. \quad (93)$$

In the above equation, the effective e-folding time-scale for star formation (accounting for SNe driven outflows), τ_{eff} , is given by

$$\tau_{\text{eff}} = \frac{\tau_*}{1 + \beta'}, \quad (94)$$

where β' measures the strength of supernovae (SNe) feedback and is defined in equation (102).

The evolution of the metal mass is treated in a similar way, assuming a constant rate of input of metals from infall, star formation from previous time-steps and star formation from the current time-step. Metals in the cold gas reservoir of a galaxy are assumed to be uniformly mixed into the gas, such that the reservoir has a uniform

¹⁷This will be approximately true if the time-step is sufficiently short that $\dot{R}\Delta t \ll \dot{R}$.

metallicity. Metals then flow from the cold gas reservoir into the stellar phase and out into the reheated reservoir at a rate proportional to the star formation rate and mass outflow rate, respectively, with the constant of proportionality being the cold gas metallicity. Material recycled from stars to the cold phase carries with it metals corresponding to the original metallicity of those stars, augmented by the appropriate metal yield. Finally, gas infalling from the surrounding halo may have been enriched in metals by previous galaxy formation and so deposits metals into the cold phase gas at a rate proportional to the mass infall rate, with proportionality equal to the (assumed uniform) metallicity of the notional profile gas. Apart from the fact that metals from stellar recycling and yields are not added instantaneously to the cold reservoir this treatment of metals remains identical to that of Cole et al. (2000). The net rate of metal mass input to the cold phase (from both cosmological infall and returned from stars) is

$$\begin{aligned} \dot{M}_{Z_i \text{ input}} = \dot{M}_{Z_i \text{ infall}} &+ \frac{\left[\frac{M_{Z_i \text{ gas0}}}{\tau_{\text{eff}}} - \frac{M_{Z_i \text{ R}}^{\text{past}}}{\Delta t} \right] I_{\text{R1}}(\Delta t, \tau_{\text{eff}}) + \frac{M_{Z_i \text{ R}}^{\text{past}}}{\Delta t} I_{\text{R0}}(\Delta t)}{\Delta t [(1 + \beta) + (I_{\text{R1}}(\Delta t, \tau_{\text{eff}}) - I_{\text{R0}}(\Delta t)) \Delta t]} \\ &+ \frac{\left[\frac{M_{\text{gas0}}}{\tau_{\text{eff}}} - \frac{M_{\text{R}}^{\text{past}}}{\Delta t} \right] I_{\text{p1}}(\Delta t, \tau_{\text{eff}}) + \frac{M_{\text{R}}^{\text{past}}}{\Delta t} I_{\text{p0}}(\Delta t)}{\Delta t [(1 + \beta) + (I_{\text{p1}}(\Delta t, \tau_{\text{eff}}) - I_{\text{p0}}(\Delta t)) \Delta t]}, \end{aligned} \quad (95)$$

where $M_{Z_i, \text{past}}$ is the mass of metal i recycled from star formation in previous time-steps and

$$I_{\text{p0}}(t) = \int_0^t p(t-t') dt', \quad (96)$$

$$I_{\text{p1}}(t, \tau) = \int_0^t \exp(-t'/\tau) p(t-t') dt'. \quad (97)$$

2.11.1 Star bursts

In previous implementations of GALFORM star bursts were assumed to have an exponentially declining star formation rate. Such a rate results from assuming an instantaneous star formation rate of

$$\dot{M}_* = \frac{M_{\text{cold}}}{\tau_*}, \quad (98)$$

where τ_* is a star formation time-scale (fixed throughout the duration of the burst), an outflow rate proportional to the star formation rate and a rate of recycling given by $R\dot{M}_*$. The resulting differential equations have a solution with an exponentially declining star formation rate.

When the instantaneous recycling approximation is dropped the rate of recycling is no longer proportional to the star formation rate and the differential equations no longer have an exponential solution. We choose to retain the original star formation law (equation 98) and solve the differential equations to determine the star formation rate, outflow rate, etc. as a function of time in the burst. The resulting set of equations have solutions identical to those in Section 2.11 but with zero cosmological infall terms. Recycled material and the effects of feedback (see Section 2.12) are applied to the gas in the burst during the lifetime of the burst. Any recycling and feedback occurring after the burst is finished are applied to the disc.

In Cole et al. (2000) while bursts were treated as having finite duration for the purposes of computing the luminosity of their stellar populations at some later time, the change in the mass of the galaxy

due to the burst occurred instantaneously. We drop this approximation and correctly follow the change in mass of each component (gas, stars, outflow) during each time-step.

2.12 Feedback

Feedback from SNe is also modified to account for the delay between star formation and supernova. In Cole et al. (2000), the outflow rate due to SNe feedback was

$$\dot{M}_{\text{out}} = \beta \dot{M}_*, \quad (99)$$

where

$$\beta = \left(\frac{V_{\text{hot}}}{V_{\text{galaxy}}} \right)^{\alpha_{\text{hot}}}, \quad (100)$$

V_{hot} and α_{hot} are parameters of the model (we allow for two different values of V_{hot} , one for quiescent star formation in discs and one for bursts of star formation) and V_{galaxy} is the circular velocity at the half-mass radius of the galaxy, determines the strength of feedback and is a function of the depth of the galaxy's potential well. We modify this to

$$\dot{M}_{\text{out}} = \beta' \dot{M}_*, \quad (101)$$

where

$$\beta' = \beta \frac{\int_0^t \dot{\phi}_*(t') \dot{N}_{\text{SNe}}(t-t') dt'}{\dot{\phi}_*(t) N_{\text{SNe}}^{(\text{II})}(\infty)} \quad (102)$$

where $N_{\text{SNe}}(t)$ is the total number of SNe (of all types) arising from a single population of stars after time t , such that the outflow rate scales in proportion to the current rate of SNe but produces the same net mass ejection after infinite time (for constant β). In fact, we compute β using the present properties of the galaxy at each time-step. The qualifier '(II)' appearing in the quantity $N_{\text{SNe}}^{(\text{II})}(\infty)$ in the denominator of equation (102) indicates that we normalize the outflow rate by reference to the number of SNe from our adopted Population II IMF (see Section 2.14). This results in the outflow correctly encapsulating any differences in the effective number of SNe between Population II and III stars. For SNe rates, we assume that all stars with initial masses greater than $8 M_{\odot}$ will result in a Type II supernova allowing the rate to be found from the lifetimes of these stars and the adopted IMF. We adopt the calculations of Nagashima et al. (2005a) to compute the Type Ia SNe rate.

Since β' appears in the gas equations of Section 2.11 but also depends on the star formation rate during the current time-step we must iteratively seek a solution for β' which is self-consistent with the star formation rate. We find that a simple iterative procedure, with an initial guess of $\beta' = \beta$ quickly converges.

When gas is driven out of a galaxy in this way, it can be either reincorporated into the M_{reheated} reservoir in the notional hot gas profile of the current halo, or it can be expelled from the halo altogether and allowed to reaccrete only further up the hierarchy once the potential well has become deeper.

We assume that the expelled fraction is given by

$$f_{\text{exp}} = \exp\left(-\frac{\lambda_{\phi} V^2}{\langle e \rangle}\right), \quad (103)$$

such that the rate of mass input to the reheated reservoir is

$$\dot{M}_{\text{reheated}} = (1 - f_{\text{exp}}) \beta' \dot{M}_*. \quad (104)$$

Here, λ_{ϕ} is a dimensionless parameter relating the depth of the potential well to V^2 (we set $\lambda_{\phi} = 1$ always), V is the circular velocity of the galaxy disc or bulge (for quiescent or bursting star

formation, respectively) and $\langle e \rangle$ is the mean energy per unit mass of the outflowing material. We further assume

$$\langle e \rangle = \frac{1}{2} \lambda_{\text{expel}} V^2, \quad (105)$$

where λ_{expel} is a parameter of the order of unity relating the energy of the outflowing gas to the potential of the host galaxy, and will be treated as a free parameter to be constrained from observations (we actually allow for λ_{expel} to have different values for quiescent and bursting star formation; see Section 3). We then proceed to the parent halo and allow a fraction

$$f_{\text{acc}} = \exp\left(-\frac{V_{\text{max}}^2}{\langle e \rangle}\right) - \exp\left(-\frac{V_v^2}{\langle e \rangle}\right) \quad (106)$$

to be reaccreted into the hot gas reservoir of the notional profile, where V_{max} is the maximum of $\sqrt{\lambda_{\text{expel}}}V$ and any parent halo V_v yet found. We then proceed to the parent's parent and repeat the accretion procedure, continuing until the base of the tree is reached. In this way, all of the gas will be reaccreted if the potential well becomes sufficiently deep.

2.13 AGN feedback

In recent years, the possibility that feedback from AGN plays a significant role in shaping the properties of a forming galaxy has come to the forefront (Croton et al. 2006; Bower et al. 2006; Somerville et al. 2008b). We adopt the black hole growth model of Malbon et al. (2007) and the AGN feedback model of Bower et al. (2006) as modified by Bower et al. (2008). The reader is referred to those papers for a full description of our implementation of AGN feedback.

2.14 Stellar populations

We consider both Pop II and Pop III stars. To compute luminosities of Population II stellar populations we employ the most recent version¹⁸ of the Conroy, Gunn & White spectral synthesis library (Conroy, Gunn & White 2009).¹⁹ We adopt a Chabrier IMF (Chabrier 2003)

$$\phi(M) \propto \begin{cases} \exp\left(-\frac{1}{2} \frac{[\log_{10} M/M_c]^2}{\sigma^2}\right) & \text{for } M \leq 1 M_{\odot} \\ M^{-\alpha} & \text{for } M > 1 M_{\odot}, \end{cases} \quad (107)$$

where $M_c = 0.08 M_{\odot}$ and $\sigma = 0.69$ and the two expressions are forced to coincide at $1 M_{\odot}$. Recycled mass fractions, yield and SNe rates are computed self-consistently from this IMF as described in Sections 2.11 and 2.12 and are shown in Fig. 6.

For Population III stars (which we assume form below a critical metallicity of $Z_{\text{crit}} = 10^{-4} Z_{\odot}$) we adopt IMF 'A' from Tumlinson (2006). Spectral energy distributions for this IMF as a function of population age were kindly provided by J. Tumlinson. Lifetimes for these stars are taken from the tabulation given by Tumlinson, Shull & Venkatesan (2003). Recycled fractions and yields and energies from pair instability SNe are computed using the data given by Heger & Woosley (2002). Recycled mass fractions, yield and SNe rates are computed self-consistently from these Population III stars as shown in Fig. 6 by green lines.

¹⁸Specifically, v2.0 downloaded from <http://www.astro.princeton.edu/~conroy/SPS/> with bug fixes up to 2010 January 7.

¹⁹For calculations of IGM evolution we *do not* use the Conroy et al. (2009) spectra because they assign stars hotter than 5×10^4 K pure blackbody spectra. This leads to an unrealistically large ionizing flux for young, metal-rich populations. We therefore instead use the Bruzual & Charlot (2003) spectral synthesis library for IGM evolution calculations.

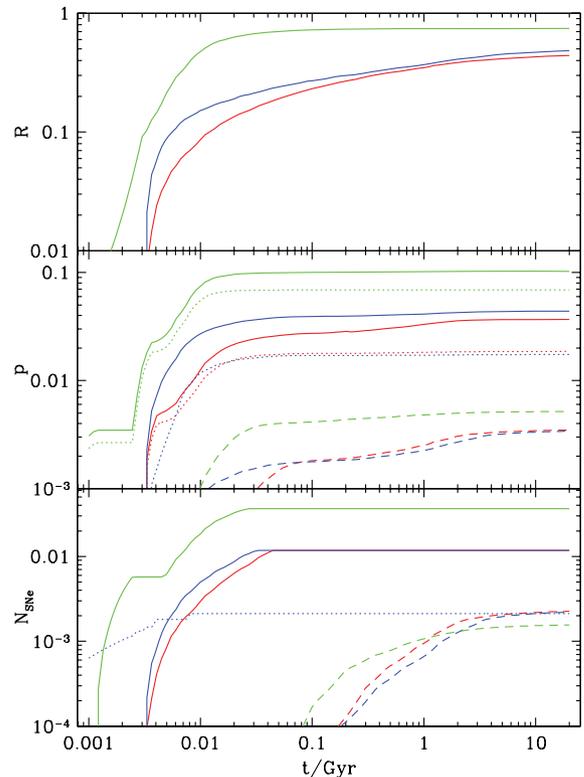


Figure 6. Upper, middle and lower panels show the recycled fraction, yield and effective number of SNe, respectively, for a Chabrier IMF (two metallicities, defined as the mass fraction of heavy elements, are shown: 0.0001 as red lines and 0.0501 as blue lines) and for metal-free Population III stars with type 'A' IMF from Tumlinson (2006) (green lines). Top panel: the fraction of mass from a single stellar population, born at time $t = 0$, recycled to the ISM after time t . Middle panel: the total metal yield from a single stellar population born at time, $t = 0$, after time t is shown by the solid lines. Dotted and dashed lines show the yield of oxygen and iron, respectively. Lower panel: cumulative energy input into the ISM, expressed as the number of equivalent SNe, per unit mass of stars formed as a function of time. The dotted line indicates the contribution from stellar winds, the solid line the contribution from Type II SNe and the dashed line the contribution from Type Ia SNe.

2.14.1 Extinction by dust

Cole et al. (2000) introduced a model for dust extinction in galaxies which significantly improved upon earlier 'slab' models. In Cole et al. (2000), the mass of dust is assumed to be proportional to the mass and metallicity of the ISM and to be mixed homogeneously with the ISM (possibly with a different scaleheight from the stars) and to have properties consistent with the extinction law observed in the Milky Way. To compute the extinction of any galaxy, a random inclination angle is selected and the extinction computed using the results of radiative transfer calculations carried out by Ferrara et al. (1999).

Following González-Perez et al. (2009), we extend this model²⁰ by assuming that some fraction, f_{cloud} , of the dust is in the form of dense molecular clouds where the stars form (see Baugh et al. 2005; Lacey et al. (2010)). Stars are assumed to form in these clouds and to escape on a time-scale of τ_{quies} (for quiescent star

²⁰An alternative method for rapidly computing dust extinction and re-emission within the GALFORM+GRASIL frameworks based on artificial neural networks is described by Almeida et al. (2010).

Table 1. Parameters of the dust model used throughout this work. The parameters are defined in Section 2.14.1.

Parameter	Value
f_{cloud}	0.25
r_{burst}	1.0
τ_{quies}	1 Myr
τ_{burst}	1 Myr
$\lambda_{1,\text{disc}}$	30 μm
$\lambda_{\text{break,disc}}$	10 000 μm
$\beta_{1,\text{disc}}$	2.0
$\beta_{2,\text{disc}}$	2.0
$\lambda_{1,\text{burst}}$	30 μm
$\lambda_{\text{break,burst}}$	100 μm
$\beta_{1,\text{burst}}$	1.6
$\beta_{2,\text{burst}}$	1.6

formation in discs) or τ_{burst} (for star formation in bursts), which is a parameter of the dust model (Granato et al. 2000), so these stars spend a significant fraction of their lifetime inside the clouds. Since massive, short-lived stars dominate the ultraviolet (UV) emission of a galaxy this enhances the extinction at short wavelengths.

To compute emission from dust we assume a far-infrared opacity of

$$\kappa = \begin{cases} \kappa_1 (\lambda/\lambda_1)^{-\beta_1} & \text{for } \lambda < \lambda_{\text{break}} \\ \kappa_1 (\lambda_{\text{break}}/\lambda_1)^{-\beta_1} (\lambda/\lambda_{\text{break}})^{-\beta_2} & \text{for } \lambda > \lambda_{\text{break}}, \end{cases} \quad (108)$$

where the opacity normalization at $\lambda_1 = 30 \mu\text{m}$ is chosen to be $\kappa_1 = 140 \text{ cm}^2 \text{ g}^{-1}$ to reproduce the dust opacity model used in GRASIL, as described in Silva et al. (1998). The dust grain model in GRASIL is a slightly modified version of that proposed by Draine & Lee (1984). Both the Draine & Lee (1984) and GRASIL dust models have been adjusted to fit data on dust extinction and emission in the local ISM (with much more extensive ISM dust emission data being used by Silva et al. 1998). The normalization is set at 30 μm because the dust opacity in the Draine & Lee (1984) and GRASIL models is well fit by a power-law longwards of that wavelength, but not shortwards. The dust luminosity is then assumed to be

$$L_\nu = 4\pi\kappa(\nu)B_\nu(T)M_{\text{Z,gas}}, \quad (109)$$

where $B_\nu(T) = [2h\nu^3/c^2]/[\exp(h\nu/kT) - 1]$ is the Planck black-body spectrum and $M_{\text{Z,gas}}$ is the mass of metals in gas. The dust temperature, T , is chosen such that the bolometric dust luminosity equals the luminosity absorbed by dust.

Values of the parameters used in dust model are given in Table 1 and were found by González-Perez et al. (2009) to give the best match to the results of the full GRASIL model.

This extended dust model, including diffuse and molecular cloud dust components, provides a better match to the detailed radiative transfer calculation of dust extinction carried out by the spectrophotometric code GRASIL (Silva et al. 1998; Baugh et al. 2004, 2005; Lacey et al. 2008) while being orders of magnitude faster, although it does not capture details such as polycyclic aromatic hydrocarbon (PAH) features.

Fontanot et al. (2009b) have explored similar models which aim to reproduce the results of GRASIL using simple, analytic prescriptions. They found that by fitting the results from GRASIL they were able to obtain a better match to the extinction in galaxies than previous, simplistic models of dust extinction had been able to attain. In this respect, our conclusions are in agreement with theirs – the model we describe here provides a significantly better match to the results

of the full GRASIL model than, for example, the dust extinction model described by Cole et al. (2000).

At high redshifts model galaxies often undergo periods of near continuous bursting as a result of experiencing disc instabilities on each subsequent time-step. This rather chaotic period of evolution is not well modelled presently – it is treated as a sequence of quiescent gas accretion periods punctuated by instability-triggered bursts while in reality we expect it to correspond more closely to a near continuous, high star formation rate mode somewhere in between the quiescent and bursting behaviour. While our model probably estimates the total amount of star formation during this period reasonably well (as it is controlled primarily by the cosmological infall rate and degree of outflow due to SNe) we suspect that it does a rather poor job of accounting for dust extinction. After each burst the gas (and hence dust) content of each galaxy is reduced to zero, resulting in no extinction. Our model therefore tends to contain too many dust-free galaxies at high redshifts. To counteract this effect we force galaxies in this regime to be observed during a bursting phase, so that they always experience some dust extinction.

Dust remains one of the most challenging aspects of galaxies to model. We will return to aspects of our model related to dust (utilizing the more detailed GRASIL model) in a future work, but note that even this is unlikely to be sufficient – what is needed is a better understanding of the complicated distribution of dust within galaxies, particularly during these early, chaotic phases.

Indeed, the distribution of star formation within galaxies at $z = 3$ to 5 has recently become within reach of observational studies (Stark et al. 2008; Elmegreen et al. 2009; Lehnert et al. 2009; Swinbank, Webb & Richard 2009). It seems that this aspect of the model is indeed supported by observational data. A future project will be to compare the internal properties of observed galaxies at these redshifts with those predicted by the model.

2.15 Absorption by the IGM

Where necessary, we model the attenuation of galaxy SEDs by neutral hydrogen in the intervening IGM using the model of Meiksin (2006).

3 MODEL SELECTION

The model described above has numerous free parameters which reflect our ignorance of the details of certain physical processes or order unity uncertainties in (e.g. geometrical) coefficients. To determine suitable values for these parameters we appeal to a broad range of observational data and search the model parameter space to find the best-fitting model.

The problem of how to implement the computationally challenging problem of fitting a complicated semi-analytic model with numerous free parameters to observational data has been considered before by Henriques et al. (2009) and Bower et al. (in preparation). To constrain model parameters in this work we use the ‘Projection Pursuit’ method of Bower et al. (in preparation). We give a brief description of that method here and refer the reader to Bower et al. (in preparation) for complete details.

Running a single set of model parameters, including all of the redshifts and wavelengths required for our analysis, is a relatively slow process. In particular, running a model with self-consistently computed IGM evolution is entirely impractical for a parameter space search. We therefore chose to run models without a self-consistently computed IGM or photoionizing background. Even then, each model takes around 2 h to run on a fast computer. To

Table 2. The allowed ranges for each parameter in our fitting parameter space. For some parameters, we choose to use the logarithm of the parameter to allow efficient exploration of several decades of parameter value.

Parameter	Minimum	Maximum
h_0	0.6750	0.7270
Ω_b	0.04320	0.04920
Λ_0	0.7142	0.7278
σ_8	0.7650	0.8690
n_s	0.9320	0.9880
$V_{\text{cut}}/\text{km s}^{-1}$	10.00	50.00
z_{cut}	5.000	13.00
$\log_{10}(\alpha_{\text{cool}})$	-1.523	0.4771
$\log_{10}(\alpha_{\text{remove}})$	-1.523	0.0000
$\log_{10}(\alpha_{\text{core}})$	-2.000	-0.5229
$\log_{10}(\epsilon_*)$	-3.523	-1.301
α_*	-4.000	1.000
$V_{\text{hot,disc}}/\text{km s}^{-1}$	100.0	550.0
$V_{\text{hot,burst}}/\text{km s}^{-1}$	100.0	550.0
α_{hot}	1.000	3.700
$\log_{10}(\lambda_{\text{expel,disc}})$	-1.523	1.000
$\log_{10}(\lambda_{\text{expel,burst}})$	-1.523	1.000
$\log_{10}(\epsilon_*)$	-2.398	-1.000
$\log_{10}(\eta_*)$	-3.000	-1.000
$\log_{10}(F_*)$	-3.000	-1.523
$\log_{10}(\alpha_{\text{reheat}})$	-1.523	0.4771
$\log_{10}(f_{\text{ellip}})$	-2.000	-0.3010
$\log_{10}(f_{\text{burst}})$	-2.000	-0.3010
$\log_{10}(f_{\text{gas,burst}})$	-1.523	-0.3010
B/T_{burst}	0.0000	1.000
A_{ac}	0.7000	1.000
w_{ac}	0.7000	1.000
$\epsilon_{\text{d,gas}}$	0.7000	1.150
$\log_{10}(\epsilon_{\text{strip}})$	-2.000	0.0000

mimic the effects of a photoionizing background we adopt the ‘ $V_{\text{cut}}-z_{\text{cut}}$ ’ model described by Font et al. (in preparation) and which they show to reproduce quite well the results of the self-consistent calculation. Briefly, this model inhibits cooling of gas in haloes with virial velocities below V_{cut} at redshifts below z_{cut} . We then include V_{cut} and z_{cut} as parameters in our fitting process.

This approach is not ideal, but is required due to computational limitations. Bower et al. (in preparation) show that local (i.e. low redshift) properties of the model are not significantly affected by the inclusion of self-consistent reionization (i.e. those data do not constrain V_{cut} or z_{cut}), and, where they are, the ‘ $V_{\text{cut}}-z_{\text{cut}}$ ’ model provides a reasonable approximation (Font et al., in preparation). In any case, as we will discuss below, some manual tuning of parameters is still required after the automated search of parameter space is completed. This manual search is then conducted using the fully self-consistent IGM calculation.

We envision the problem in terms of a multi-dimensional parameter space into which constraints from observational data are mapped. Given the large number of model parameters and the fact that running a single realization of the model requires a significant amount of computer time, we cannot perform a simple grid-search of the parameter space on a sufficiently fine grid. Instead, we begin by specifying plausible ranges for model parameters. The ranges considered for each parameter are listed in Table 2 – for some parameters we choose to consider the logarithm of the parameter as the variable in our parameter space to allow for efficient explo-

ration of several decades of parameter value. We scale each model parameter such that it varies between 0 and 1 across this allowed range. We then generate a set of points in this limited and scaled model parameter space using Latin hypercube sampling (McKay, Beckman & Conover 1979), thereby ensuring an efficient coverage of the parameter space. A model is run for each set of parameters and a goodness of fit measure computed.

The choice of a goodness of fit measure is important and non-trivial (see Bower et al., in preparation). We do not expect our model to fit all of the constraints in a statistically rigorous manner, as the model is clearly approximate. The Bayesian approach to this issue is to assign a prior assessment of the reliability of the model to each of the data set comparisons and to define a correlation matrix reflecting the a priori connections between data sets. This concept (referred to as ‘model discrepancy’ in the statistical literature) is discussed in detail for $z = 0$ luminosity function constraints in Bower et al. (in preparation). However, in the present paper, we needed a simpler approach to the problem. We therefore adopted a non-Bayesian methodology of simply summing χ^2 for each data set that we used. This has the advantage of simplicity, but clearly there may be more appropriate choices for the relative weighting of different data sets: we will explore this issue in a future paper. There is little doubt that a better measure of goodness of fit could be found. In particular, the relative weightings given to each data set should really reflect how well we think the model performs in that particular quantity, how accurately we think that we have been able to match any observational selection and, inevitably, how much we believe the data themselves. These are extremely thorny issues to which, at present, we do not have a good answer.

Specifically, in this work, the goodness of fit measure is taken to be

$$\tilde{\chi}^2 = \sum_i w_i \frac{\chi_i^2}{N_i}, \quad (110)$$

where χ_i^2 is the usual goodness of fit measure for data set i , N_i is the number of degrees of freedom in that data set and w_i is a weight assigned to each data set. The sum is taken over all data sets shown in Section 4 and, additionally, cosmological parameters were allowed to vary within the 2σ intervals permitted by the Dunkley et al. (2009) constraints, and were included in the goodness of fit measure using a Gaussian prior. When computing χ^2 for each data set we estimate the error in each datum to be the sum in quadrature of the experimental error and any statistical error present in the model due to the finite number of Monte Carlo merger tree realizations that we are able to carry out. This ensures that two models which differ by an amount comparable to the random noise in the models have similar values of χ^2 . The specific data sets used, along with the weights assigned to them (estimated using our best judgement of the reliability of each data set and the GALFORM’s ability to model it), are listed in Table 3.

Once a set of models have been run, a principal components analysis is performed on the goodness of fit values of those models with $\tilde{\chi}^2$ values in the lower 10th percentile of all models to find which linear combinations of parameters provide the *minimum* variance in goodness of fit. These are the parameter combinations that are most tightly constrained by the observational data. A principal component with low variance implies that this particular combination of the parameters is tightly constrained if the model is likely to produce an acceptable fit. Of course, even if this constraint is satisfied, a good model is not guaranteed; rather we can be confident that

Table 3. The set of data sets used as constraints on our model, together with a reference to where the data set is shown in this paper and the value of the weight, w_i , assigned to each constraint.

Constraint	Reference	Weight (w_i)
Star formation history	Section 4.2	1.00
b _J -band $z = 0$ luminosity function	Fig. 9	2.00
<i>K</i> -band $z = 0$ luminosity function	Fig. 10	2.00
Morphologically segregated $z = 0$ luminosity function	Fig. 13	1.00
60 μm $z = 0$ luminosity function	Fig. 11	1.00
Evolving <i>K</i> -band luminosity function	Fig. 12	1.00
$z = 3$ UV luminosity function	Fig. 14	1.00
$z = 5$ UV luminosity function	Fig. 15	0.75
$z = 6$ UV luminosity function	Fig. 15	0.75
Tully–Fisher relation	Section 4.5	2.00
Gas-phase metallicities	Fig. 20	1.00
Colour distributions	Section 4.4	2.00
Half-light radius distributions	Fig. 18	1.50
Disc scalelength distributions	Fig. 19	2.00
Supermassive black hole mass distributions	Section 4.9	1.00
Stellar metallicities	Fig. 21	1.00
Gas-to-light ratios	Fig. 22	1.00
Clustering	Section 4.8	1.50
Local Group luminosity function	Fig. 25	1.00
Local Group satellite galaxy sizes	Fig. 26	1.00
Local Group satellite galaxy metallicities	Fig. 27	1.00

if it is not satisfied the fit will not be good.²¹ When analysing the acceptable region in this way, we also need to bear in mind that the PCA assumes that the relationships are linear, whereas Bower et al. (in preparation) show that the actual acceptable space is curved. This will prevent any of the suggested projections being arbitrarily thin and limit the accuracy of constraints. Nevertheless, the procedure substantially cuts down the volume of parameter space where model evaluations need to be run. These linear combinations are used to define rotated axes in the parameter space within which we select a new set of points again using the Latin hypercube sampling. The process is repeated until a suitably converged model is found.²² This process is not fast, requiring around 150 000 CPU hours,²³ but does produce a model which is a good match to the input data.

Fig. 7 demonstrates the efficacy of our method using four 2D slices through the multi-dimensional parameter space. The colour scale in each panel shows constraints on two of the model parameters, while the projections below and to the left of the panel indicate the constraints on the indicated single parameter. Contours illustrate

the relative number of model evaluations which were performed at each point in the plane. It can be clearly seen that our ‘Projection Pursuit’ methodology concentrates model evaluations in those regions which are most likely to provide a good fit. The nominal best-fitting model is indicated by a yellow star in each panel. Despite the large number of models run we do not believe that this precise point should be considered as the ‘best’ model – the dimensionality of the parameter space is so large that we do not believe that it has been sufficiently well mapped to draw this conclusion. Additionally, we also need a model discrepancy matrix – without this, we cannot say whether a model is acceptable (in the sense that it should only agree with the data as well as we expect given the level of approximation in the model). Without the discrepancy term, we will tend to overfit the model. Instead, we utilize these results to suggest the region of parameter space in which the best model is likely to be found. We then adjust parameters manually to find the final model (utilizing our intuition of how the model will respond to changes in parameters).

Interesting constraints and correlations can be seen in Fig. 7. For example, the combination $\alpha_{\text{hot}} - V_{\text{hot,disc}}$ is quite well constrained and somewhat anti-correlated (such that an increase in α_{hot} can be played off against a decrease in $V_{\text{hot,disc}}$). It is immediately clear, for example, that no good model can be found with $\lambda_{\text{expel,disc}} \gtrsim 1.5$ while $\lambda_{\text{expel,bulge}}$ is much less well constrained, but must be larger than about 1.5.

The principal component vectors from the final set of 36 017 models are shown in Table 4. We note here that these vectors are quite different from those found by Bower et al. (in preparation). This is not too surprising as our implementation of GALFORM is quite different from theirs and we constrain our model to a much broader collection of data sets. We will examine the PCA vectors in greater detail in a future paper, and so restrict ourselves to a brief discussion here. Taking the first PCA vector for example, we see that it is dominated by z_{cut} , α_* and α_{hot} . These parameters all have strong effects on the faint end of luminosity functions. Luminosity functions are abundant in our set of constraints and have

²¹This is only strictly true if the relationships between $\tilde{\chi}^2$ and the parameters are approximately linear and unimodal. If there exists a separate small island of good values somewhere, our PCA+Latin Hypercube method might happen to miss the region, or it might not exert sufficient pull on the PCA compared to the large region and might be subsequently ignored. The advantage of the emulator approach used by Bower et al. (in preparation) is that it gives an estimate of the error made by excluding regions from further evaluations.

²²In practice, these calculations were run on distributed computing resources (including machines at the ICC in Durham, TeraGrid and Amazon EC2). Each machine was given an initial small set of models to run. After running each model, the results were transferred back to a central server. Periodically, the server would collate all available results, perform the PCA and generate a new set of models which it then distributed to all active computing resources.

²³The authors, feeling the need to help preserve our own small region of one realization of the Universe, purchased carbon offsets to counteract the carbon emissions resulting from this large investment of computing time.

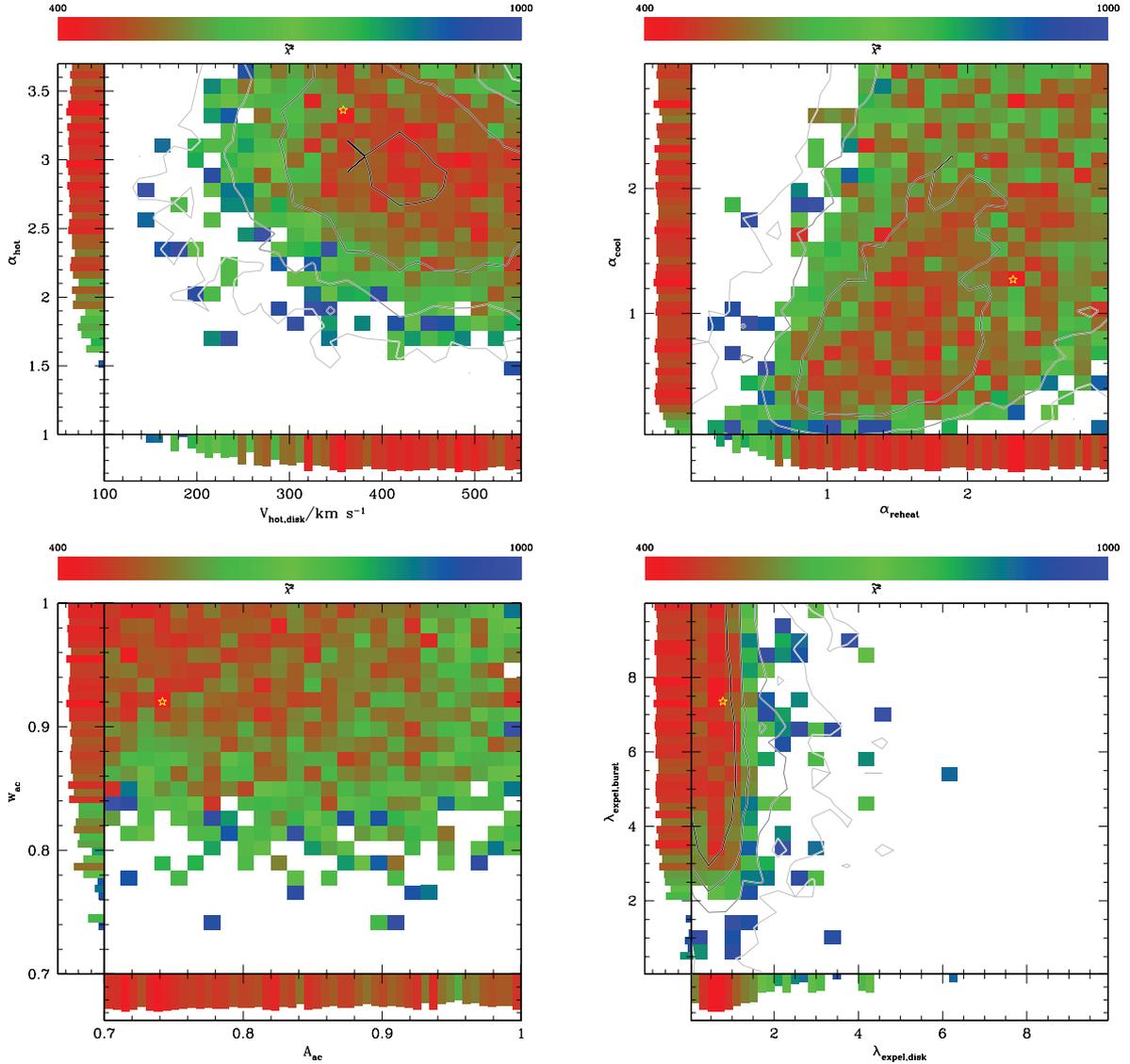


Figure 7. Constraints on model parameters shown as 2D slices through the multi-dimensional parameter space. In each panel, the colour scale indicates the value of $\tilde{\chi}^2$ as shown by the bar above the panel, with the yellow star indicating the best-fitting model. Each point in the plane is coloured to correspond to the minimum value of $\tilde{\chi}^2$ found when projecting over all other dimensions of the parameter space. Contours illustrate the relative number of model evaluations at each point in the plane – from lightest to darkest line colour they correspond to 10, 30, 100 and 300 evaluations per grid cell. Most evaluations are carried out when the best model fits are found, indicating that our method is efficient in concentrating resources where good models are most likely to be found. To each side of the plane, the distribution of $\tilde{\chi}^2$ is projected over one of the remaining dimensions to show constraints on the indicated parameter. Top left panel: the main parameters of the SNe feedback model, $V_{\text{hot,disk}}$ and α_{hot} . Top-right panel: critical parameters controlling the cooling and AGN feedback models, α_{reheat} and α_{cool} . Lower-left panel: parameters of the adiabatic contraction model which have important consequences for the sizes of galaxies, A_{ac} and w_{ac} . Lower-right panel: parameters of the SNe feedback model that control the amount of material expelled from haloes, $\lambda_{\text{expel,disk}}$ and $\lambda_{\text{expel,burst}}$.

been well measured. As such, they provide some of the strongest constraints on the model. It can be seen that an increase in α_{hot} , which will flatten the slope of the faint of a luminosity function, has a similar effect as a decrease in α_{*} , which will preferentially reduce rates of star formation in low-mass galaxies and so also flatten the faint-end slope. The second PCA component shows a strong but opposite dependence on Ω_{b} and $\lambda_{\text{expel,burst}}$. Increasing Ω_{b} results in more fuel for galaxy formation, while increasing $\lambda_{\text{expel,burst}}$ causes material to be lost by being expelled from haloes. As we continue to further PCA vectors the parameter combinations they represent become more complicated and difficult to interpret – the advantage of our methodology is that these complex interactions

can be taken into account when exploring the model parameter space.

The differences between our results and those of Bower et al. (in preparation) are interesting in their own right. For example, Bower et al. (in preparation) found two ‘islands’ of good fit in the SNe feedback parameter space ($V_{\text{hot,disk}}$ and $V_{\text{hot,burst}}$): a strong feedback island (corresponding approximately to what we find in this work) and a weak feedback island (which we do not find). The weak feedback island is ruled out in the present work as, while a good fit to the galaxy luminosity function can be found in it (as demonstrated by Bower et al., in preparation), no good fit to, for example, galaxy sizes can be found.

Table 4. The principal components, rank ordered by their contribution to the variance, σ^2 , from our models. In each row, the dominant elements (those with an absolute value in excess of 0.33) are shown in bold.

PCA	σ	h_0	Ω_b	Λ_0	σ_8	n_s	V_{cut}	z_{cut}	α_{cool}	α_{remove}	α_{core}	ϵ_*	α_*	$V_{\text{hot,disc}}$	$V_{\text{hot,burst}}$	α_{hot}
PCA 1	0.025	0.186	-0.049	0.027	-0.168	0.036	0.022	-0.352	0.045	-0.227	-0.123	-0.261	0.425	-0.097	-0.051	-0.379
PCA 2	0.043	0.115	-0.627	-0.283	-0.040	-0.283	-0.032	0.182	0.152	0.050	0.095	-0.111	0.011	-0.143	0.097	0.008
PCA 3	0.045	-0.185	-0.116	0.086	0.203	0.122	0.364	-0.039	0.147	0.093	-0.029	-0.010	0.166	0.420	-0.213	-0.027
PCA 4	0.051	-0.138	0.004	-0.032	-0.032	0.008	-0.055	-0.056	-0.011	0.410	0.047	-0.424	0.392	0.030	-0.023	-0.284
PCA 5	0.064	-0.066	0.058	-0.128	-0.140	0.158	-0.076	0.174	0.014	-0.050	-0.102	0.046	-0.230	0.203	0.107	-0.270
PCA 6	0.083	0.045	-0.117	0.031	0.135	0.141	0.217	-0.081	0.112	-0.159	-0.260	-0.129	-0.092	0.216	0.053	-0.060
PCA 7	0.104	-0.070	-0.290	-0.658	0.033	0.149	0.121	-0.027	-0.008	0.019	-0.004	0.112	-0.103	0.056	-0.023	-0.152
PCA 8	0.106	-0.137	-0.109	-0.059	0.041	0.009	0.145	-0.225	-0.616	0.028	-0.149	-0.205	0.040	0.108	0.271	0.471
PCA 9	0.111	0.123	0.132	0.032	0.070	-0.058	0.579	0.027	0.255	0.064	0.066	0.299	0.138	-0.288	0.423	-0.038
PCA 10	0.114	-0.054	0.098	-0.050	-0.047	-0.053	0.089	-0.152	0.411	0.222	-0.032	-0.236	-0.116	-0.296	-0.212	0.407
PCA 11	0.122	-0.015	-0.001	-0.012	0.003	0.025	-0.043	0.047	0.000	0.013	0.049	0.008	0.004	-0.029	0.041	0.007
PCA 12	0.128	0.092	0.043	-0.037	-0.072	-0.263	0.037	-0.006	-0.010	0.018	-0.122	0.069	0.000	0.034	-0.012	-0.001
PCA 13	0.137	0.002	-0.178	-0.027	0.098	0.018	0.258	0.056	-0.118	-0.177	0.200	-0.020	-0.124	-0.050	-0.503	0.047
PCA 14	0.142	0.128	-0.031	0.077	0.063	0.035	-0.022	-0.017	-0.032	0.467	-0.158	0.195	-0.015	-0.022	-0.482	0.032
PCA 15	0.147	0.112	-0.036	-0.059	0.038	-0.074	-0.307	-0.680	0.101	-0.105	0.256	0.151	-0.224	0.015	-0.007	-0.033
PCA 16	0.155	0.223	0.021	0.037	0.016	-0.091	0.139	0.140	-0.162	-0.505	0.194	-0.079	0.231	-0.135	-0.287	0.101
PCA 17	0.157	-0.089	0.080	0.040	0.014	0.075	-0.110	0.011	-0.003	0.014	0.039	0.096	0.011	-0.076	-0.030	-0.017
PCA 18	0.168	0.030	-0.122	-0.123	-0.013	0.068	-0.216	-0.025	0.302	-0.074	-0.131	0.061	0.384	0.147	0.080	0.474
PCA 19	0.174	0.727	-0.017	0.029	0.123	0.160	0.000	0.093	-0.191	0.313	0.204	-0.075	-0.074	0.143	0.106	0.007
PCA 20	0.182	-0.099	-0.458	0.551	0.334	-0.087	-0.014	-0.077	0.094	-0.033	-0.075	-0.117	-0.237	-0.063	0.135	-0.109
PCA 21	0.189	-0.131	0.204	-0.150	-0.162	-0.023	0.290	-0.146	-0.050	0.032	0.032	-0.375	-0.360	-0.178	-0.011	-0.064
PCA 22	0.196	-0.017	-0.322	0.291	-0.773	0.318	0.130	-0.047	-0.016	0.062	0.150	0.134	-0.040	0.009	0.002	0.085
PCA 23	0.204	-0.152	0.001	0.058	-0.110	-0.015	-0.195	0.339	0.005	-0.107	-0.246	-0.008	-0.044	-0.068	-0.066	-0.028
PCA 24	0.231	0.099	0.151	0.011	-0.043	0.053	-0.033	0.157	0.346	-0.148	0.292	-0.374	-0.176	0.431	0.074	0.132
PCA 25	0.242	-0.122	-0.032	-0.067	0.232	0.359	-0.154	-0.005	-0.022	-0.068	0.091	0.041	0.022	-0.266	0.025	-0.027
PCA 26	0.279	-0.049	0.084	0.018	0.111	0.314	-0.011	0.007	-0.015	-0.018	0.154	0.051	0.086	-0.077	-0.005	0.042
PCA 27	0.343	0.131	-0.042	-0.041	0.120	0.596	0.005	-0.042	0.111	-0.097	-0.233	-0.082	-0.051	-0.183	0.009	0.054
PCA 28	0.366	0.111	-0.046	0.031	0.011	0.071	-0.151	0.227	-0.039	0.040	-0.047	-0.326	-0.063	-0.333	0.069	0.023
PCA 29	0.436	-0.345	-0.051	0.034	0.125	0.098	-0.041	0.081	-0.047	0.109	0.598	-0.009	0.155	-0.038	0.092	-0.004

Table 4 – continued

PCA	$\lambda_{\text{expel,disc}}$	$\lambda_{\text{expel,burst}}$	ϵ_{\bullet}	η_{\bullet}	F_{\bullet}	α_{reheat}	f_{ellip}	f_{burst}	$f_{\text{gas,burst}}$	B/M_{burst}	A_{ac}	w_{ac}	$\epsilon_{\text{d,gas}}$	ϵ_{strip}
PCA 1	0.093	0.047	-0.162	0.033	-0.168	0.297	-0.288	0.187	-0.024	0.006	0.009	-0.016	0.249	-0.003
PCA 2	0.093	-0.486	-0.053	-0.059	0.073	0.109	0.145	0.048	0.025	-0.105	-0.052	0.052	0.092	-0.054
PCA 3	0.199	0.006	0.395	0.049	0.285	0.138	0.208	0.297	0.054	0.037	-0.033	-0.020	0.194	0.007
PCA 4	-0.296	-0.021	0.006	-0.073	0.230	-0.405	0.060	-0.233	0.019	-0.008	-0.027	0.010	-0.117	-0.057
PCA 5	-0.234	-0.234	-0.128	0.639	0.147	-0.173	-0.158	0.258	-0.045	-0.008	0.033	0.012	0.135	0.046
PCA 6	0.050	-0.241	0.231	0.209	-0.265	0.072	-0.142	-0.629	-0.018	-0.011	-0.065	-0.025	-0.200	-0.117
PCA 7	0.102	0.575	-0.116	-0.001	0.057	-0.024	-0.088	-0.102	-0.049	-0.007	-0.006	0.013	-0.071	-0.074
PCA 8	-0.053	-0.044	-0.157	0.066	0.115	0.029	-0.068	-0.043	-0.001	-0.001	0.008	0.002	0.289	-0.046
PCA 9	-0.080	0.015	-0.030	0.000	0.049	-0.219	-0.070	-0.109	0.009	-0.026	0.063	-0.022	0.311	-0.016
PCA 10	-0.154	0.074	-0.051	0.267	0.287	0.319	-0.237	-0.018	0.007	-0.021	-0.006	-0.024	0.131	0.008
PCA 11	-0.007	-0.019	-0.006	-0.022	0.010	0.001	-0.002	0.033	-0.308	0.108	-0.453	-0.817	0.040	-0.085
PCA 12	-0.066	0.010	0.064	-0.005	0.043	-0.012	-0.038	0.025	-0.252	0.678	-0.411	0.409	0.003	-0.139
PCA 13	-0.560	-0.061	0.107	-0.171	-0.252	-0.140	-0.219	0.092	-0.009	-0.001	0.036	-0.016	0.166	-0.045
PCA 14	0.327	-0.087	-0.335	0.153	-0.192	-0.179	0.019	-0.178	-0.037	0.027	0.003	-0.011	0.317	0.043
PCA 15	-0.012	-0.062	0.255	0.034	0.205	-0.194	0.067	-0.149	-0.112	-0.039	0.047	0.003	0.252	0.025
PCA 16	0.237	0.043	-0.109	0.303	0.352	-0.250	0.124	-0.140	-0.022	0.025	0.022	-0.029	-0.157	-0.025
PCA 17	0.047	-0.044	-0.004	0.024	0.032	0.042	-0.011	0.058	0.112	-0.024	0.134	-0.001	0.049	-0.950
PCA 18	0.055	0.030	0.134	0.068	-0.279	-0.435	-0.202	0.236	-0.011	0.007	0.038	-0.007	-0.085	-0.018
PCA 19	0.051	0.023	0.189	-0.023	0.124	0.049	-0.322	0.124	0.024	0.007	0.050	-0.019	-0.143	-0.042
PCA 20	0.041	0.269	-0.219	0.046	0.057	-0.203	-0.140	0.147	-0.049	0.005	-0.047	0.034	-0.118	-0.047
PCA 21	0.430	-0.168	0.116	-0.116	-0.161	-0.335	-0.089	0.269	-0.010	0.001	-0.022	0.005	-0.127	-0.029
PCA 22	0.010	0.082	0.058	0.001	0.034	-0.031	0.038	-0.093	0.085	0.058	-0.072	0.013	-0.048	-0.012
PCA 23	0.183	0.006	0.245	-0.293	0.357	-0.062	-0.484	-0.190	-0.174	-0.071	0.109	0.020	0.306	0.051
PCA 24	0.032	0.104	-0.348	-0.174	-0.035	-0.025	0.055	-0.146	0.105	0.077	-0.072	0.042	0.338	-0.057
PCA 25	0.074	-0.126	0.020	-0.014	0.073	-0.025	-0.144	-0.036	0.657	0.322	-0.280	-0.020	0.074	0.127
PCA 26	0.030	-0.066	-0.052	-0.015	0.009	0.021	-0.041	0.034	-0.270	-0.521	-0.577	0.384	0.002	-0.034
PCA 27	-0.083	-0.179	-0.158	-0.232	0.139	-0.017	0.281	0.070	-0.359	0.236	0.277	0.031	-0.044	0.016
PCA 28	-0.025	0.335	0.414	0.301	-0.249	0.003	0.329	-0.031	-0.042	-0.006	-0.025	0.053	0.350	-0.013
PCA 29	0.173	-0.080	0.017	0.176	-0.176	0.158	-0.200	-0.085	-0.344	0.257	0.270	0.055	-0.020	0.077

4 RESULTS

In this section we will begin by identifying the best-fitting model and will then show results from that model compared to the observational data that was used to constrain the model parameters. With the exception of results shown in Section 4.12, all of the data shown in this section were used to constrain the model and, as such, the results do not represent predictions of the model. (In Section 4.12.1, we examine the distribution of gas between different phases as a function of halo mass, while in Section 4.12.2 we explore the fraction of stellar mass in the intracluster light component of haloes. The data shown in these comparisons were *not* used as constraints when searching for the best-fitting model.) The overall best-fitting model (i.e. that which best describes the union of all data sets) is shown by blue lines. Additionally, we show as magenta lines the best-fitting model to each individual data set (as described in the figure captions) for comparison. We do not claim that the following represents a complete census of the observational data that *could* be used to constrain our galaxy formation model. Instead, we have selected data which span a range of physical characteristics and redshifts that we think best constrain the physics of our model, while remaining within the limited (although substantial) computational resources at our disposal.

In addition to these best-fitting models, we will, where possible, compare our current results with those from the previous implementation of GALFORM described by Bower et al. (2006). Results from the Bower et al. (2006) model are shown by green lines in each figure. We have not included figures for every constraint used in this work – specifically, in many cases we show examples of the constraints only for a limited number of magnitude or redshift ranges. However, all of the constraints used are listed in Table 3 and are discussed in the text.

4.1 Best-fitting model

The resulting set of best-fitting parameters are listed in Table 5. We will not investigate the details of these results here, leaving an exploration of which data constrain which parameters and the possibility of alternative, yet acceptable, parameter sets to a future work. The best-fitting model turns out to be a reasonably good match to local luminosity data, galaxy colours, metallicities, gas content, supermassive black hole masses and constraints on the epoch of reionization, but to perform less well in matching galaxy sizes, clustering and the Tully–Fisher relation. In addition, luminosity functions become increasingly more discrepant with the data as we move to higher redshifts. In the remainder of this section we will briefly discuss some important aspects of the best-fitting parameter set.

The cosmological parameters are all close to the *WMAP* five-year expectations (by construction). The parameters of the gas cooling model are all quite reasonable: the three parameters α_{reheat} and α_{cool} are all of the order of unity as expected, α_{remove} is somewhat smaller but still plausible, while the core radius a_{core} is around 22 per cent of the virial radius. The parameters of the adiabatic contraction model differ from those proposed by Gnedin et al. (2004) but are within the range of values found by Gustafsson, Fairbairn & Sommer-Larsen (2006) when fitting the profiles of dark matter haloes in simulations including galaxy formation with feedback. The disc stability parameter, $\epsilon_{\text{d,gas}}$, is close to, albeit lower than, the value of 0.9 suggested by the theoretical work of Christodoulou et al. (1995). The stripping parameter, ϵ_{strip} , is of the order of unity as expected.

The star formation parameters are reasonable, implying a low efficiency of star formation. The feedback parameters, $V_{\text{hot,disc|burst}}$

Table 5. Parameters of the best-fitting model used in this work and of the Bower et al. (2006) model. Note that the best-fitting model listed here is one that includes self-consistent reionization and evolution of the IGM (see Section 2.10) and which has been adjusted to also produce a reasonable reionization history (see Section 4.11). It therefore does not correspond to the location of the best-fitting model indicated in Fig. 7. Where appropriate, references are given to the paper, or section of this work, in which the parameter is described.

Parameter	This work	Value		Reference
		This work	Bower et al. (2006)	
<i>Cosmological</i>				
Ω_0	0.284	0.250		
Λ_0	0.716	0.750		
Ω_b	0.04724	0.04500		
h_0	0.691	0.730		
σ_8	0.807	0.900		
n_s	0.933	1.000		
<i>Gas cooling model</i>				
α_{reheat}	2.32	1.260		Section 2.6.2
α_{cool}	0.550	0.580		Section 2.13
α_{remove}	0.102	N/A		Section 2.6.2
a_{core}	0.163	0.100		Section 2.6.3
<i>Adiabatic contraction</i>				
A_{ac}	0.742	1.000		Section 2.7
w_{ac}	0.920	1.000		Section 2.7
<i>Star formation</i>				
ϵ_*	0.0152	0.0029		Cole et al. (2000)
α_*	−3.28	−1.50		Cole et al. (2000)
<i>Disc stability</i>				
$\epsilon_{\text{d,gas}}$	0.743	0.800 ²⁴		Section 2.4.1
<i>SNe feedback</i>				
$V_{\text{hot,disc}}$	358.0 km s ^{−1}	485.0 km s ^{−1}		Section 2.12
$V_{\text{hot,burst}}$	328.0 km s ^{−1}	485.0 km s ^{−1}		Section 2.12
α_{hot}	3.36	3.20		Section 2.12
$\lambda_{\text{expel,disc}}$	0.785	N/A		Section 2.12
$\lambda_{\text{expel,burst}}$	7.36	N/A		Section 2.12
<i>Ram-pressure stripping</i>				
ϵ_{strip}	0.335	N/A		Section 2.9.1
<i>Merging</i>				
f_{ellip}	0.0214	0.3000		Cole et al. (2000)
f_{burst}	0.335	0.100		Cole et al. (2000)
$f_{\text{gas,burst}}$	0.331	0.100		Section 2.3
B/T_{burst}	0.672	N/A		Section 2.3
<i>Black hole growth</i>				
ϵ_{\bullet}	0.0134	0.0398		Section 2.13
η_{\bullet}	0.0163	N/A		Section 2.13
F_{SMBH}	0.0125	0.00500		Malbon et al. (2007)

are much lower than the value of 485 km s^{−1} required by Bower et al. (2006) and significantly closer to the value of 200 km s^{−1} adopted by Cole et al. (2000). This is desirable as values around 200 km s^{−1} already stretch the SNe energy budget. We also note that the value of α_{hot} is lower than that required by Bower et al. (2006) and closer to the ‘natural’ value of 2, which would imply an efficiency of SNe energy coupling into feedback that was independent of galaxy properties. The expulsion parameters, $\lambda_{\text{expel,disc|burst}}$, are close to unity as expected.

²⁴ The Bower et al. (2006) model used a single value of ϵ_{d} for both gaseous and stellar discs.

The parameters of the merging model imply that mass ratios of 1:10 or greater are required for a major merger, a little low, but within the range of plausibility, while only 1:5 or greater mergers trigger a burst. Minor mergers in which the primary galaxy has at least 34 per cent gas by mass and at least 34 per cent of its mass in a disc can also lead to bursts.

Finally, the black hole growth parameters are quite reasonable: black holes radiate at about 9 per cent of the Eddington luminosity, 5 per cent of cooling gas reaches the black hole during radio mode feedback and around 0.5 per cent of gas in a merging event is driven into the black hole.

Overall, the parameters of the best-fitting model seem reasonable on physical grounds. Given the large dimensionality of the parameter space, the complexity of the model and the various assumptions used in modelling complex physical processes we would not consider these values to be either precise or accurate (which is why we do not quote error bars here), but to merely represent the most plausible values within the context of the GALFORM semi-analytic model of galaxy formation.

In addition to this overall best-fitting model, we show in Table 6 the parameters which produced the best fit to subsets of the data (as indicated). We caution that these models were selected from runs without self-consistent reionization and also with relatively few realizations of merger trees, making them noisy. This means that, after re-running these models with many more merger tree realizations it is possible that they will not be such good fits to the data. We do, in fact, find such cases as we will highlight below. Nevertheless, we will refer to this table in the remainder of this section when exploring the ability of our model to match each data set. We also point out that there is no guarantee that any of these

models that provide a good match to an individual data set are good matches overall – for example, the model which best matches galaxy sizes may produce entirely unacceptable $z = 0$ luminosity functions.

4.2 Star formation history

Fig. 8 shows the star formation rate per unit volume as a function of redshift, with symbols indicating observational estimates and lines showing results from our model. Dotted and dashed lines show quiescent star formation in discs and bursts of star formation, respectively, while solid lines indicate the sum of these two. The quiescent mode dominates at all redshifts, although we note that at high redshifts model discs are typically unstable and undergo frequent instability events. These galaxies may therefore not look like typical low-redshift disc galaxies. The best-fitting model is in excellent agreement with the star formation rate data from $z = 1$ to $z = 8$, reproducing the sharp decline in star formation below $z = 2$ while maintaining a relatively high star formation rate out to the highest redshifts. Our model lies below the data at $z \lesssim 1$ despite being a good match to the b_J -band luminosity function (see Section 4.3). This suggests some inconsistency in the data analysis, perhaps related to the choice of IMF or the calibration of star formation rate indicators. Indeed, the model which best fits this particular data set (shown as magenta lines in Fig. 8) does so by virtue of having a large value of ϵ_* (see Table 6; this increases star formation rates overall) and a small value of α_{cool} (which alters the critical mass scale for AGN feedback and thereby delays the truncation of star formation at low redshifts). While these changes

Table 6. Parameters of the overall best-fitting model compared to those of models which best fit individual data sets (as indicated by column labels). Parameters which play a key role (as discussed in the relevant subsections of Section 4) in helping to obtain a good fit to each data set are shown in bold type.

Parameter	Overall	Star formation rate	b_J & K LFs	60 μm LF	K20 LFs	Morphological LFs	$z = 3$ UV LF	$z = 5$ & 6 UV LFs	Colours	Tully–Fisher
Λ_0	0.716	0.723	0.723	0.717	0.721	0.720	0.717	0.721	0.723	0.722
Ω_b	0.04724	0.0445	0.0465	0.0479	0.0471	0.0491	0.0452	0.0477	0.0482	0.0441
h_0	0.691	0.677	0.711	0.714	0.707	0.726	0.700	0.703	0.689	0.724
σ_8	0.807	0.799	0.786	0.805	0.785	0.788	0.779	0.765	0.808	0.783
n_s	0.933	0.955	0.957	0.939	0.952	0.960	0.947	0.946	0.951	0.959
α_{reheat}	2.32	2.68	1.98	1.47	1.76	2.34	2.38	2.16	2.26	1.91
α_{cool}	0.550	0.571	2.31	1.12	1.50	2.67	2.10	2.81	0.588	1.06
α_{remove}	0.102	0.842	0.692	0.133	0.0986	0.547	0.0607	0.228	0.162	0.125
a_{core}	0.163	0.128	0.168	0.155	0.0695	0.187	0.0515	0.142	0.109	0.216
A_{ac}	0.742	0.920	0.819	0.764	0.780	0.770	0.804	0.746	0.880	0.860
w_{ac}	0.920	0.792	0.954	0.941	0.957	0.989	0.968	0.972	0.868	0.817
ϵ_*	0.0152	0.0695	0.0453	0.0375	0.00520	0.0281	0.0223	0.300	0.0153	0.0427
α_*	−3.28	−2.11	−2.73	−2.65	−2.05	−3.43	−2.95	−3.68	−0.392	−1.69
$\epsilon_{\text{d,gas}}$	0.743	0.734	0.743	0.726	0.829	0.774	0.804	0.957	0.808	0.812
$V_{\text{hot,disc}}$	358.0	425.0	532.0	421.0	491.0	411.0	506.0	546.0	459.0	389.0
$V_{\text{hot,burst}}$	328.0	130.0	470.0	413.0	539.0	498.0	544.0	488.0	242.0	370.0
α_{hot}	3.36	2.61	2.81	2.73	3.57	3.50	3.53	3.15	2.95	2.58
$\lambda_{\text{expel,disc}}$	0.785	0.738	0.252	0.920	0.273	0.477	0.571	0.266	0.607	0.551
$\lambda_{\text{expel,burst}}$	7.36	6.49	7.90	5.39	5.23	7.46	6.62	9.23	6.55	2.13
ϵ_{strip}	0.335	0.951	0.696	0.248	0.207	0.0997	0.739	0.355	0.145	0.101
f_{ellip}	0.0214	0.184	0.0946	0.0658	0.0250	0.327	0.0246	0.118	0.0250	0.308
f_{burst}	0.335	0.260	0.310	0.477	0.297	0.286	0.281	0.451	0.183	0.263
$f_{\text{gas,burst}}$	0.331	0.209	0.0817	0.0553	0.164	0.349	0.236	0.225	0.452	0.160
B/T_{burst}	0.672	0.538	0.889	0.890	0.517	0.367	1.00	0.215	0.928	0.409
ϵ_{\bullet}	0.0134	0.0437	0.00596	0.0363	0.00877	0.0542	0.00423	0.0407	0.0130	0.0857
η_{\bullet}	0.0163	0.0596	0.00476	0.00711	0.00188	0.0137	0.00728	0.0307	0.00788	0.0893
F_{\bullet}	0.0125	0.00818	0.00289	0.00628	0.0206	0.0233	0.00190	0.0256	0.00271	0.0164
V_{cut}	N/A	17.0	36.5	28.4	38.7	43.9	32.9	27.7	34.5	12.7
z_{cut}	N/A	10.1	11.7	10.9	12.7	12.4	10.8	11.9	12.8	10.2

Table 6 – *continued*

Parameter	Overall	Tully–Fisher	Sizes	Metallicity	M_{H_2}/L_B	Clustering	SMBHs	Local Group LF	Local Group Sizes	Local Group Z 's
Δ_0	0.716	0.722	0.720	0.724	0.715	0.714	0.716	0.722	0.718	0.722
Ω_b	0.04724	0.0441	0.0447	0.0453	0.0458	0.0470	0.0437	0.0475	0.0492	0.0467
h_0	0.691	0.724	0.698	0.720	0.711	0.688	0.699	0.710	0.682	0.685
σ_8	0.807	0.783	0.809	0.775	0.788	0.795	0.778	0.771	0.769	0.773
n_s	0.933	0.959	0.961	0.948	0.935	0.957	0.945	0.938	0.933	0.942
α_{reheat}	2.32	1.91	2.33	2.37	2.52	0.922	2.96	2.43	2.62	1.81
α_{cool}	0.550	1.06	0.0955	2.11	1.48	0.855	1.28	2.30	2.25	1.49
α_{remove}	0.102	0.125	0.917	0.334	0.146	0.466	0.848	0.0814	0.0825	0.0508
a_{core}	0.163	0.216	0.0905	0.0772	0.0281	0.105	0.222	0.127	0.0940	0.0210
A_{ac}	0.742	0.860	0.964	0.765	0.766	0.795	0.876	0.741	0.736	0.737
w_{ac}	0.920	0.817	0.809	0.945	0.989	0.871	0.919	0.908	0.928	0.985
ϵ_*	0.0152	0.0427	0.00735	0.00272	0.00329	0.0295	0.0420	0.00751	0.0322	0.0175
α_*	-3.28	-1.69	-2.83	-3.60	-3.07	-2.65	-2.51	-3.32	-2.65	-1.52
$\epsilon_{\text{d,gas}}$	0.743	0.812	0.716	0.774	0.773	0.736	0.743	0.957	0.784	0.800
$V_{\text{hot,disc}}$	358.0	389.0	341.0	497.0	449.0	353.0	393.0	374.0	452.0	543.0
$V_{\text{hot,burst}}$	328.0	370.0	125.0	498.0	496.0	341.0	271.0	507.0	533.0	467.0
α_{hot}	3.36	2.58	3.12	3.32	3.53	2.37	3.18	3.25	3.14	2.48
$\lambda_{\text{expel,disc}}$	0.785	0.551	0.412	0.283	0.380	1.06	0.646	0.438	0.659	0.622
$\lambda_{\text{expel,burst}}$	7.36	2.13	5.62	8.97	7.87	7.24	9.86	9.60	8.16	6.38
ϵ_{strip}	0.335	0.101	0.607	0.0184	0.200	0.288	0.359	0.0787	0.975	0.595
f_{ellip}	0.0214	0.308	0.360	0.0925	0.0204	0.107	0.203	0.454	0.0672	0.0212
f_{burst}	0.335	0.263	0.242	0.348	0.483	0.239	0.435	0.379	0.388	0.436
$f_{\text{gas,burst}}$	0.331	0.160	0.0937	0.171	0.264	0.361	0.120	0.410	0.225	0.450
B/T_{burst}	0.672	0.409	0.681	0.734	0.825	0.500	0.695	0.545	0.251	0.718
ϵ_{\bullet}	0.0134	0.0857	0.0232	0.0266	0.0914	0.0201	0.0560	0.0419	0.00481	0.00823
η_{\bullet}	0.0163	0.0893	0.0588	0.00928	0.0912	0.0216	0.0248	0.0139	0.0119	0.00538
F_{\bullet}	0.0125	0.0164	0.00970	0.00807	0.0293	0.00352	0.0287	0.0133	0.0279	0.00585
V_{cut}	N/A	12.7	27.2	26.5	43.0	42.9	28.8	45.5	47.5	35.5
z_{cut}	N/A	10.2	9.31	11.0	12.5	12.7	11.0	12.8	12.9	12.7

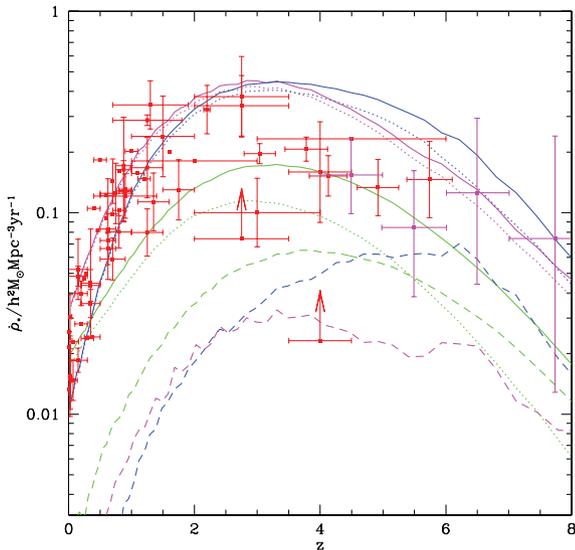


Figure 8. The star formation rate per unit comoving volume in the Universe as a function of redshift. Red points show observational estimates from a variety of sources as compiled by Hopkins (2004) while magenta points show the star formation rate inferred from gamma-ray bursts by Kistler et al. (2009). The solid lines show the total star formation rate density from our models, while the dotted and dashed lines show the contribution to this from quiescent star formation in discs and starbursts, respectively. Blue lines show the overall best-fitting model, while magenta lines indicate the best-fitting model to this data set and the green lines show results from the Bower et al. (2006) model.

result in a better fit to the star formation rate, they produce very unacceptable fits to the luminosity functions (which have too many bright galaxies) and galaxies which are far too depleted of gas.

The Bower et al. (2006) model has a much lower star formation rate density than our best-fitting model at $z > 0.5$, although it shows a comparable amount of star formation in bursts. [The Bower et al. (2006) model still manages to obtain a good match to the K -band luminosity function at $z = 0$ however by virtue of the fact that at $z \lesssim 1$, where much of the build up of stellar mass occurs, the two models have comparable average star formation rates, and because it uses a different IMF which results in a different mass-to-light ratio. Our best-fitting model produces 65 per cent more mass in stars at $z = 0$ than the Bower et al. (2006) model, but produces only 35 per cent more K -band luminosity density, as will be shown in Fig. 10, mostly from faint galaxies.] Our best-fitting model can be seen to be in significantly better agreement with the data than the Bower et al. (2006) model and nicely reproduces the sharp decline in star formation rate at low redshifts.

4.3 Luminosity functions

Luminosity functions have traditionally represented an important constraint for galaxy formation models. We therefore include a variety of luminosity functions, spanning a range of redshifts in our constraints.

Figs 9 and 10 show local ($z \approx 0$) luminosity functions from the Two-degree Field Galaxy Redshift Survey (2dFGRS) (Norberg et al. 2002; b_J band) and the Two-Micron All Sky Survey (2MASS) (Cole et al. 2001; K band), respectively, together with model predictions.

It is well established that the faint-end slope of the luminosity function, which is flatter than would be naively expected from the

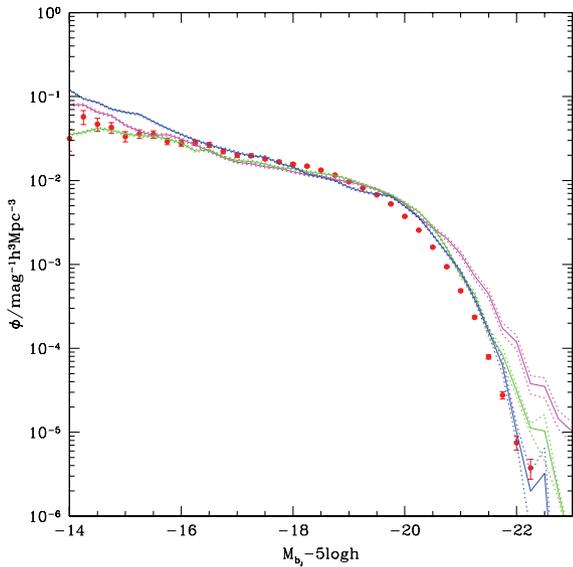


Figure 9. The $z = 0$ b_J -band luminosity function from our models: the solid lines show the luminosity function after dust extinction is applied while the dotted lines show the statistical error on the model estimate. Red points indicate the observed luminosity function from the 2dFGRS (Norberg et al. 2002). Blue lines show the overall best-fitting model, while magenta lines indicate the best-fitting model to this data set and the $z = 0$ K -band luminosity function (see Fig. 10; note that the requirement that this model be a good match to the $z = 0$ K -band luminosity function is the reason why the fit here is not as good as that of the overall best-fitting model) and green lines show results from the Bower et al. (2006) model.

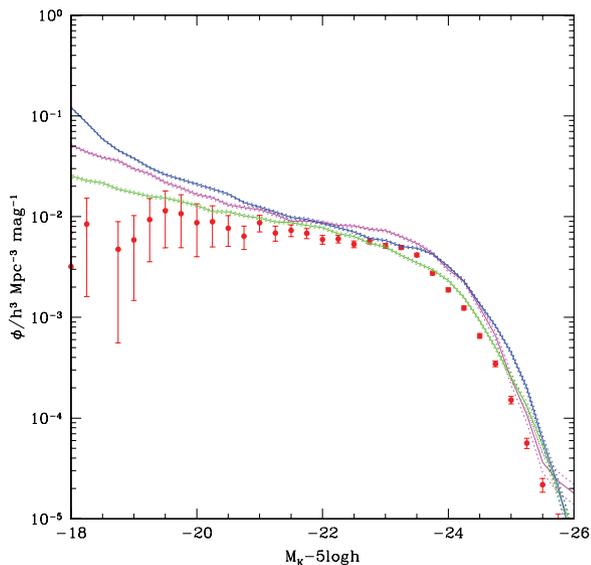


Figure 10. The $z = 0$ K -band luminosity function from our models: the solid lines show the luminosity function after dust extinction is applied while the dotted lines show the statistical error on the model estimate. Red points indicate data from the 2dFGRS+2MASS (Cole et al. 2001). Blue lines show the overall best-fitting model, while magenta indicate the best-fitting model to this data set and the $z = 0$ b_J -band luminosity function (see Fig. 9) and green show results from the Bower et al. (2006) model.

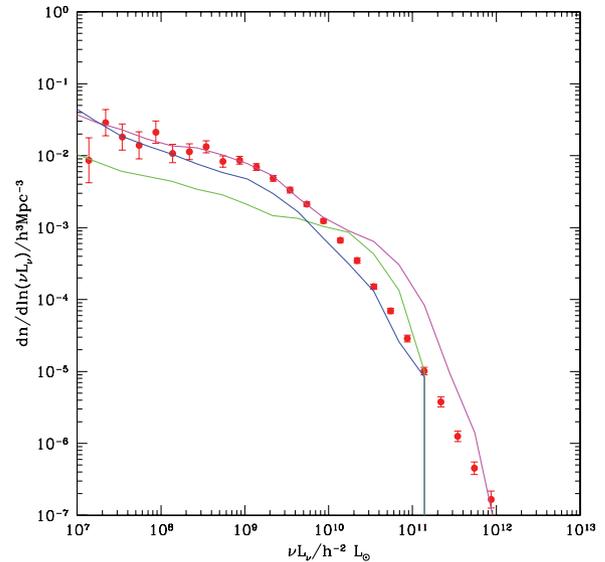


Figure 11. The $z = 0$ $60 \mu\text{m}$ luminosity functions from our models are shown by the solid lines. Red points indicate data from Saunders et al. (1990). Blue lines show the overall best-fitting model, while magenta lines indicate the best-fitting model to this data set and the green lines show results from the Bower et al. (2006) model.

slope of the dark matter halo mass function, requires some type of feedback in order to be reproduced in models. The SNe feedback present in our model is sufficient to flatten the faint-end slope of the local luminosity functions and bring it into good agreement with the data in the b_J band, except perhaps at the very faintest magnitudes shown. The K band shows an even flatter faint-end slope and this is not as well reproduced by our model.

Both our best-fitting model and the Bower et al. (2006) model produce good fits to these luminosity functions (although our best-fitting model produces a break which is slightly too bright in the K band, indicating that the galaxy colours are not quite right – see Section 4.4). This is not surprising of course as these were primary constraints used to find parameters for the Bower et al. (2006) model. The Bower et al. (2006) model does give a noticeably better match to the faint end of the K -band luminosity function (although it is far from perfect), due to the higher value of α_{hot} that it adopts (see Table 5). Unfortunately, this large value of α_{hot} adversely affects the agreement with other data sets and so our best-fitting model is forced to adopt a lower value. The important point here is that the Bower et al. (2006) model was designed to fit just these luminosity functions, while the current model is being asked to simultaneously fit a much larger compilation of data sets. This point is further illustrated by the magenta lines in Figs 9 and 10 which show the model that best matches these two data sets. It achieves a flatter faint-end slope by virtue of having quite large values of α_{hot} and α_{cool} . This improved match to the faint end is at the expense of the bright end though (χ^2 fitting gives more weight to the faint end, which has more data points with smaller error bars).

Fig. 11 shows the $60 \mu\text{m}$ infrared luminosity function from Saunders et al. (1990) (red points) and the corresponding model results (lines). The $60 \mu\text{m}$ luminosity function constrains the dust absorption and reemission in our model and so is complementary to the optical and near-IR luminosity functions discussed above. Our best-fitting model produces a very good match to the data at low luminosities – the sharp cut off at $10^{11} h^{-2} L_{\odot}$ is artificial and due to the limited number of merger trees which we are able to run and the

scarcity of these galaxies (which are produced by massive bursts of star formation). The Bower et al. (2006) model matches well at high luminosities but underpredicts the number of faint galaxies. This is due to the higher frequency of starbursts at low redshifts in the Bower et al. (2006) model (see Fig. 8), which populate the bright end of the 60 μm luminosity function. It must be kept in mind that absorption and re-emission of starlight by dust is one of the most challenging processes to model semi-analytically, and we expect that approximations made in this work may have significant effects on emission at 60 μm . A more detailed study, utilizing GRASIL, will be presented in a future work. The best-fitting model to this specific data set is a good fit to the data although it has somewhat too many 60- μm -bright galaxies. This is achieved by adopting a much lower value of $f_{\text{gas, burst}}$ which lets minor mergers trigger bursts more easily. This increases the abundance of bursting galaxies with high star formation rates and fills in the bright end of the 60 μm luminosity function.

Fig. 12 shows the K_s -band luminosity function from the K20 survey (Pozzetti et al. 2003) at $z = 1.0$. (The data at $z = 0.5$ and 1.5

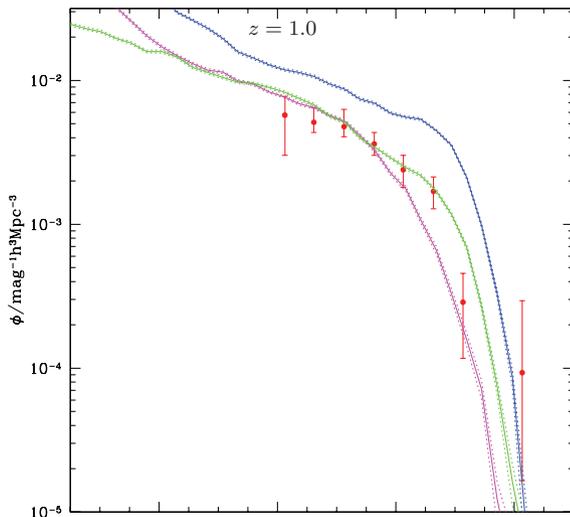


Figure 12. The $z = 1$ K_s -band luminosity function from our models is shown by the solid lines with dotted lines indicating the statistical uncertainty on the model estimates. Red points indicate data from Pozzetti et al. (2003). Blue lines show the overall best-fitting model, while magenta lines indicate the best-fitting model to this data set and the green lines show results from the Bower et al. (2006) model.

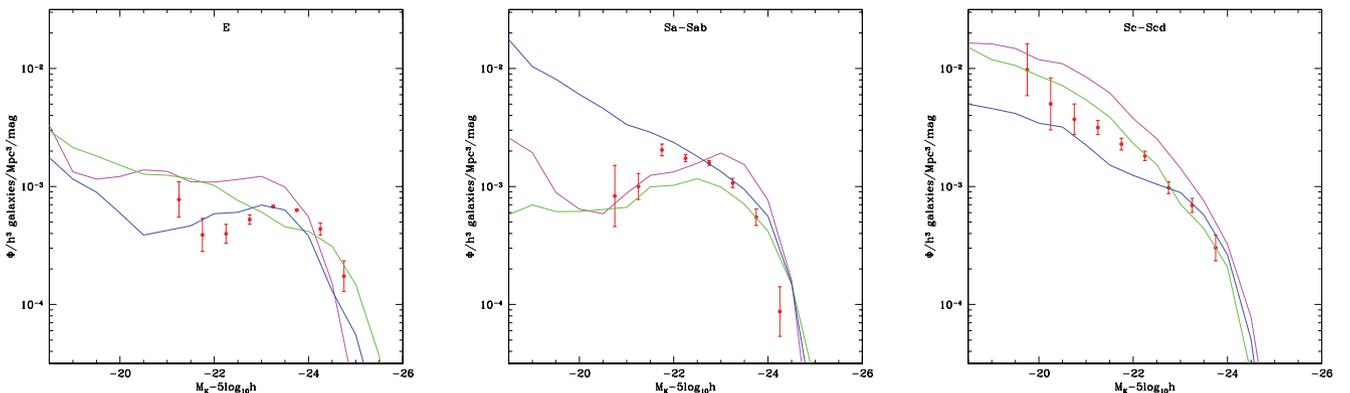


Figure 13. The $z = 0$ morphologically segregated K -band luminosity functions from our models. Points indicate the observed luminosity function from Devereux et al. (2010) for morphological classes as indicated in each panel. Blue lines show the overall best-fitting model, while magenta lines indicate the best-fitting model to this data set and the green lines show results from the Bower et al. (2006) model.

were used as constraints also.) The model traces the evolution of the luminosity function quite well but overpredicts the abundance at all redshifts. This is in contrast to the Bower et al. (2006) model which matches these luminosity functions quite well. This is partly due to the tension between luminosity functions and the star formation rate density of Fig. 8 which would be better fit if the model produced an even higher star formation rate density. This constraint forces our best-fitting model to build up more stellar mass than the Bower et al. (2006) model, consequently, to overpredict the abundance of galaxies at these redshifts. This tension between luminosity function and star formation constraints may in part be due to the difficulties involved with estimating the latter observationally [due to uncertainties in the IMF calibration of star formation rate indicators and so on; see Hopkins & Beacom (2006) for a detailed examination of these issues]. The best-fitting model to this specific data set successfully matches the data at all three redshifts. It achieves this through a combination of relatively high (i.e. less negative) α_* and a high value of α_{hot} . Together, this combination allows for a flatter faint-end slope while maintaining the normalization of the bright end.

In addition to these luminosity functions that include all galaxy types, in Fig. 13 we show the morphologically selected luminosity function of Devereux et al. (2010) overlaid with model results. We base morphological classification of model galaxies on bulge-to-total ratio (B/T) in dust-extinguished K -band light. We determine the mapping between B/T and morphology by requiring that the relative abundance of each type in the model agrees with the data in the interval $-23.5 < M_K - 5 \log_{10} h \leq -23.0$ but the morphological mix is not enforced outside this magnitude range. Our best-fitting model reproduces the broad trends seen in these data – although we find that too many Sb-Sbc galaxies are produced at the highest luminosities. The Bower et al. (2006) model gives a better match to these data overall. The best fit to the particular data set (magenta lines in Fig. 13) has a relatively large value of f_{ellip} , but is not significantly better than our best-fitting model.

In addition to these relatively low-redshift constraints, we are particularly interested here in examining constraints from the highest redshifts currently observable. Therefore, Fig. 14 shows the luminosity function of $z \approx 3$ Lyman-break galaxies together with the expectation from our best-fitting model (blue line). Model galaxies are drawn from the entire sample of galaxies at $z = 3$ found in the model. The model significantly overpredicts the number of luminous galaxies even when internal dust extinction is taken into account (the dashed line in Fig. 14 shows the luminosity function

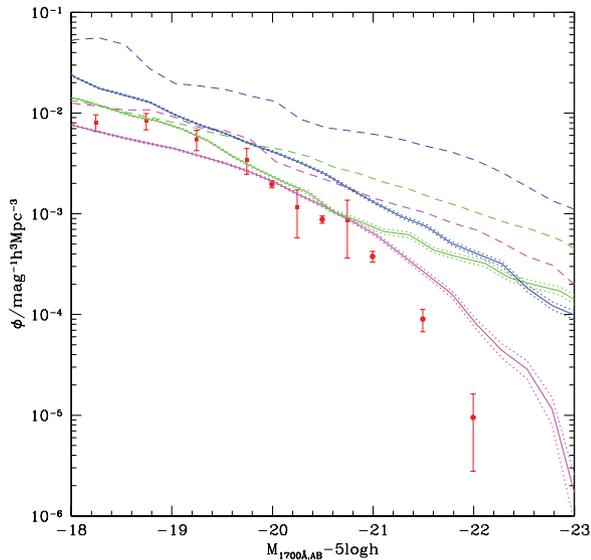


Figure 14. The $z = 3$ 1700-Å luminosity functions from our models are shown by the solid lines with dotted lines showing the statistical uncertainty on the model estimates. The dashed lines indicate the luminosity function when the effects of dust extinction are neglected. Red points indicate the observed luminosity function from Steidel et al. (1999; circles) and Dickinson (1998; squares). Blue lines show the overall best-fitting model, while magenta lines indicate the best-fitting model to this data set and the green lines show results from the Bower et al. (2006) model.

without the effects of dust extinction). The Bower et al. (2006) model gives a similarly bad match to these data at the bright-end (although is slightly better at low luminosities), producing too many highly luminous galaxies. The best-fitting model to this specific data set turns out to be not such a good fit, although it is better than either of the other models shown. The problem here is one of noise. The models run for our parameter space search utilized relatively small numbers of merger tree realizations (to permit them to run in a reasonable amount of time). In this particular case, the model run during the parameter space search looked like a good match to the $z \approx 3$ Lyman-break galaxy luminosity function, but, when re-run with many more merger trees, it turned out that the apparently good fit was partly a result of fortuitous noise. This luminosity function is particularly sensitive to such effects, as the bright end is dominated by rare starburst galaxies.

Finally, at the highest redshifts for which we presently have statistically useful data, Fig 15 shows rest-frame UV luminosity function at $z = 5$ from McLure et al. (2009). These highest redshift luminosity functions in principle place a strong constraint on the model. However, the effects of dust become extremely important at these short wavelengths and so our model predictions are less reliable. As such, these constraints are less fundamental than most of the others which we consider. We use our more detailed dust modelling for the Bower et al. (2006) model here even though the original Bower et al. (2006) used the simpler dust model of Cole et al. (2000). However, as noted in Section 2.14.1, in our current model we ensure that high- z galaxies which are undergoing near continuous instability-driven bursting are observed during the dust phase of the burst. In the Bower et al. (2006) model shown here this is not the case – such systems are almost always observed in a gas and dust free state, making them appear much brighter. It is clear that the treatment of these galaxies in terms of punctuated equilibrium

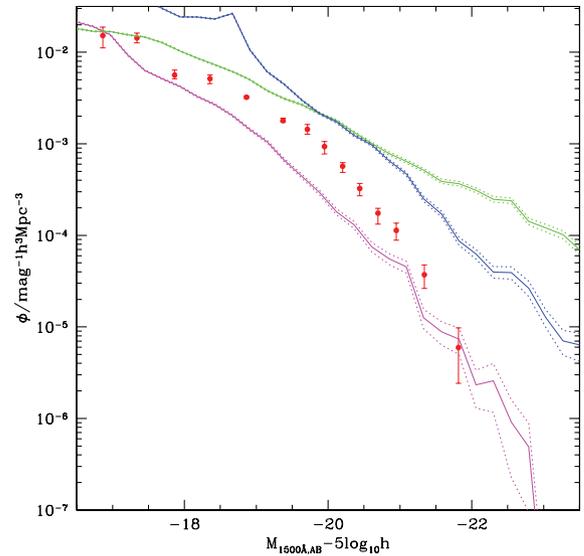


Figure 15. The $z = 5$ rest-frame 1500 Å luminosity function from our models are shown by the solid lines, with statistical errors indicated by the dotted lines. Red points indicate data from McLure et al. (2009). Blue lines show the overall best-fitting model, while magenta lines indicate the best-fitting model to this data set and the green lines show results from the Bower et al. (2006) model.

of discs is inadequate and we will return to this issue in more detail in a future work.

The best-fitting model again overpredicts the number and/or luminosities of galaxies at these redshifts. The Bower et al. (2006) model performs much worse here however – drastically overpredicting the number of luminous galaxies. The majority of this difference is due to the treatment of dust in bursts in our current model. Additionally, however, this difference simply reflects the fact that high- z constraints were not considered when selecting the parameters of the Bower et al. (2006) model – the improved agreement here illustrates the benefits of considering a wide range of data sets when constraining model parameters. The best-fitting model to these specific data sets shows a steeper decline at high luminosities and a lower normalization over all luminosities. Once again, the best fit here is not particularly good, for the same reasons that the $z = 3$ UV luminosity function is not too well fit (i.e. that the models run to search parameter space use relatively few merger trees, leading to significant noise in these luminosity functions which depend on galaxies that form in rare haloes). This is achieved through a combination of strong feedback (i.e. high $V_{\text{hot,disc}}$) and highly efficient star formation with a very strong dependence on galaxy circular velocity. However, this achieves only a relatively small improvement over the overall best-fitting model, at the expense of significantly worse fits to other data sets.

4.4 Colours

The bimodality of the galaxy colour–magnitude diagram has long been understood to convey important information regarding the evolutionary history of different types of galaxy. Recently, semi-analytic models have paid close attention to this diagnostic (Croton et al. 2006; Bower et al. 2006). In particular, Font et al. (2008) found that the inclusion of detailed modelling of ram-pressure stripping of hot gas from satellite galaxy haloes is crucial for obtaining an accurate determination of the colour–magnitude relation. That same model of ram-pressure stripping is included in the present work.

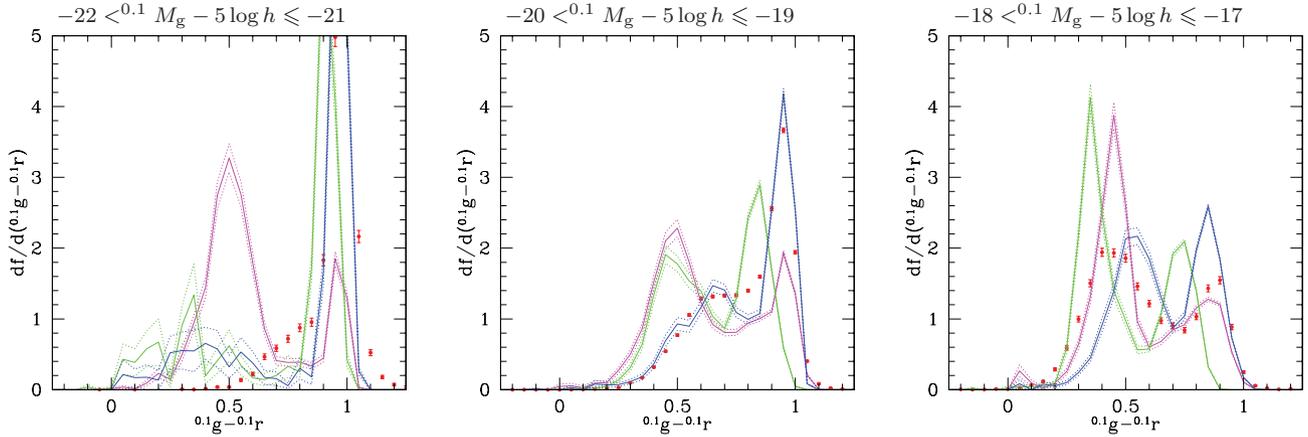


Figure 16. $0.1g-0.1r$ colour distributions for galaxies at $z = 0.1$ split by g -band absolute magnitude (see above each panel for magnitude range). Solid lines indicate the distributions from our models while the red points show data from the Sloan Digital Sky Survey (SDSS) (Weinmann et al. 2006). Blue lines show the overall best-fitting model, while magenta lines indicate the best-fitting model to this data set and the green lines show results from the Bower et al. (2006) model. Note that the magenta model is selected on the basis of more panels than are shown here.

Fig. 16 shows slices of constant magnitude through the colour–magnitude diagram of Weinmann et al. (2006), overlaid with results from our model. The model is very successful in matching these data, showing that at bright magnitudes the red galaxy component dominates, shifting to a mix of red and blue galaxies at fainter magnitudes. The median colours of the blue and red components of the galaxy population are reproduced better in our current model than by that of Bower et al. (2006), although there is clearly an offset in the blue cloud at faint magnitudes (model galaxies in the blue cloud are slightly too red). Our model reproduces the colours of galaxies reasonably well, so this offset may be partly due to the limitations of stellar population synthesis models. This problem with the Bower et al. (2006) model was noted by Font et al. (2008) who demonstrated that a combination of a higher yield of $p = 0.04$ in the instantaneous recycling approximation [Bower et al. (2006) assumed a yield of $p = 0.02$] and ram-pressure stripping of cold gas in galaxy discs lead to a much better match to galaxy colours. The yield is not a free parameter in our model, instead it is determined from the IMF and stellar metal yields directly (see Fig. 6), potentially rising as high as $p = 0.04$ after several Gyr. This is very close to the value adopted by Font et al. (2008), and our model is able to produce a good match to the colours. As we will see later (in Section 4.7), the Bower et al. (2006) model has more serious problems with galaxy metallicities which are somewhat rectified in our present model thereby helping us obtain a better match to the galaxy colours. The best-fitting model to this specific data set is a better match than our overall best-fitting model for fainter

galaxies, although it performs less well at brighter magnitudes. At faint magnitudes it produces a bluer blue-cloud which better matches that which is observed. It achieves this success by having a much larger value (i.e. less negative) of α_* . This parameter controls how star formation rates scale with galaxy mass, with this model having less dependence than any other. This improves the match to galaxy colours (at the expense of steepening the faint-end slope of the luminosity function), particularly for fainter galaxies.

4.5 Scaling relations

Fitting the Tully–Fisher relation simultaneously with the luminosity function has been a long-standing challenge for models of galaxy formation (see Dutton, van den Bosch & Courteau 2008 and references therein). Fig. 17 shows the Tully–Fisher relation from the Sloan Digital Sky Survey (SDSS) as measured by Pizagno et al. (2007) together with the result from our best-fitting model. The model is in reasonable agreement with zero-point, although somewhat offset to higher velocities, and in good agreement with the luminosity dependence and width of the Tully–Fisher relation. Our new model is a significantly better match to the Tully–Fisher relation than that of Bower et al. (2006), which produces galaxies with rotation speeds that are systematically too large (particularly for the brightest galaxies). For example, for the most luminous galaxies shown the Bower et al. (2006) predicts a population of galaxies with circular velocities of $300\text{--}400\text{ km s}^{-1}$ or greater – strongly ruled out by observations. The new model on the other hand

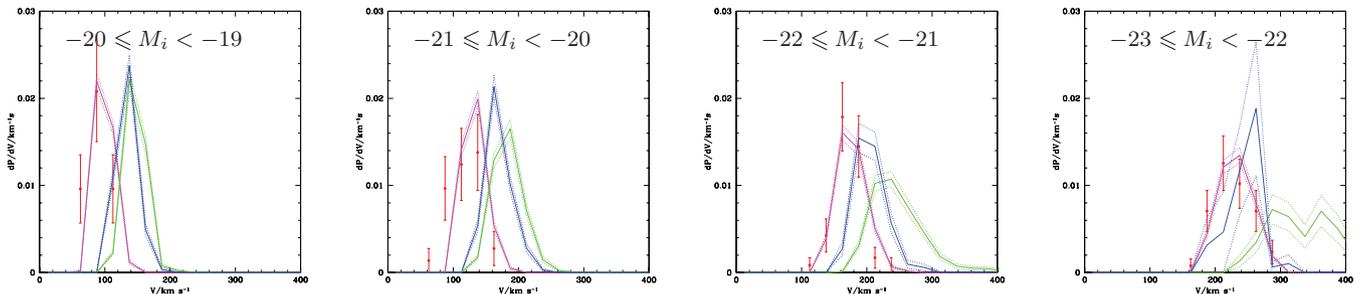


Figure 17. Slices through the i -band Tully–Fisher relation from the SDSS (Pizagno et al. 2007) at constant absolute magnitude are shown by red points. Solid lines show results from our models with dotted lines indicating the statistical error on the model estimate. Blue lines show the overall best-fitting model, while magenta lines indicate the best-fitting model to this data set and the green lines show results from the Bower et al. (2006) model.

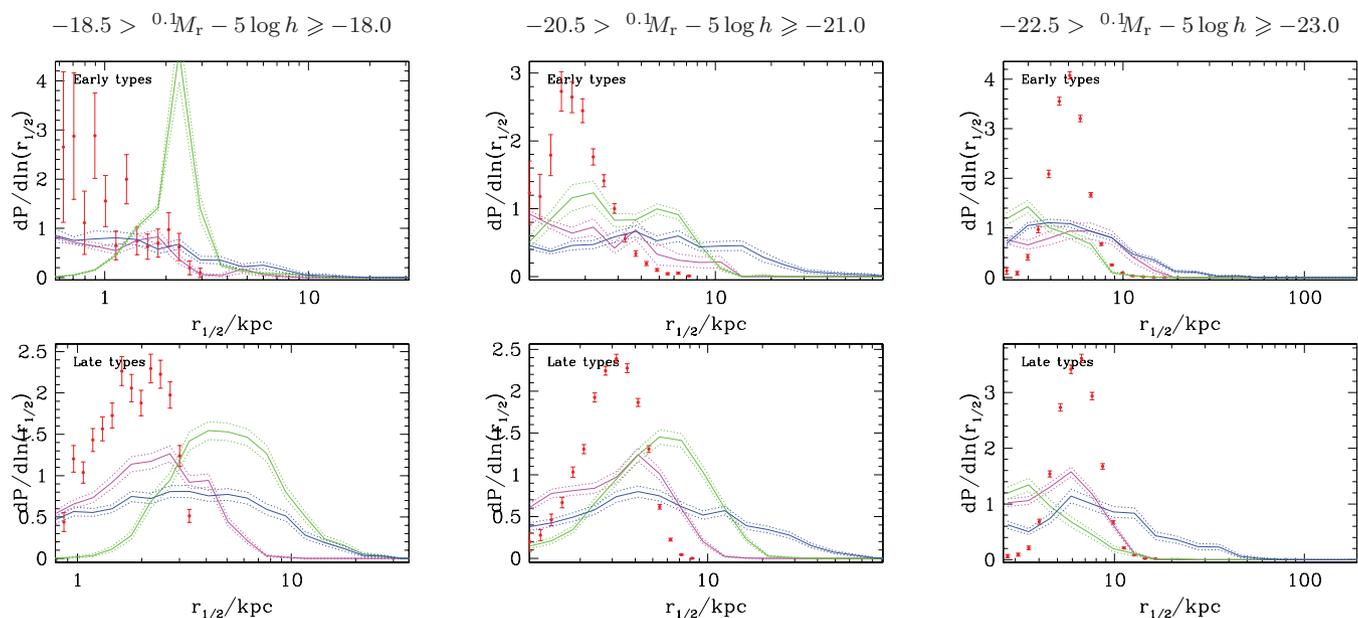


Figure 18. Distributions of galaxy half-light radii (measured in the dust-extinguished face-on r -band light profile) at $z = 0.1$ segregated by r -band absolute magnitude and by morphological class. Solid lines show results from our models while dotted lines show the statistical error on the model estimates. Red points data from the SDSS (Shen et al. 2003). Blue lines show the overall best-fitting model, while magenta lines indicate the best-fitting model to this data set and the green lines show results from the Bower et al. (2006) model.

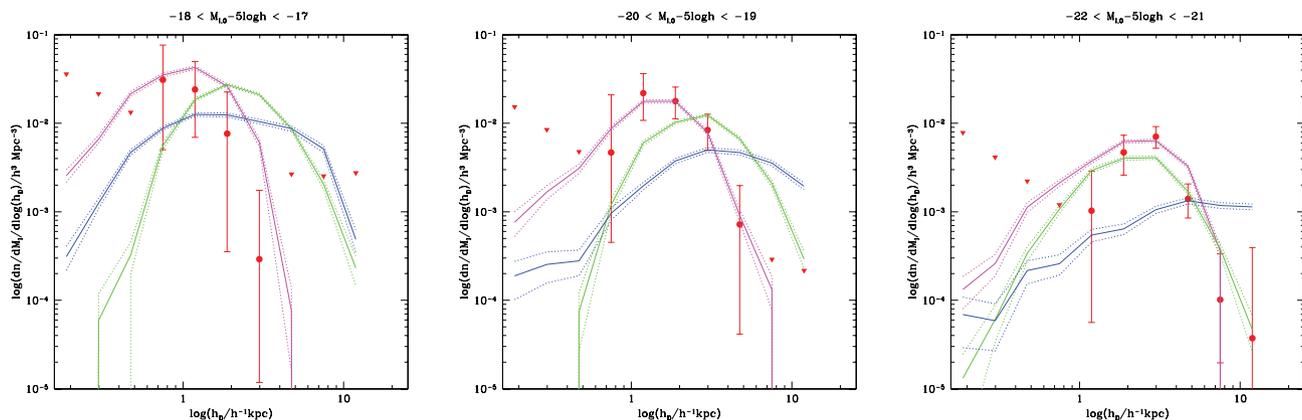


Figure 19. Distribution of disc scalelengths for galaxies at $z = 0$ segregated by face-on I -band absolute magnitude. Solid lines show results from our models while dotted lines indicate the statistical uncertainty on the model estimates. Red circles show data from de Jong & Lacey (2000) with upper limits indicated by red triangles. Blue lines show the overall best-fitting model, while magenta lines indicate the best-fitting model to this data set and the green lines show results from the Bower et al. (2006) model.

predicts essentially no galaxies in this velocity range. The best-fitting model to this particular data set is a significantly better match than our overall best-fitting model. No single parameter is responsible for the improvement, but $\lambda_{\text{expel, burst}}$ plays an important role – it is much lower in the best-fitting model to the Tully–Fisher data.

4.6 Sizes

Fig. 18 shows the distribution of galaxy sizes, split by morphological type and magnitude, from the SDSS (Shen et al. 2003). To morphologically classify model galaxies we utilize the bulge-to-total ratio in dust-extinguished $0.1r$ -band light. From the K -band morphologically segregated luminosity function (see Section 4.3) we find that E and S0 galaxies are those with $B/T > 0.714$ for the best-fitting model. There is no convincing reason to expect this value to correspond precisely to the morphological selection used

by Shen et al. (2003), but it is currently our best method to choose a division between early and late types in our model. For simplicity, we employ the same morphological cut for all three models plotted in Fig. 18. Model results are overlaid as lines. Model galaxies are too large compared to the data, by factors of about 2, and the distribution of model galaxy sizes is too broad. This problem is more significant for the fainter galaxies.

Fig. 19 shows the distribution of disc sizes from de Jong & Lacey (2000) with model results overlaid as lines. This permits a more careful comparison with the model as it does not require us to assign morphological types to model galaxies. Model discs are somewhat too large in all luminosity bins considered, and the width of the distribution of disc sizes is broader than that observed.

The Bower et al. (2006) model produces galaxies which are systematically smaller than those in our current best-fitting model at bright magnitudes, but larger at faint magnitudes. It also produces

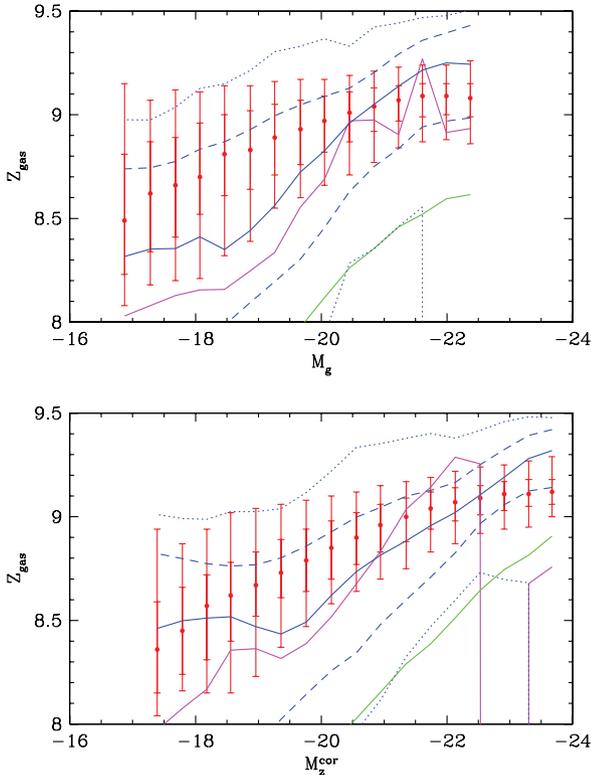


Figure 20. Gas-phase metallicity as a function of absolute magnitude from the SDSS (Tremonti et al. 2004) is shown by the red points. Points show the median value, while error bars indicate the 2.5, 16, 84 and 97.5 percentiles of the distribution. Lines indicate results from our best-fitting model. Solid lines indicate the median model relation, dashed lines the 16 and 84 percentiles and dotted lines the 2.5 and 97.5 percentiles, corresponding to the error bars on the data. Blue lines show the overall best-fitting model, while magenta lines indicate the best-fitting model to this data set and the green lines show results from the Bower et al. (2006) model. (Note that dashed and dotted lines are shown only for the best-fitting model for clarity.)

a narrower distribution of disc sizes. Our best-fitting model to these combined size data sets is a rather poor match to the distribution of disc sizes. We find that it is challenging to obtain realistic sizes for discs in our model while simultaneously matching other observational constraints. This problem, which may reflect inaccuracies in the angular momentum of cooling gas, angular momentum loss during cooling or merging, or internal processes which transfer angular momentum out of galaxies will be addressed in greater detail in a future work.

4.7 Gas and metal content

The star formation and SNe feedback prescriptions in our model can be constrained by measurements of the gas and metal content of galaxies. Fig. 20 shows the distribution of gas-phase metallicities from the SDSS (Tremonti et al. 2004) compared with results from our best-fitting model. Model galaxies are drawn from the entire population of galaxies at $z = 0.1$. Tremonti et al. (2004) select star-forming galaxies – essentially those with well-detected $H\beta$, $H\alpha$ and $[\text{NII}] \lambda 6584$ lines – and also reject galaxies with a significant AGN component. We have not attempted to reproduce these observational

selection criteria here,²⁵ but note that excluding galaxies with very low star formation rates makes negligible difference to our results. The model clearly produces a strong trend of increasing metallicity with increasing luminosity, just as is observed, although the relation is somewhat too steep, resulting in metallicities which are around a factor of 2 too low at the lowest luminosities plotted. This relation is driven, in the model, by SNe feedback: in low-luminosity galaxies feedback is more efficient at ejecting material from a galaxy making it less efficient at self-enriching. The trend is somewhat steeper in the model than is observed and therefore underpredicts the metallicity of low-luminosity galaxies. The spread in metallicity at fixed luminosity is larger than that which is observed. The best-fitting model to the metallicity data sets presented in this section can be seen to actually be a worse fit to the gas phase metallicity, a consequence of tensions between fitting these data and stellar metallicities and gas fractions.

Fig. 21 shows distributions of mean stellar metallicity in various bins of absolute B -band magnitude. Data, shown by points, are taken from Zaritsky et al. (1994), while results from our best-fitting model are shown by lines. For model galaxies, we plot the luminosity-weighted mean metallicity of all stars (i.e. both disc and bulge stars). Although the data are quite noisy, there is, in general, good agreement of the model with these data. The Bower et al. (2006) model fails to match the scaling of metallicity with stellar mass seen in these data. An increase in the yield in this model (from $p = 0.02$ to $p = 0.04$ as required to better match galaxy colours; Font et al. 2008) would improve this situation significantly, but some reduction in the dependence of SNe feedback on galaxy mass is likely still required to obtain the correct scaling.

Finally, Fig. 22 shows the distribution of gas-to-light ratios from a compilation of data compared to results from our best-fitting model. Model galaxies are selected to have bulge-to-total ratios in B -band light of 0.4 or less and gas fractions of 3 per cent or more in order to attempt to match the morphological selection (Sa and later types) in the observations. The results are somewhat sensitive to the morphological criteria used, a fact which must be taken into account when considering the comparison with the observational data. The model ratio is somewhat too high (too much gas per unit light), but displays approximately the correct dispersion. The Bower et al. (2006) model gets closer to the observed mean for bright galaxies, but shows a dramatic downturn at low luminosities (a result of its very strong SNe feedback). The best-fitting model to this specific data set is an excellent match to both the mean and dispersion in the gas fraction data. This is achieved primarily via a very low efficiency of star formation (allowing gas fractions to stay high) coupled with strongly velocity-dependent feedback which helps obtain the measured slope in this relation.

Overall, the Bower et al. (2006) performs much less well in matching metallicity and gas content properties. This problem can be traced to the very strong scaling of SNe feedback strength with galaxy circular velocity adopted in the Bower et al. (2006) model and the low yield. This strongly suppresses the effective yield in low-mass galaxies, resulting in them being too metal poor, and likewise strongly suppresses the gas content of those same low-mass galaxies. These constraints are among the primary drivers causing our best-fitting model to adopt a lower value of α_{hot} .

²⁵Both because we cannot, at present, include the AGN component in the spectra and because it would involve constructing mock catalogues which is too expensive during our parameter space search.

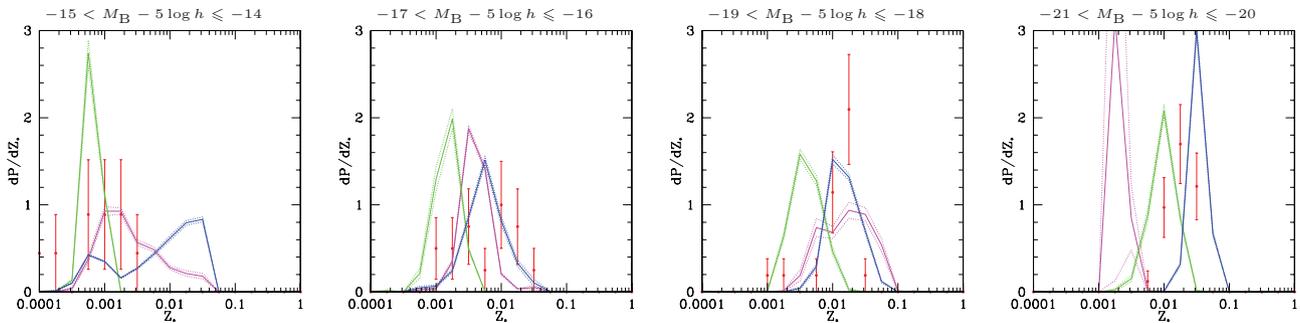


Figure 21. Distributions of mean stellar metallicity at different slices of absolute magnitude. Red points show observational data compiled by Zaritsky, Kennicutt & Huchra (1994). Solid lines indicate results from our models while dotted lines show the statistical uncertainty on the model estimate. Blue lines show the overall best-fitting model, while magenta lines indicate the best-fitting model to this data set and the green lines show results from the Bower et al. (2006) model.

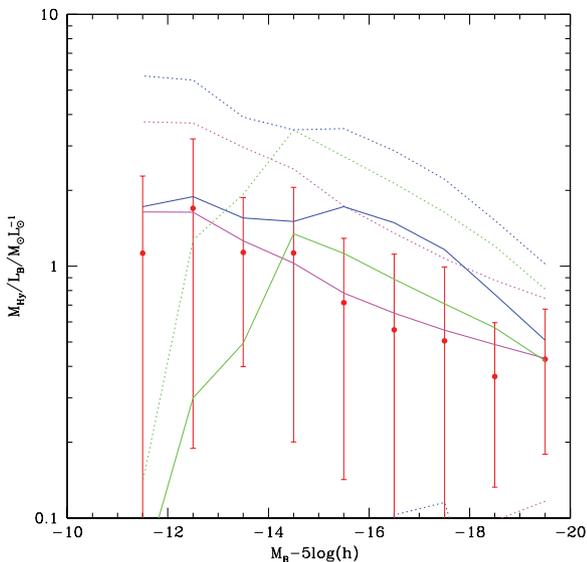


Figure 22. Gas (hydrogen) to B -band light ratios at $z = 0$ as a function of B -band absolute magnitude. The solid lines show the mean ratio from our models while the dotted lines show the dispersion around the mean. Red points show the mean ratio from a compilation of data from Huchtmeier & Richter (1988) and Sage (1993) with error bars indicating the dispersion in the distribution. Blue lines show the overall best-fitting model, while magenta lines indicate the best-fitting model to this data set and the green lines show results from the Bower et al. (2006) model. Model galaxies were selected to have bulge-to-total ratios in B -band light of 0.4 or less and gas fractions of 3 per cent or more in order to attempt to match the morphological selection (Sa and later types) in the observations.

4.8 Clustering

Galaxy clustering places strong constraints on the occupancy of galaxies within dark matter haloes and, therefore, the merger rate (amongst other things). To compute the clustering properties of galaxies we make use of the fact that halo occupation distributions are naturally predicted by the GALFORM model. We therefore extract halo occupation distributions directly from our best-fitting model. We then employ the halo model of galaxy clustering (Cooray & Sheth 2002) to compute two-point correlation functions in redshift space. These are compared to measured redshift-space correlation functions from the 2dFGRS (Norberg et al. 2002) in Fig. 23.

There is excellent agreement between the model and data on large scales (where the two halo term dominates). On small scales, in the one halo regime, the model systematically overestimates the corre-

lation function. This discrepancy, which is due to the model placing too many satellite galaxies in massive haloes, has been noted and discussed previously by Kim et al. (2009). In their study, Kim et al. (2009) demonstrated that this problem might be resolved by invoking destruction of satellite galaxies by tidal forces and by accounting for satellite-satellite mergers (both processes reduce the number of satellites). The current model includes both of these processes and treats them in a significantly more realistic way than did Kim et al. (2009). We find that they are not enough to bring the model correlation function into agreement with the data on small scales (although they do help), in our particular model. This may indicate that these processes have not been modelled sufficiently accurately, or that our model simply begins with too many satellites. We note that the Bower et al. (2006) model performs similarly well on large scales and somewhat better on small scales (the stronger feedback in this model helps reduce the number of satellite galaxies of a given luminosity in high-mass haloes), although it still overpredicts the small-scale clustering, as has been noted by Kim et al. (2009). The best-fitting model to the clustering data alone is not very successful. This is again due to the difficulty of computing accurate correlation functions using the relatively small sets of merger trees that we are able to utilize for parameter space searches, and serves as an excellent example of the need to include better estimates of the model uncertainty (i.e. the variance in predictions from the model due to the limited number of merger trees utilized) when computing goodness of fit measures.

4.9 Supermassive black holes

The inclusion of AGN feedback in semi-analytic models of galaxy formation necessitates the inclusion of the supermassive black holes that are responsible for that feedback. As such, it is important to constrain the properties of these black holes to match those that are observed. Fig. 24 shows the distribution of supermassive black hole masses in three slices of galaxy bulge mass. Points show observational data from Häring & Rix (2004) while lines show results from our best-fitting model. The model is in excellent agreement with the current data. The Bower et al. (2006) model produces nearly identical results for the black hole masses. This is not surprising since, as pointed out by Bower et al. (in preparation), the F_{\bullet} parameter can be adjusted to achieve a good fit here without significantly affecting any other predictions. For this same reason, the best-fitting model to these black hole data is not significantly better than either the Bower et al. (2006) or the overall best-fitting model.

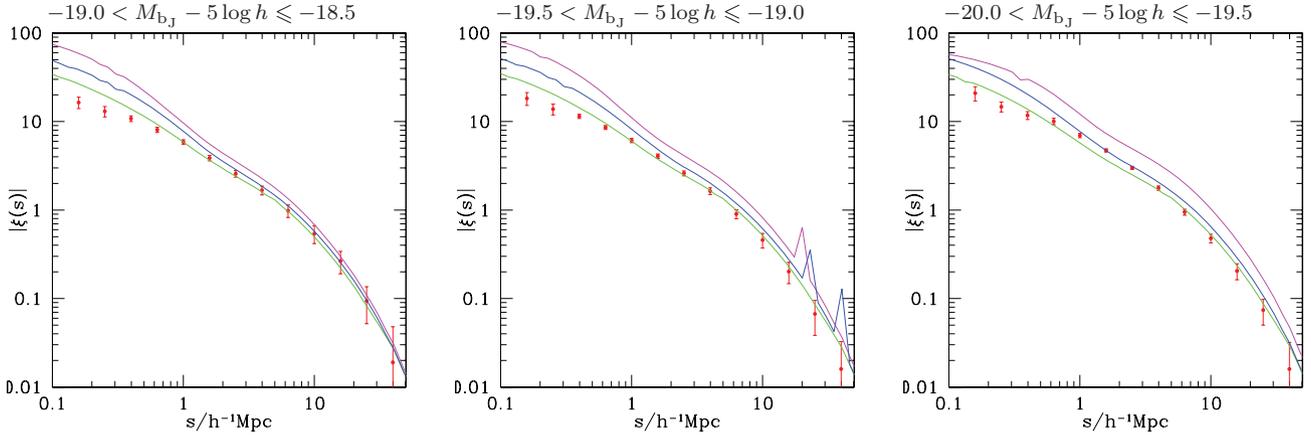


Figure 23. Redshift space two-point correlation functions of galaxies selected by their b_J -band absolute magnitude. Solid lines show results from our models while red points indicate data from the 2dFGRS (Norberg et al. 2002). Model correlation functions are computed using the halo model of clustering (Cooray & Sheth 2002) with the input halo occupation distributions computed directly from our best-fitting model. Blue lines show the overall best-fitting model, while magenta lines indicate the best-fitting model to this data set and the green lines show results from the Bower et al. (2006) model.

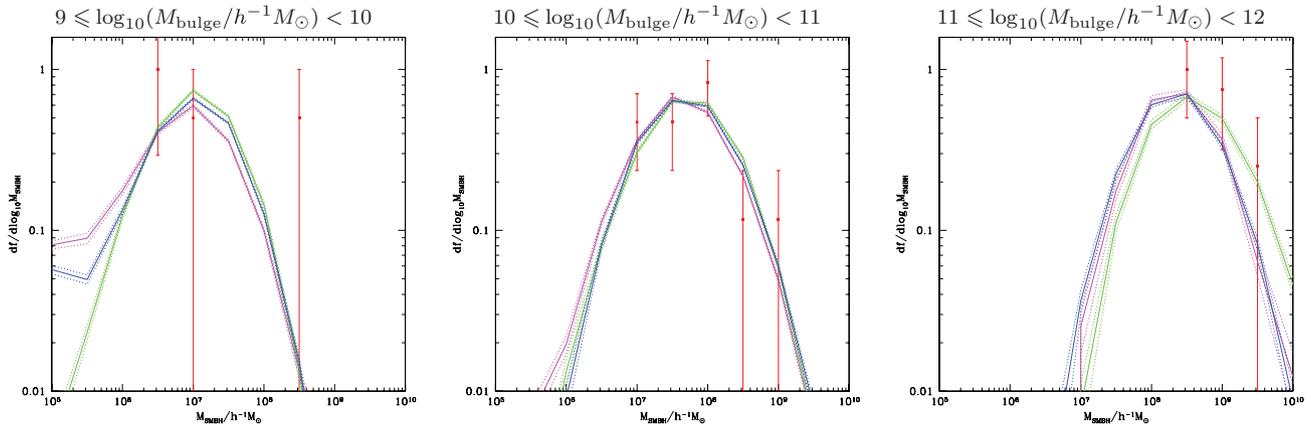


Figure 24. The distribution of supermassive black hole mass in three slices of galaxy bulge mass. Data are taken from Häring & Rix (2004) and are shown by red points. Solid lines indicate results from our models with dotted lines showing the statistical uncertainty on the model estimate. Blue lines show the overall best-fitting model, while magenta lines indicate the best-fitting model to this data set and the green lines show results from the Bower et al. (2006) model.

4.10 Local Group

The recent discovery of several new satellite galaxies of the Milky Way has led to their abundance and properties being more robustly known and therefore acting as a strong constraint on models of galaxy formation and has attracted significant attention recently (Bullock, Kravtsov & Weinberg 2000; Benson et al. 2002a; Somerville 2002; Gnedin & Kravtsov 2006; Madau, Diemand & Kuhlen 2008a; Madau et al. 2008b; Muñoz et al. 2009; Bovill & Ricotti 2009; Busha et al. 2009; Macciò et al. 2010). Our model is the only one of which we are aware that follows the formation of these galaxies within the context of a self-consistent model of the IGM and the global galaxy population which fits a broad range of experimental constraints on galaxies and the IGM.

To compute the expected properties of Milky Way satellites in our model we simulate a large number of dark matter haloes with masses at $z=0$ in the range $2 \times 10^{11} - 3 \times 10^{12} h^{-1} M_\odot$. From these, we select only those haloes with a virial velocity in the range $125 - 180 \text{ km s}^{-1}$ (consistent with recent estimates; Dehnen, McLaughlin & Sachania 2006; Xue et al. 2008) and which contain a central galaxy with a bulge-to-total ratio between 5 and 20 per cent to approximately match the properties of the Milky Way. This step is potentially important, as it ensures that the satellite populations

that we consider are consistent with the formation of a Milky Way-like galaxy.²⁶ In practice, we find that the morphological selection has little effect on the satellite luminosity function. However, the selection of suitable haloes based on virial velocity produces a significant reduction (by about a factor of 2) in the number of satellites compared to the common practice of selecting haloes with masses of approximately $10^{12} h^{-1} M_\odot$. Halo selection is clearly of great importance when addressing the missing satellite problem. We prefer to use a selection on halo virial velocity here rather than a selection on galaxy stellar mass, as was used by Benson et al. (2002a) for example, since we know that the Tully–Fisher relation in our model is incorrect (see Section 4.5) and so selecting on galaxy mass would result in an incorrect sample of halo masses.

Fig. 25 shows the V -band luminosity function of Milky Way satellite galaxies from our best-fitting model compared with the latest observational estimate. Our model is able to produce a sufficient

²⁶The merging history of a halo will affect both the properties of the central galaxy and the population of satellite galaxies. By selecting only haloes whose merger history was suitable to produce a Milky Way we ensure that we are looking only at satellite populations consistent with the presence of such a galaxy.

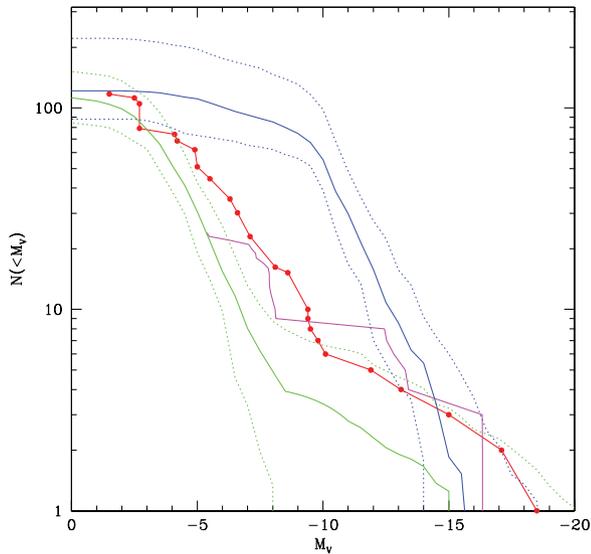


Figure 25. The luminosity function of Local Group satellite galaxies in our models. Red points show current observational estimates of the luminosity function from Koposov et al. (2008) including corrections for sky coverage and selection probability from Tollerud et al. (2008). Solid lines show the median luminosity functions of model satellite galaxies located in Milky Way-hosting haloes, while dotted lines indicate the 10th and 90th percentiles of the distribution of model luminosity functions. Blue lines show the overall best-fitting model, while magenta lines indicate the best-fitting model to this data set and the green lines show results from the Bower et al. (2006) model.

number of the brightest satellites in a small fraction of realizations, although the median lies below the observed luminosity function for the Milky Way. At lower luminosities, our best-fitting model overpredicts the observed number of satellites by factors of up to 5. It has recently been pointed out (Busha et al. 2009; Font et al., in preparation) that inhomogeneous reionization (namely the reionization of the Lagrangian volume of the Milky Way halo by Milky

Way progenitors) is an important consideration when computing the abundance of Local Group satellites. In particular, Font et al. (in preparation) find a similar level of discrepancy in the luminosity function when they ignore this effect (as we do here) and use a similar feedback model, but demonstrate that consideration of inhomogeneous reionization can reconcile the predicted and observed abundance of satellites. We do not consider inhomogeneous reionization here, but will return to it in greater detail in a future work. It must be noted, however, that this may have an impact on the luminosity function of Local Group satellites. The Bower et al. (2006) model gives a reasonably good match to the data, producing slightly fewer satellites than are observed at all luminosities. The best-fitting model to this specific data set is in good agreement with the observations down to $M_V = -5$, but fails to produce fainter satellites. (It also produces very few halo/galaxy pairs which meet our criteria to be deemed ‘Milky Way-like’, resulting in poor statistics for this model. The models utilized during the parameter space search happened to produce more faint galaxies, resulting in them being judged a good fit – this is another example of where understanding the model uncertainty is of crucial importance.)

Fig. 26 shows the distribution of half-mass radii for Milky Way satellites split into four bins of V -band absolute magnitude (only two of the bins are shown). The data are sparse, but the model produces galaxies that are too small compared to the observed satellites by factors of around 3–6. The Bower et al. (2006) model has the opposite problem, producing faint satellites that are too large but doing well at matching the sizes of brighter satellites. The best-fitting model to the Local Group size data alone is not significantly better than the overall best-fitting model – the sizes tend to be rather insensitive to most parameters.

Fig. 27 shows the distribution of stellar metallicities for Milky Way satellites split into the same four bins of V -band absolute magnitude (of which only two are shown). Once again, the data are sparse, but the model is seen to predict distributions of metallicity that are too broad compared to those observed. The Bower et al. (2006) model performs poorly here, significantly underestimating

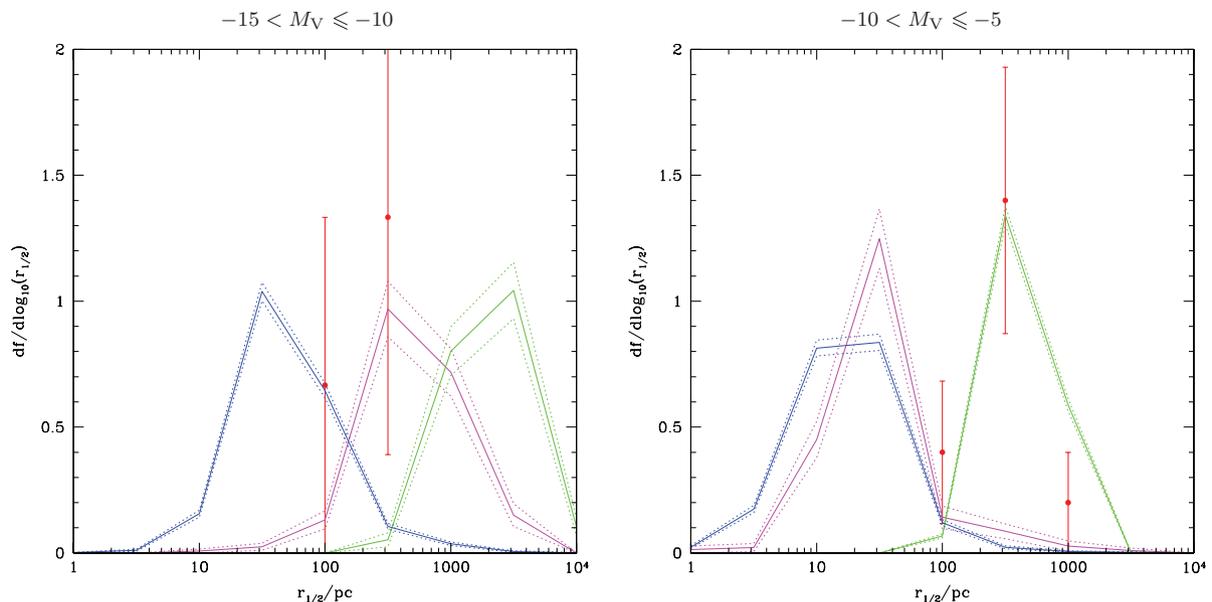


Figure 26. The size distribution of Local Group satellite galaxies in our models. Red points show current observational estimates of the size distribution from Tollerud et al. (2008). Solid lines show the size distribution of model satellite galaxies located in Milky Way-hosting haloes with dotted lines showing the statistical uncertainty on the model estimate. Blue lines show the overall best-fitting model, while magenta lines indicate the best-fitting model to this data set and the green lines show results from the Bower et al. (2006) model.

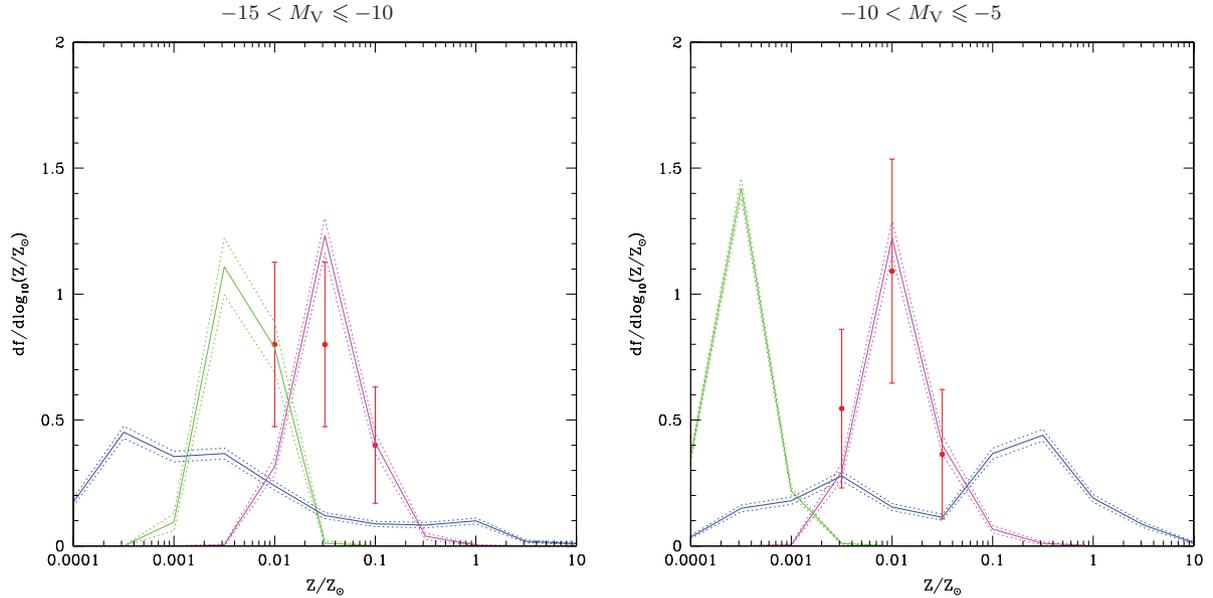


Figure 27. The metallicity distribution of Local Group satellite galaxies in our models. Red points show current observational estimates of the metallicity distribution from the compilation of Mateo (1998) and from Kirby et al. (2008). Solid lines show the metallicity distribution of model satellite galaxies located in Milky Way-hosting haloes with dotted lines showing the statistical uncertainty on the model estimate. Blue lines show the overall best-fitting model, while magenta lines indicate the best-fitting model to this data set and the green lines show results from the Bower et al. (2006) model.

the metallicities of the fainter satellites. This problem can be directly traced to the high value of α_{hot} used by the Bower et al. (2006) model which results in exceptionally strong SNe feedback, and consequently very low effective yields, for low-mass galaxies. The best-fitting model to the Local Group metallicity data alone performs much better than the Bower et al. (2006) and significantly better than the overall best-fitting model in reproducing both the trend with luminosity and scatter at fixed luminosity. This is achieved through a combination of relatively weakly velocity-dependent feedback (i.e. a low value of α_{hot}) and a weak scaling of star formation efficiency with velocity. Together, these parameters determine the trend of effective yield with mass and the degree of self-enrichment in these galaxies. However, this weaker feedback and low α_{hot} also result in a steeper faint-end slope for the global luminosity function compared to Bower et al. (2006), thereby giving less success in matching the data in that particular statistic.

4.11 IGM evolution

As described in Section 2.10, our model self-consistently evolves the properties of the IGM along with those of galaxies. In this section we discuss basic properties of the IGM (and related quantities) from our best-fitting model.

Photoheating of the IGM begins to raise its temperature above the adiabatic expectation at $z \approx 25$, reaching a peak temperature of approximately 15 000 K when hydrogen becomes fully reionized before cooling to around 2000 K by $z = 0$. Hydrogen is fully reionized by $z = 8$. Helium is singly ionized at approximately the same time. There follows an extended period during which helium is partially doubly ionized, but is not fully doubly ionized until much later, around $z = 4$.

Fig. 28 shows the Gunn–Peterson (Gunn & Peterson 1965) and electron scattering optical depths as a function of redshift. The Gunn–Peterson optical depth rises sharply at the epoch of reionization becoming optically thick at $z = 8$. The rise in Gunn–Peterson optical depth is offset from that seen in observations of high-redshift

quasars, suggesting that reionization of hydrogen occurs somewhat too early in our model, although Becker, Rauch & Sargent (2007) have argued that this trend in optical depth does not necessarily coincide with the epoch of reionization, but is instead consistent with a smooth extrapolation of the Lyman- α forest from lower redshifts (our model does not include the Lyman α forest). The electron scattering optical depth is an excellent match to that inferred from *WMAP* observations of the cosmic microwave background (i.e. consistent within the errors) suggesting that our model reionizes the Universe at the correct epoch.

One of the key effects of the reionization of the Universe is to suppress the formation of galaxies in low-mass dark matter haloes. We find that the accretion temperature, T_{acc} , remains approximately constant at around 30 000 K below $z = 3$, corresponding to a mass scale increasing with time. The filtering mass rises sharply during reionization and remains large until the present day.

We note that the model predicts too much flux at 912 Å in the photon background. We suspect that this is due to the fact that our IGM model is uniform. Inclusion of a non-uniform IGM (i.e. the Lyman α forest) would result in a greater mean optical depth and would reduce the model flux.

4.12 Additional results

In this section, we present two additional results that were not used to constrain the model, and therefore represent predictions.

4.12.1 Gas phases

While not included in our fitting procedure, it is interesting to examine the distribution of gas between different phases as a function of dark matter halo mass. Fig. 29 shows the fraction of baryons in hot (including reheated gas), galaxy (cold gas in discs plus stars in discs and spheroids) and ejected (lost from the halo) phases. The Bower et al. (2006) model (which has no ejected material) shows

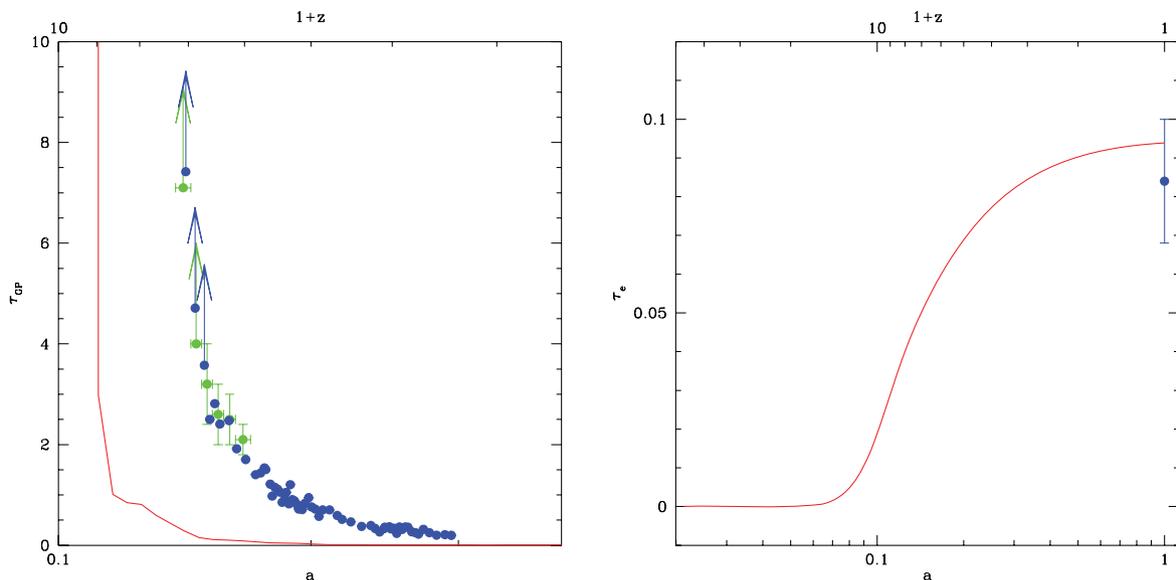


Figure 28. Left-hand panel: the Gunn–Peterson (Gunn & Peterson 1965) optical depth as a function of expansion factor and redshift in our best-fitting model. Points show observational constraints from Songaila (2004; blue points) and Fan et al. (2006; green points). Right-hand panel: the electron scattering optical depth to the CMB as a function of redshift in our best-fitting model. The blue point shows the *WMAP* 5 constraint (Dunkley et al. 2009).

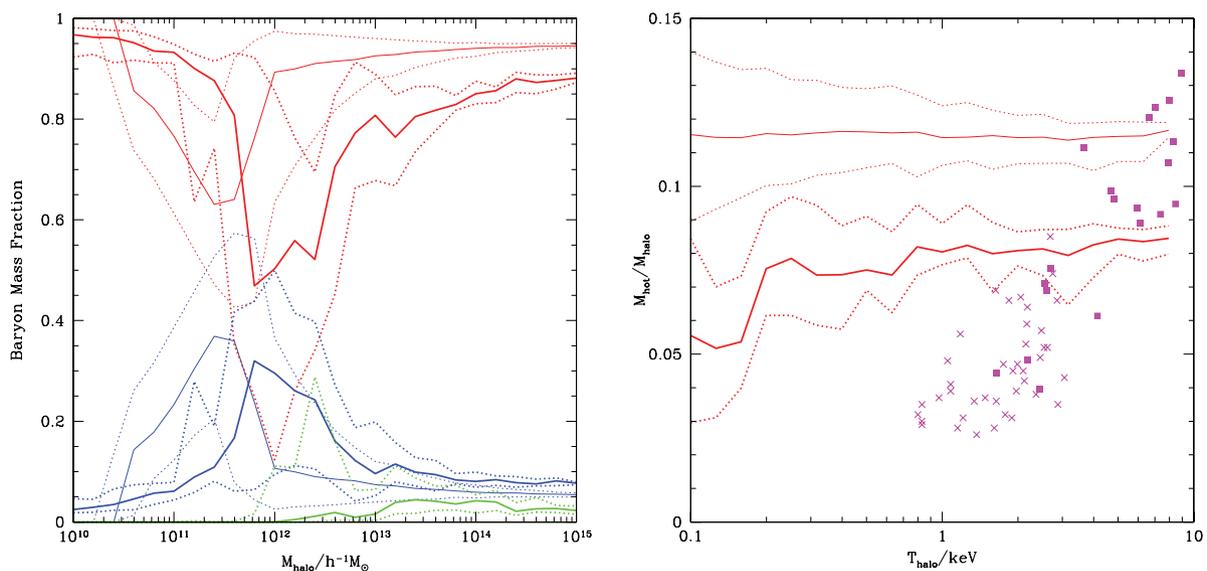


Figure 29. Left-hand panel: solid lines show the median fraction of baryons in different phases as a function of halo mass, while dotted lines indicate the 10th and 90th percentiles of the distribution. Red lines show gas in the hot phase (which includes any gas in the M_{reheated} reservoir), blue lines gas in the galaxy phase and green lines gas which has been ejected from the halo. Thin lines indicate results from the Bower et al. (2006) model while thick lines show results from the best-fitting model used in this work. Right-hand panel: the ratio of hot gas mass to total halo mass as a function of halo virial temperature is shown by the solid red line. Magenta points show data from Sun et al. (2009) (crosses) and Vikhlinin et al. (2009) (squares). Both the observed data and the model results are measured within r_{2500} (the radius enclosing an overdensity of 2500). These data were not included as constraints in our search of the model parameter space.

a peak in galaxy phase fraction at $M_{\text{halo}} \approx 2 \times 10^{11} h^{-1} M_{\odot}$ with a rapid decline to lower mass and asymptoting to a constant fraction of 5 per cent in higher mass haloes. This follows the general trend found in semi-analytic models (see e.g. Benson et al. 2000b) in which SNe feedback suppresses galaxy formation in low-mass haloes, while inefficient cooling and AGN feedback do the same in the highest mass haloes. In contrast, our best-fitting model shows modest ejection of gas in massive haloes and a corresponding suppression in the hot gas fraction, although the trends are qualitatively the same as in Bower et al. (2006). This is different from the de-

pendence of hot gas fraction on halo mass found by Bower et al. (2008) – our current model produces less ejection than found by Bower et al. (2008) resulting in the hot gas fraction being too high in intermediate-mass haloes. In particular, the right-hand panel of Fig. 29 shows the gas fraction in model haloes as a function of hot gas temperature. Model gas fractions were computed within a radius enclosing an overdensity of 2500, just as were the observed data. This radius, and the gas fraction within it, is computed using the dark matter and gas density profiles described in Sections 2.5.4 and 2.6.3, respectively. Compared to the data (magenta points), the

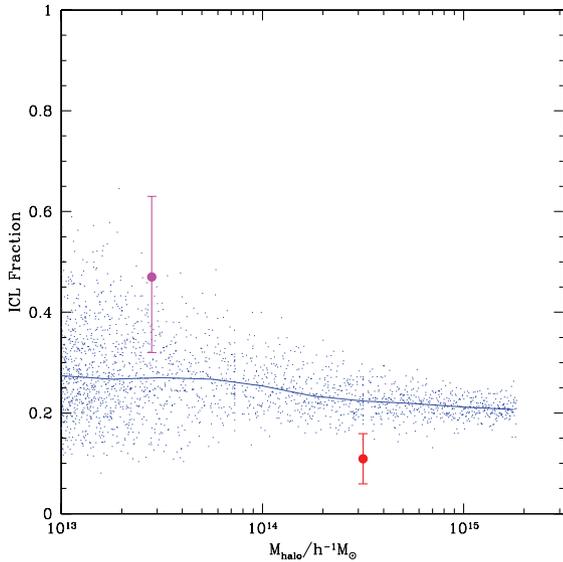


Figure 30. The fraction of stars which are part of the intrahalo light as a function of halo mass. Blue points show individual model haloes, while the blue line shows the running median of this distribution. The magenta and red points indicate the observational determinations of McGee & Balogh (2010) and Zibetti et al. (2005) for groups and clusters, respectively.

Bower et al. (2006) model is a very poor match, showing almost no trend with temperature. Our best-fitting model also performs poorly, and it is clear that the suppression in hot gas fraction does not have the correct dependence on halo mass.²⁷ In contrast, the Bower et al. (2008) model produced an excellent match to these data (as it was designed to do). We therefore expect that our best-fitting model will not give a good match to the X-ray luminosity–temperature relation, and would instead require more efficient ejection, with a stronger dependence on halo mass in the relevant range, to achieve a good fit. We reiterate that these data were not included as a constraint when searching parameter space for the best-fitting model. We will return to this issue in future work, including these constraints directly.

4.12.2 Intrahalo light

Stars that are tidally stripped from model galaxies become part of a diffuse intrahalo component which we assume fills the host halo. We can therefore predict the fraction of stars which are found in this intrahalo light as a function of halo mass and compare it to measurements of this quantity. Zibetti et al. (2005) have measured this quantity for clusters, while McGee & Balogh (2010) have measured it for galaxy groups. In Fig. 30, we show their results overlaid on results from our model. Blue points show individual model haloes, while the blue line shows the running median of this distribution. The magenta and red points indicate the above mentioned observational determinations for groups and clusters, respectively. Our model predicts an intrahalo light fraction which is a very weak function of halo mass, remaining at 20–25 per cent over two orders of magnitude in halo mass. At fixed halo mass, there is significant scatter, particularly for the lower mass haloes. Our predictions are in agreement with the current observational determinations, given

²⁷Given the hot gas profile assumed in our model and the baryon fraction, the largest ratio of hot gas to dark matter mass we could find here in massive haloes is 0.10 (since the gas profile is cored, but the dark matter profile is not).

their rather large error bars, and it is clear that in the future such measurements have the potential to provide valuable constraints on models of tidal stripping.

5 EFFECTS OF PHYSICAL PROCESSES

In the previous section, we have explored the effects of varying parameters of the model and their effect on key galaxy properties. We will now instead briefly explore the effects of certain physical processes (those which either are new to this work or have not been extensively examined in the past) on the results of our galaxy formation model. The intent here is not to assess whether these models are ‘better’ than our standard model – they all utilize less realistic physical models – but to examine the effects of ignoring certain physical processes or of making certain assumptions. This emphasizes one of the key strengths of the semi-analytic approach: the ability to rapidly investigate the importance of different physical processes on the properties of galaxies. Rather than showing all model results in each case, we will show a small selection of model results which best demonstrate the effects of the updated model.

5.1 Reionization and photoheating

Our standard model includes a fully self-consistent treatment of the evolution of the IGM and its back reaction on galaxy formation. Two key physical processes are at work here. The first is the suppression of baryonic infall into haloes due to the heating of the IGM by the photoionizing background (see Section 2.10.4). The second is the reduction in cooling rates of gas in haloes as a result of photoheating by the same background (see Section 2.6.7). Here, we compare this standard model to a model with identical parameters, but with these two physical processes switched off. (We retain Compton cooling and molecular hydrogen cooling, but revert to collisional ionization equilibrium cooling curves since there is no photon background in this model.)

Fig. 31 shows some of the key effects of making these changes to our best-fitting model. In panel ‘a’ we show the $z = 0$ b_J -band luminosity function. The model with no baryonic accretion suppression or photoheating (green line) shows a small excess of very bright galaxies relative to the best-fitting model (blue line) due to slightly different cooling rates in this model which affect the efficiency of AGN feedback. As shown in panel ‘b’ of Fig. 31, the $z = 5$ and $z = 6$ UV luminosity functions are almost identical in this variant model and our best-fitting model. At these higher redshifts AGN feedback has yet to become a significant factor in galaxy evolution. A small excess of galaxies is seen in the model with no baryonic accretion suppression or photoheating at the faintest magnitudes plotted. This is as expected – those mechanisms preferentially suppress the formation of very low mass galaxies.

The effects of this change in the AGN feedback can be seen also in panel ‘c’, where we show the star formation history of the Universe. At high redshifts, the two models are nearly identical. However, below $z \approx 1.5$ when AGN feedback begins to come into play, the two models diverge (primarily due to differences in their quiescent star formation rates – the rates of bursting star formation remain quite similar), due to the weakened AGN feedback in this variant model.

Finally, in panel ‘d’, we show the luminosity function of Local Group satellites. There is little difference between this variant model and the best-fitting model for satellites brighter than about $M_V = -10$ – photoheating and baryonic suppression play only a minor role in shaping the properties of these brighter satellites. At fainter

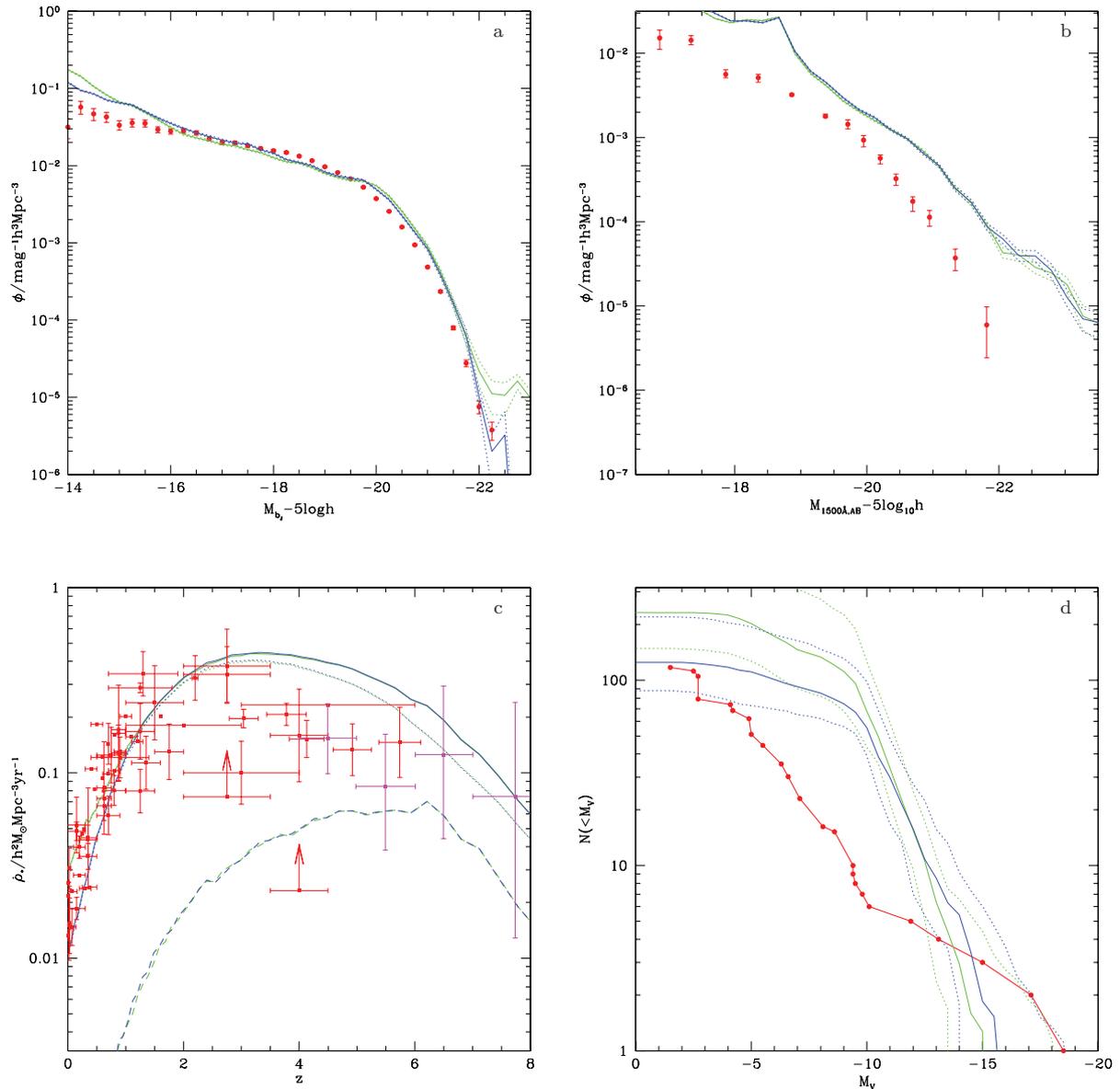


Figure 31. Comparisons between our best-fitting model (blue lines) and the same model without the effects of suppression of baryonic accretion or photoionization equilibrium cooling (green lines). (a) The $z = 0$ b_J -band luminosity function as in Fig. 9. (b) The $z = 5$ 1500 \AA luminosity function as in Fig. 15. (c) The mean star formation rate density in the Universe as a function of redshift as in Fig. 8. (d) The luminosity function of Local Group satellite galaxies as in Fig. 25.

magnitudes, the variant model predicts more satellites than the best-fitting model – by about a factor of 2. Suppression of baryonic accretion and photoheating are clearly then important mechanisms for determining the number of satellites in the Local Group, but other baryonic effects (namely SNe feedback) are clearly at work in reducing the number of satellites below the number of dark matter subhaloes.

5.2 Orbital hierarchy

In our standard model, the full hierarchy of substructures (i.e. haloes within haloes within haloes...) is followed (see Section 2.8). This is in contrast to all previous semi-analytic treatments, in which only the first level of the hierarchy has been considered (i.e. only subhaloes, no sub-subhaloes, etc.). Fig. 32 compares results from this variant model (green lines) with those from our best-fitting standard

model (blue lines). Panel ‘a’ of this figure shows the $z = 0$ b_J -band luminosity function of galaxies. Without a hierarchy of substructures we find that this luminosity function is unchanged over most of the range of luminosities shown. The exception is for the brightest galaxies, which become slightly brighter when no hierarchy of substructures is used. These galaxies grow primarily through merging, and this suggests therefore that including a hierarchy of substructures reduces the rate of merging on to these galaxies. At first sight, this seems counter intuitive as galaxies should have more opportunity to merge as they pass through each level of the hierarchy. In fact, this is not the case. A subhalo may sink within the potential well of a halo and then be tidally stripped, releasing any sub-subhaloes it may contain into the halo. These sub-subhaloes (which become subhaloes in their new host) are placed on to new orbits consistent with their orbital position and velocity at the time at which their subhalo was disrupted. The merging time-scale for

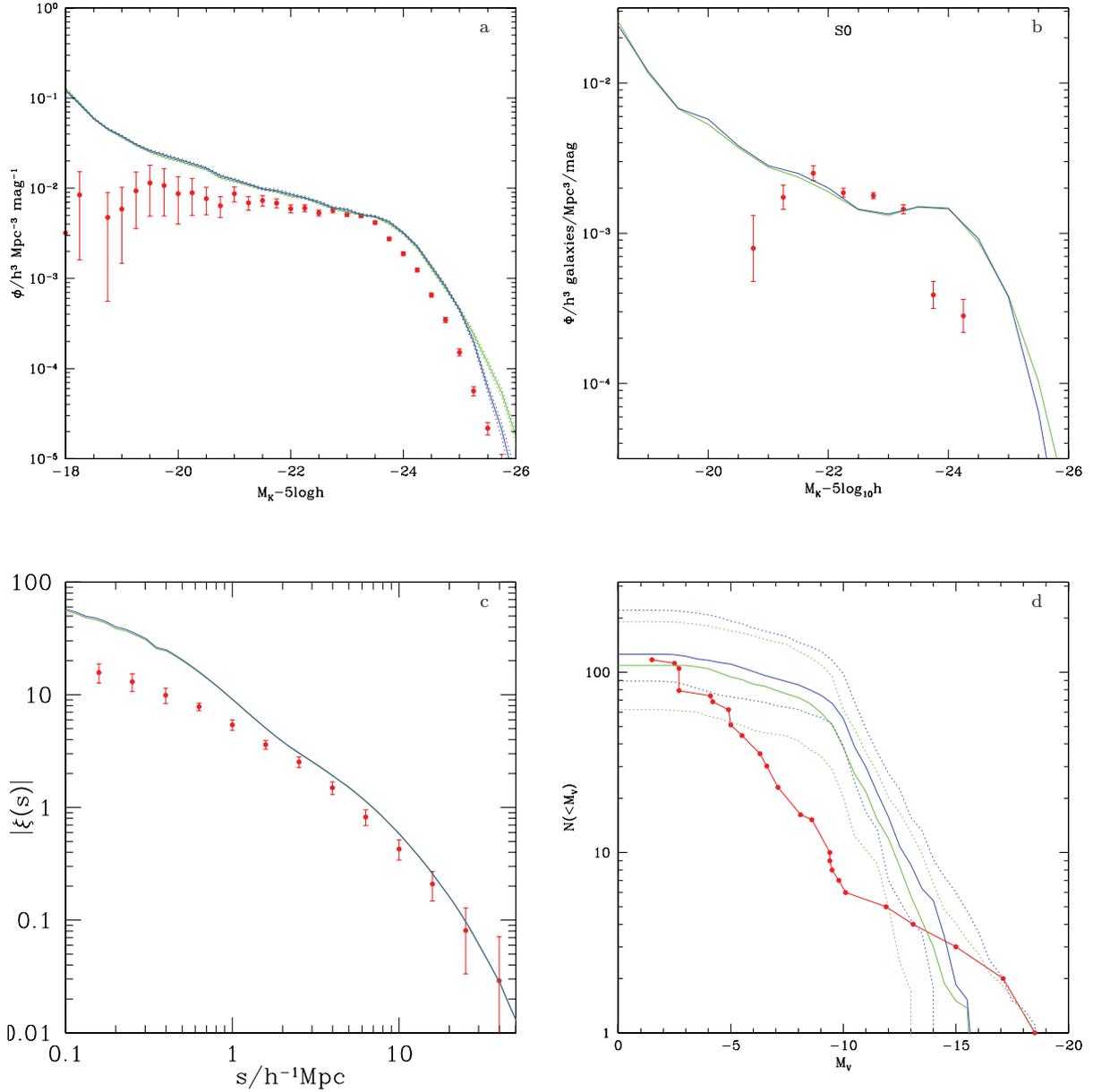


Figure 32. Comparisons between our best-fitting model (blue lines) and the same model without a full hierarchy of substructures (green lines). (a) The $z=0$ b_J -band luminosity function as in Fig. 9. (b) The K -band $z=0$ luminosity function of S0 galaxies as in Fig. 13. (c) The redshift space two-point correlation function of galaxies with $-18.5 < b_J \leq -17.5$ as in Fig. 23. (d) The luminosity function of Local Group satellite galaxies as in Fig. 25.

these orbits plus the time they have already spent orbiting with a subhalo can be longer than the merging time-scale they would have received if they had been made subhaloes as soon as they crossed the virial radius of the host halo. This is due in part to the relatively weak dependence of merging time-scale on $r_C(E)$ in the Jiang et al. (2008) fitting formula²⁸ and partly due to the fact that sub-subhaloes are ejected on to relatively energetic orbits (since they effectively gain a kick in velocity as their subhalo no longer holds them in place).

Panel ‘b’ in Fig. 32 shows that most of the increase in luminosity when the orbital hierarchy is ignored occurs in the S0 morphological

class, which, in this model, makes up a significant part of the bright end of the luminosity function. Panel ‘c’ shows that the inclusion of the orbital hierarchy makes little difference to the correlation function of galaxies. Mergers between galaxies remain dominated by subhalo–halo interactions, such that this new physics has little impact on the number of pairs of galaxies in massive haloes. Finally, panel ‘d’ shows the luminosity function of Local Group galaxies. Their numbers are slightly reduced when the orbital hierarchy is ignored, a direct consequence of the slightly increased merger rate.

5.3 Tidal and ram-pressure stripping

Our standard model incorporates both ram-pressure and tidal stripping of gas and stars from galaxies and their hot gaseous atmospheres. We compare this standard model to one in which both

²⁸We note that this formula has not been well-tested in the regime in which we are employing it. A more detailed study of the merging time-scales and orbits of sub-subhaloes is clearly warranted.

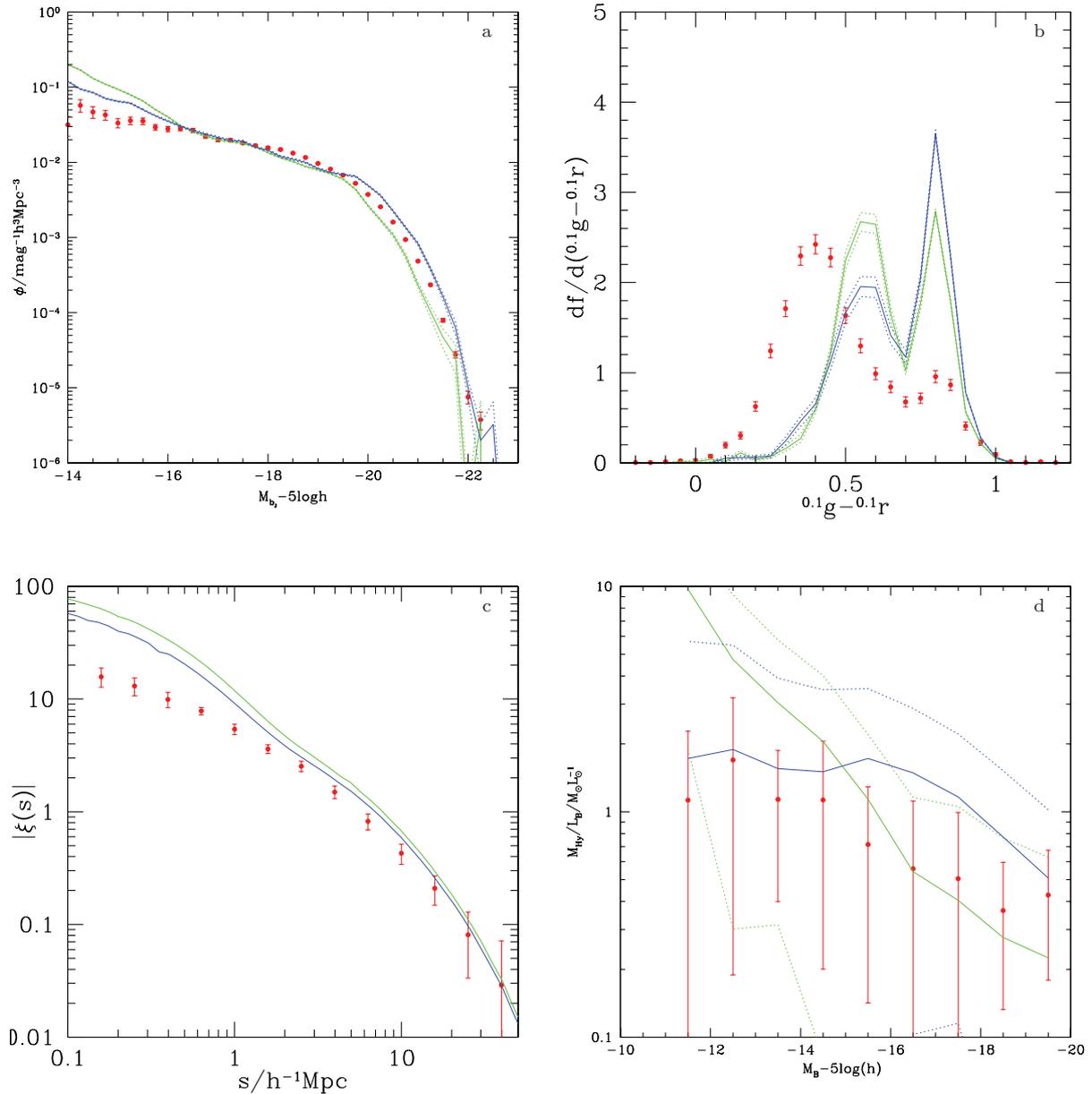


Figure 33. Comparisons between our best-fitting model (blue lines) and the same model without the effects of tidal or ram-pressure stripping of gas and stars from galaxies and their hot atmospheres (green lines). (a) The $z = 0$ b_J -band luminosity function as in Fig. 9. (b) The $^{0.1g-0.1r}$ colour distribution for galaxies at $z = 0.1$ with $-17 < M_{0.1g} \leq -16$ as in Fig. 16. (c) The redshift space two-point correlation function of galaxies with $-18.5 < b_J \leq -17.5$ as in Fig. 23. (d) Gas (hydrogen) to B -band light ratios at $z = 0$ as a function of B -band absolute magnitude as in Fig. 22.

of these stripping mechanisms have been switched off. In general, tidal stripping of stars will reduce the luminosity of satellite galaxies. Ram-pressure or tidal stripping of gas from galaxies or their hot atmospheres will also reduce the luminosity of satellites and, additionally, may increase the luminosity of central galaxies (since the stripped gas is added to their supply of potential fuel).

Fig. 33 compares results from the model with no tidal or ram-pressure stripping (green lines) with our standard, best-fitting model (blue lines). In panel ‘a’ we show the b_J -band luminosity function. At the faintest magnitudes, the model without stripping shows an excess of galaxies relative to the standard model. This is due to low-mass galaxies in groups and clusters being stripped of a significant fraction of their stars in the standard model. Conversely, the model without stripping produces fewer of the brightest galaxies (or, more

correctly, the bright galaxies that it produces are not quite as luminous as in the standard model). This is a consequence of the fact the ram-pressure stripping is able to remove some gas from low-mass galaxies, making it available for later accretion on to massive galaxies, allowing those massive galaxies to grow somewhat more luminous. In panel ‘b’, we examine the colour distribution of faint galaxies. The model with no stripping produces a shift of galaxies to the blue cloud as expected – with stripping included these galaxies lose their gas supply and quickly turn red.

A further effect of stripping can be seen in panel ‘c’ which shows the correlation function of faint galaxies. Without stripping, this is increased on small scales since a greater number of galaxies in massive haloes now make it into the luminosity range selected. Tidal stripping of stars (and, to some extent, ram-pressure removal of gas)

reduces the luminosities of cluster galaxies and thereby reduces the number of galaxy pairs on small scales in a given luminosity range, thereby helping to reduce small-scale correlations. Finally, we show in panel ‘d’ the gas-to-light ratio in a model without stripping. In low-mass galaxies the resulting ratio is much higher than in our standard case, a direct result of this gas no longer being removed by ram-pressure forces. In more massive galaxies there is, instead, a reduction in the gas-to-light ratio relative to the standard model arising because much of the gas is now locked away in smaller systems and so not available for incorporation into larger galaxies.

Although not shown in Fig. 33 stripping processes have an effect on Local Group galaxies – in the absence of stripping there is a modest increase (by around 50 per cent) in the number of galaxies brighter than $M_V = -10$, but the total number of galaxies is mostly unchanged. Additionally, the sizes of Local Group satellites are larger when stripping processes are ignored as expected (many of the satellites lose their outer portions due to tidal stripping), while metallicities are mostly unaffected.

5.4 Non-instantaneous recycling, enrichment and supernovae feedback

Our standard model utilizes a fully non-instantaneous model of recycling and chemical enrichment from stellar populations and of feedback from SNe. We compare this model with one in which the instantaneous recycling approximation is used and in which SNe feedback occurs instantaneously after star formation. In this model, cooling rates are computed from the total metallicity (rather than accounting for the abundances of individual elements as described in Section 2.6) since we cannot track individual elements in this approximation. We adopt a yield of $p = 0.04$ and a recycled fraction of $R = 0.39$ for this instantaneous recycling model. (These values correspond approximately to the values expected for a single stellar population with a Chabrier IMF and an age of approximately 10 Gyr.)

Figs 34 and 35 compare the results of this model with our best-fitting standard model. In Fig. 34, panel ‘a’ shows that at $z = 0$ the bright-end of the b_J -band luminosity function is shifted brightwards in the instantaneous model. This is a consequence of the increased metal enrichment in this model which increases cooling rates (which both increases the amount of gas that can cool and increases the mass scale at which AGN feedback becomes effective). This trend is reversed at higher redshifts for the UV luminosity function that we consider. Here, the luminosity function is shifted fainter in the instantaneous model. This effect is due to increased dust extinction in the instantaneous model (which is able to build up metals more rapidly, particularly at high redshifts and so results in dustier galaxies).

Panel ‘b’ shows the star formation rate density as a function of redshift. The instantaneous model shows a lower star formation rate at high redshift, and a higher rate at low redshift compared to our standard model. At high redshift this can be seen to be due almost entirely to a change in the rate of bursty star formation. The cause of this is rather subtle: in the non-instantaneous model gas is rapidly locked up into stars at high redshifts and is only slowly returned to the ISM of galaxies. This, coupled with somewhat reduced feedback in the non-instantaneous model (since it takes some time for the SNe to occur after star formation happens), makes discs more massive and therefore more prone to instabilities (see Section 2.4.1). The non-instantaneous model has more instability-triggered bursts of star formation at high redshift and there is more gas available to burst in those events. At low redshifts, differences in metal enrichment

in hot gas in the instantaneous model result in slightly less efficient AGN feedback and, therefore, a higher star formation rate.

Instantaneous enrichment has a big effect on galaxy colours as indicated in panels ‘c’ and ‘d’ of Fig. 34. At faint magnitudes, we find a somewhat better fit to the data in the instantaneous model (the blue and red peaks are more widely separated and the red peak is less populated). However, at bright magnitudes the instantaneous model produces too many blue galaxies and too few red ones, resulting in significant disagreement with the data.

Panel ‘a’ of Fig. 35 shows the sizes of galaxy discs. Remarkably, the instantaneous models show a much better match to the data than our standard model.²⁹ This can be traced to a corresponding difference in the distributions of specific angular momenta of discs in the two models, which, in turn, can be traced to the different rates of instability-triggered bursts at high redshifts in the two models. In the non-instantaneous model these happen at a high rate. As a result, the low angular momentum material of these discs is locked up into the spheroid components. Later accretion then results in the formation of discs from higher angular momentum material, resulting in discs that are too large. The stochasticity of this process likewise leads to a large dispersion in disc-specific angular momenta and, therefore, sizes. In the instantaneous model, the rate of instability-triggered bursts is greatly reduced, allowing discs to retain their early accreted, low angular momentum material, giving smaller discs with less variation in size.

Panel ‘b’ shows an example of the distribution of stellar metallicities. Stars in the instantaneous model are enriched to higher metallicities as expected – in the non-instantaneous model it takes time for stars to evolve and produce metals, allowing less enrichment overall. Panels ‘c’ and ‘d’ show the effects on gas content and metallicity, respectively. The gas content is reduced in the instantaneous model and is in excellent agreement with the data. This is a result of the late-time replenishment of the ISM in the non-instantaneous model by material recycled from stars. The instantaneous model produces lower gas phase metallicities, again as a result of the lack of this late-time replenishment which consists of relatively low metallicity material.

5.5 Adiabatic contraction

Adiabatic contraction of dark matter haloes in response to the condensation of baryons is included in our standard model as described in Section 2.7. In Fig. 36 we compare our standard model with one in which this adiabatic contraction is switched off such that dark matter haloes profiles are unchanged by the presence of baryons. Such a change may be expected to result in galaxies which are somewhat larger and more slowly rotating. Panel ‘a’ shows the effects on Local Group satellite galaxy sizes. A slight increase in size is seen as expected. For larger galaxies, we see a similar effect. Rotation speeds of galaxies are less affected though – panel ‘b’ shows a slice through the Tully–Fisher and indicates that switching off adiabatic contraction has actually had little effect on this statistic.

6 DISCUSSION

We have described a substantially revised implementation of the GALFORM semi-analytic model of galaxy formation. This version

²⁹It is worth noting that the Bower et al. (2006) model uses the instantaneous recycling approximation and also does better at matching galaxy sizes than our current best-fitting model.

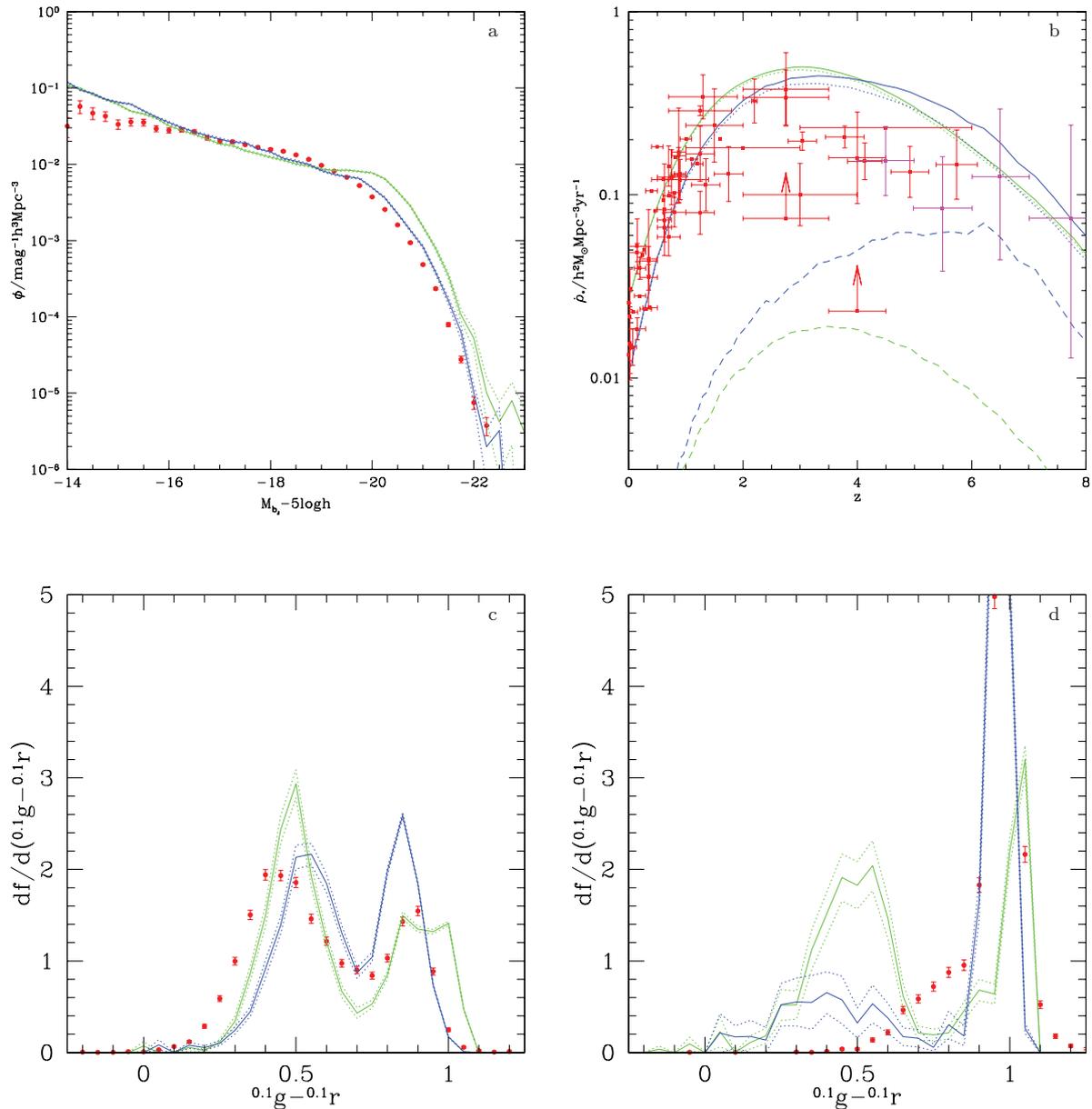


Figure 34. Comparisons between our best-fitting model (blue lines) and the same model using an instantaneous approximation for recycling, chemical enrichment and SNe feedback (green lines). (a) The $z=0$ b_J -band luminosity function as in Fig. 9. (b) The star formation rate density as a function of redshift as in Fig. 8. (c) The $0.1g-0.1r$ colour distribution for galaxies at $z=0.1$ with $-18 < M_{0.1g} \leq -17$ as in Fig. 16. (d) The $0.1g-0.1r$ colour distribution for galaxies at $z=0.1$ with $-22 < M_{0.1g} \leq -21$ as in Fig. 16.

incorporates the numerous developments in our understanding of galaxy formation since the last major review of the code (Cole et al. 2000). Together with changes to the code to implement black hole feedback (Bower et al. 2006, 2008), ram-pressure stripping (Font et al. 2008) and to track the formation of black holes (Malbon et al. 2007), we have made fundamental improvements to key physical processes (such as cooling, re-ionization, galaxy merging and tidal stripping) and removed a number of limiting assumptions (in particular, instantaneous recycling and chemical enrichment are no longer assumed). In addition to computing the properties of galaxies, the model now self-consistently solves for the evolution of the IGM and its influence on later epochs of galaxy formation.

The goals of these changes have been three-fold. First, a prime motivation has been to remove the code's explicit dependence on

discrete halo formation events. In the older code, the mass-doubling events were used to reset halo properties and re-initialize the cooling and free fall accretion calculations. In turn, this leads to abrupt changes in the supply of cold gas to the central galaxy which was often not associated with any particular merging event in the haloes' history. The new method avoids such artificial dependencies and leads to smoothly varying gas accretion rates in haloes with smooth accretion histories, and only leads to abrupt changes during sufficiently important merging events. The new scheme explicitly tracks the energetics of material expelled from galaxies by feedback, and also allows the angular momentum of the feedback and accreted material to be self-consistently propagated through the code. Secondly, we have aimed to enhance the range of physical processes treated in the code so that it incorporates the full range of effects that are

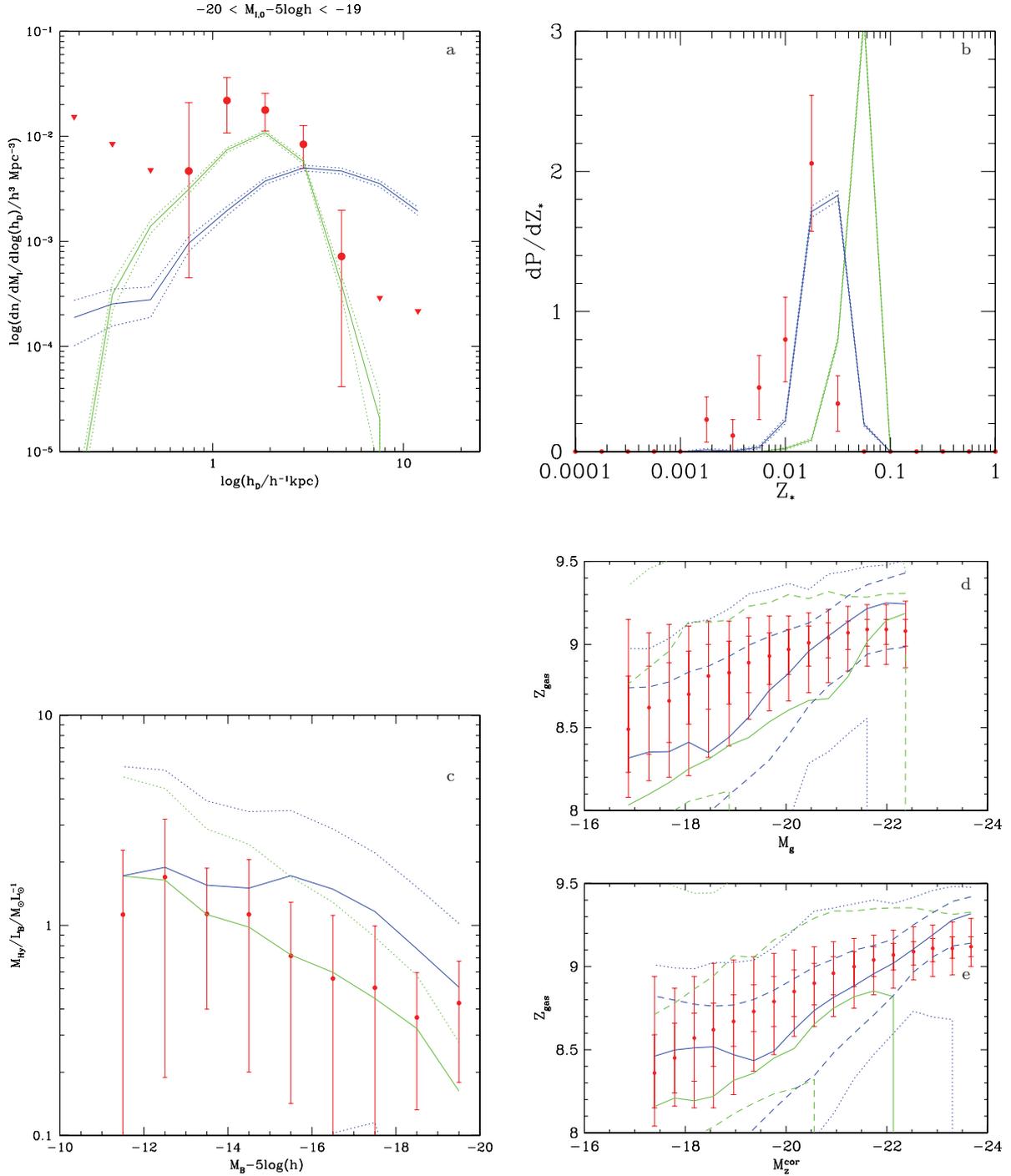


Figure 35. Comparisons between our best-fitting model (blue lines) and the same model with instantaneous recycling, chemical enrichment and SNe feedback (green lines). (a) The distribution of disc sizes for galaxies in the range $-20 < M_{I,0} - 5 \log h \leq -19$ as in Fig. 19. (b) The distribution of stellar metallicities for galaxies in the range $-20 < M_B - 5 \log h \leq -19$ as in Fig. 21. (c) The ratio of hydrogen gas mass to B -band luminosity as in Fig. 22. (d) and (e): the gas phase metallicity as a function of absolute magnitude as in Fig. 20.

likely to be key in determining galaxy properties. In particular, we now include careful treatments of galaxy–environment interactions (tidal and ram-pressure stripping), taking into account the sub-halo hierarchy present within each halo; we take into account the self-consistent re-ionization of the IGM and the impact that this has on gas supply to early galaxies; and we allow for material to be ejected from haloes (both by star formation and by AGN), broadening the range of plausible feedback schemes included in the model. Finally,

the version of the code described may be driven by accurate Monte Carlo realizations of halo merger trees. This allows the uncertainty in the background cosmological parameters to be factored into the model parameter constraints.

We have also advanced the methodology by which we test the model’s performance by simultaneously comparing the model to a wide range of observational data. In addition to our conventional approach of primarily comparing to local optical and near-IR

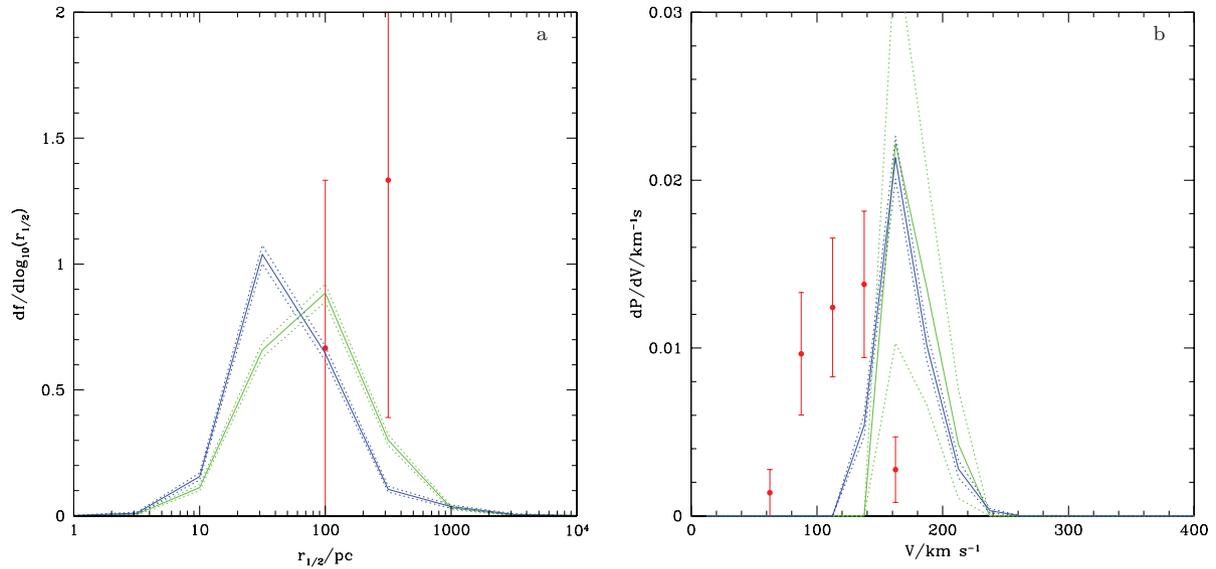


Figure 36. Comparisons between our best-fitting model (blue lines) and the same model without adiabatic contraction of dark matter haloes (green lines). (a) The distribution of half-light radii for Local Group satellites in the magnitude range $-15 < M_V \leq -10$ as in Fig. 26. (b) The Tully–Fisher relation for galaxies in the magnitude range $-21 < M_i \leq -20$ as in Fig. 17.

luminosity functions, we now include luminosity function data covering a much greater a range of redshift and wavelength, the star formation history of the universe, the distribution of galaxies in colour–space, their gas and metal content, the Tully–Fisher relation and various observational measurements of the galaxy size distribution. In addition to these galaxy properties we also use the thermal evolution of the IGM as an additional constraint.

The drawback of introducing additional physical processes is that this introduces additional parameters into the model. However, we now believe that we have the tools to efficiently explore high dimensional parameter spaces and thus identify strongly constrained parameter combinations, and the additional model freedom is much less than the sum of the observational constraints. We performed an extensive search of the new model’s parameter space utilizing the ‘parameter pursuit’ methodology of Bower et al. (in preparation) to rapidly search the high-dimensional space.

This allowed us to find a model which is an adequate description of many of the data sets which were used as constraints. In particular, the model is a good match to local luminosity functions and the overall rate of star formation in the Universe while simultaneously producing reasonable distributions of galaxy colours, metallicities, gas fractions and supermassive black hole masses all while predicting a plausible reionization history. In many of the original data comparisons, the model gives comparable results to Bower et al. (2006). In other comparisons (particularly, colours, metallicities and gas fractions) it greatly improves on the older model.

Additionally, most of the model parameters have shifted relatively little compared to the older model. Where parameters have changed significantly, it is possible to identify a direct cause. For example, the minimum time-scale on which feedback material can be re-accreted by a galaxy (which is set by α_{reheat}) is shorter for the new model. This makes good sense since a fraction of feedback material is now expelled from the system through the new expulsive feedback channel (see Section 2.12). Far from indicating a lack of progress, the comparability of the models is a tremendous success. We cannot emphasize enough how much many of the internal algorithms of the model have been revised: the near stability of the end results

suggests a high degree of convergence, and that adding additional detailing of many aspects of the model is not required.

Despite this encouraging success, significant discrepancies between the model and the data remain in many areas. In particular, the sizes of galaxies are too large in our model (and there is too much dispersion in galaxy sizes). This may reflect a break down in certain model assumptions (e.g. the conservation of angular momentum of gas during the cooling and collapse phase), or that we are still lacking some key physics in this part of the model (e.g. dissipative effects during spheroid formation; Covington et al. 2008). In addition to the sizes, our model continues to produce too many satellite galaxies in high-mass haloes, leading to an overprediction of the small-scale clustering amplitude of faint galaxies; and predicts a Tully–Fisher relation offset from that which is observed, despite using the latest models of adiabatic contraction. [We note that Dutton et al. (2007) have demonstrated the difficulty of obtaining a match to the Tully–Fisher relation quite clearly, and have advocated adiabatic expansion or transfer of angular momentum from gas to dark matter to alleviate this problem.] Additionally, at high redshifts the agreement with luminosity function data is relatively poor, but these results are highly sensitive to the very uncertain effects of dust on galaxy magnitudes.

The overall aim of this work was to construct a model that incorporates the majority of our current understanding of galaxy formation and explore the extent to which such a model can reproduce a large body of observational data spanning a range of physical properties, mass scales and redshifts. This is far from being the final word on the progress of this model. Numerous improvements remain to be made – such as the inclusion of a physics-based model of star formation. Nevertheless, the current version has been demonstrated to produce good agreement with a very wide range of observational data. Despite the large number of adjustable parameters current observational data are more than sufficient to constrain this model – the good agreement with that data should be seen as a confirmation of current galaxy formation theory.

We have not attempted, in this work, to explore in detail which physical processes are responsible for which observed phenomena.

That, and an investigation of which data provided constraints on which parameters, will be the subject of a future work. The parameter space searching methodology described in this paper is quite efficient and successful, but is presently limited by two factors. The first is the available computing time and speed of model calculations which limit how fine-grained any parameter space search can be. Further optimization of our galaxy formation code coupled with more and faster computers will alleviate this problem, but it will remain a limitation for the near future. The second limitation is our ignorance about how best to combine constraints from different data sets. Some of the observational data that we would like to use are undoubtedly affected by poorly understood systematic errors. As a result it is unclear how a precedence should be assigned to each data set. For example, given the robustness of the measurements, are we more interested in the class of models that accurately match the $z = 5$ luminosity, or those that perform better in clustering measurements? Ideally, the model would match both equally well, but underlying systematic errors may make this impossible. Furthermore, to utilize the observational data in a statistically correct way we often require more information (e.g. the full covariance matrix rather than just errors on each data point) than is available.

The most formidable challenge, however, is to better understand the uncertainty in each model prediction. This is a combination of the variance introduced by the limited number of dark matter halo merger trees that we are able to simulate and the accuracy of the approximations made in computing a given property in the model. The first of these is relatively straightforward to estimate (e.g. via a bootstrap resampling approach), but the second is much more difficult. For example, we are quite sure that calculations of dust extinction in rapidly evolving high-redshift galaxies are very uncertain, while calculations of galaxy stellar masses at $z = 0$ are much more robust. The difficulty arises in assigning a numerical ‘weight’ to the model predictions for these different constraints. Beyond simply making an educated guess, one might envisage comparing predictions of dust extinction from our model with a matched sample of simulated high-redshift galaxies in which the complicated dynamics geometry and radiative transfer could be treated more accurately. The variance between the semi-analytic and numerical simulation results would then give a quantitative estimate of the model uncertainty. The problem with such an approach is that creating such a matched sample is extremely difficult and time consuming.

In addition to these uncertainties, we should really include uncertainties arising from non-galaxy formation aspects of the calculation. Good examples of these include the IMF (which we are not explicitly trying to predict in our work, but which is uncertain and makes a significant difference to many of our results) and the spectra of stellar populations which have significant uncertainties in some regimes. Understanding these various model uncertainties is extremely challenging, but is crucial if serious parameter space searching in semi-analytic models is to take place.

However, even in the absence of a well-synthesized approach, it is clear from the data sets we have considered that certain key problems remain to be tackled in order to produce a model of galaxy formation consistent with a broad range of observed data. First, the sizes of model galaxies are too large, suggesting a lack of understanding of the physics of angular momentum in galaxies (see Section 2.7). It is known that the simple energy-conserving model for merger remnant sizes proposed by Cole et al. (2000) systematically overpredicts the sizes of spheroids and results in too much

scatter in their sizes (Covington et al. 2008), but it remains unclear how much this will affect the sizes of discs³⁰ and, furthermore, many spheroids in our model are formed through disc-instabilities rather than mergers – there is, as yet, no good systematic study of how to accurately determine the sizes of such instability-formed spheroids. The disc-instability process itself has significant consequences for the angular momentum content of discs and, as such, a careful examination of this process is called for. Secondly, despite the inclusion of tidal stripping and satellite–satellite merging, the number of satellite galaxies in high-mass haloes seems to remain too high, as evidenced by the clustering of galaxies (see Section 4.8). Thirdly, the clear tension between luminosity function constraints and those from the inferred star formation rate density must be reconciled.

The model described in this work will provide the basis for further improvements to our modelling of galaxy formation. In the near future we intend to return to the following outstanding issues and examine their importance for the constraints and results presented here in greater detail:

- (i) when, exactly, do disc instabilities occur and precisely what effect do they have on the galaxies in which they happen;
- (ii) improved modelling of the sizes of galaxies and how different physical processes affect these sizes;
- (iii) the X-ray properties and hot gas fractions in haloes and how these constrain the amount and type of feedback from galaxies;
- (iv) the effects of patchy reionization on Local Group galaxy properties and on the galaxy population as a whole;
- (v) the importance of the cold mode of gas accretion and how this affects the build up of galaxies at high redshifts (cf. Brooks et al. 2009);
- (vi) improved modelling of AGN feedback utilizing recent estimates of jet power, spin-up rates and the effects of mergers on black hole spin and mass (Boyle, Kesden & Nissanke 2008; Benson & Babul 2009);
- (vii) examination of physically motivated models of star formation and SNe feedback utilizing the framework of Stringer & Benson (2007).

7 CONCLUSIONS

In this paper, we have presented recent developments of the galaxy formation model GALFORM. This extends the model presented in Cole et al. (2000) and Bower et al. (2006) adding many additional physical process (such as environmental interactions and additional feedback channels), improving the treatment of other key processes (including cooling, re-ionization and galaxy merging) and removing unnecessary limiting assumptions (such the instantaneous recycling approximation).

The new code is compared to wide range of observational constraints from both the local and distant universe and across a wide range of wavelengths. We navigate through the high dimensional parameter space using the ‘projection pursuit’ method suggested in Bower et al. (in preparation), identifying a model that performs well in many of the observational comparisons. We find it impossible to identify a model that matches all the available data sets well and there are inherent tensions between the data sets pointing to some remaining inadequacies in our understanding and implementation. In particular, the model as it stands fails to correctly account for the

³⁰Discs feel the gravitational potential of any embedded spheroid, so their sizes will be somewhat reduced if the sizes of spheroids are systematically reduced.

observed distribution of galaxy sizes and the observed Tully–Fisher relation.

Galaxy formation is an inherently complex and highly non-linear process. As such, it is clear that our understanding of it remains incomplete and our ability to model it imperfect. Nevertheless, huge progress has been made in both of these areas, and we expect that progress will continue at a rapid pace. The model described in this work provides an excellent match to many data sets and is in reasonable agreement with many others; it represents a solid foundation upon which to base further calculations of galaxy formation. In particular, with its parameters well constrained by current data it can be used to make predictions for as yet unprobed regimes of galaxy formation.

The present work is clearly not the last word on the subjects covered herein, however. In fact, we expect to constantly revise our model in response to new constraints and improved understanding of the physics.³¹ This simply reflects the current state of galaxy formation theory – it is a rapidly developing field about which we are constantly gaining new insight.

ACKNOWLEDGMENTS

AJB acknowledges support from the Gordon and Betty Moore Foundation and would like to acknowledge the hospitality of the Kavli Institute for Theoretical Physics at the University of California, Santa Barbara, where part of this work was completed. This research was supported in part by the National Science Foundation under Grant No. NSF PHY05-51164. We thank the GALFORM team (Carlton Baugh, Shaun Cole, Carlos Frenk, John Helly and Cedric Lacey) for allowing us to use the collaboratively developed GALFORM code in this work. This work has benefited from conversations with numerous people, including Juna Kollmeier, Aparna Venkatesan, Annika Peter, Alyson Brooks and Yu Lu. We thank Simon White and the anonymous referee for suggestions which helped improve the clarity of the original manuscript. We thank Shiyin Shen for providing data in electronic form. We are grateful to the authors of RECAST and CLOUDY for making these valuable codes publicly available and to Charlie Conroy, Jim Gunn, Martin White and Jason Tumlinson for providing SEDs of single stellar populations. Lauren Porter and Tom Fox contributed code to compute galaxy clustering and IGM evolution, respectively. We gratefully acknowledge the Institute for Computational Cosmology at the University of Durham for supplying a large fraction of the computing time required by this project. This research was supported in part by the National Science Foundation through TeraGrid (Catlett 2007) resources provided by the NCSA and by Amazon Elastic Compute Cloud resources provided by a generous grant from the Amazon in Education program.

REFERENCES

Abadi M. G., Moore B., Bower R. G., 1999, *MNRAS*, 308, 947
 Almeida C., Baugh C. M., Lacey C. G., Frenk C. S., Granato G. L., Silva L., Bressan A., 2010, *MNRAS*, 402, 544
 Angulo R. E., Lacey C. G., Baugh C. M., Frenk C. S., 2009, *MNRAS*, 399, 983

Arrigoni M., Trager S. C., Somerville R. S., Gibson B. K., 2010, *MNRAS*, 402, 173
 Athanassoula E., 2008, *MNRAS*, 390, L69
 Barnes J., Efstathiou G., 1987, *ApJ*, 319, 575
 Barnes E. I., Williams L. L. R., Babul A., Dalcanton J. J., 2007a, *ApJ*, 654, 814
 Barnes E. I., Williams L. L. R., Babul A., Dalcanton J. J., 2007b, *ApJ*, 655, 847
 Baugh C. M., Cole S., Frenk C. S., Lacey C. G., 1998, *ApJ*, 498, 504
 Baugh C. M., Benson A. J., Cole S., Frenk C. S., Lacey C. G., 1999a, *MNRAS*, 305, L21
 Baugh C. M., Lacey C. G., Cole S., Frenk C. S., 1999b, in *The Most Distant Radio Galaxies*. Royal Netherlands Academy of Arts and Science, Amsterdam, p. 265
 Baugh C. M., Lacey C. G., Frenk C. S., Benson A. J., Cole S., Granato G. L., Silva L., Bressan A., 2004, *New Astron. Rev.*, 48, 1239
 Baugh C. M., Lacey C. G., Frenk C. S., Granato G. L., Silva L., Bressan A., Benson A. J., Cole S., 2005, *MNRAS*, 356, 1191
 Becker G. D., Rauch M., Sargent W. L. W., 2007, *ApJ*, 662, 72
 Benson A. J., 2005, *MNRAS*, 358, 551
 Benson A. J., 2008, *MNRAS*, 388, 1361
 Benson A. J., Babul A., 2009, *MNRAS*, 397, 1302
 Benson A. J., Baugh C. M., Cole S., Frenk C. S., Lacey C. G., 2000a, *MNRAS*, 316, 107
 Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000b, *MNRAS*, 311, 793
 Benson A. J., Nusser A., Sugiyama N., Lacey C. G., 2001, *MNRAS*, 320, 153
 Benson A. J., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2002a, *MNRAS*, 333, 177
 Benson A. J., Lacey C. G., Baugh C. M., Cole S., Frenk C. S., 2002b, *MNRAS*, 333, 156
 Benson A. J., Bower R. G., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2003, *ApJ*, 599, 38
 Benson A. J., Lacey C. G., Frenk C. S., Baugh C. M., Cole S., 2004, *MNRAS*, 351, 1215
 Benson A. J., Sugiyama N., Nusser A., Lacey C. G., 2006, *MNRAS*, 369, 1055
 Bett P., Eke V., Frenk C. S., Jenkins A., Helly J., Navarro J., 2007, *MNRAS*, 376, 215
 Blaizot J., Guiderdoni B., Devriendt J. E. G., Bouchet F. R., Hatton S., 2003, *Astrophys. Space Sci.*, 284, 373
 Blaizot J., Guiderdoni B., Devriendt J. E. G., Bouchet F. R., Hatton S. J., Stoehr F., 2004, *MNRAS*, 352, 571
 Blaizot J. et al., 2006, *MNRAS*, 369, 1009
 Blumenthal G. R., Faber S. M., Flores R., Primack J. R., 1986, *ApJ*, 301, 27
 Bovill M. S., Ricotti M., 2009, *ApJ*, 693, 1859
 Bower R. G., Benson A. J., Lacey C. G., Baugh C. M., Cole S., Frenk C. S., 2001, *MNRAS*, 325, 497
 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
 Bower R. G., McCarthy I. G., Benson A. J., 2008, *MNRAS*, 390, 1399
 Boyle L., Kesden M., Nissanke S., 2008, *Phys. Rev. Lett.*, 100, 151101
 Brooks A. M., Governato F., Quinn T., Brook C. B., Wadsley J., 2009, *ApJ*, 694, 396
 Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
 Bullock J. S., Kravtsov A. V., Weinberg D. H., 2000, *ApJ*, 539, 517
 Busha M. T., Alvarez M. A., Wechsler R. H., Abel T., Strigari L. E., 2009, *The Impact of Inhomogeneous Reionization on the Satellite Galaxy Population of the Milky Way*. <http://adsabs.harvard.edu/abs/2009arXiv0901.3553B>
 Cardone V. F., Piedipalumbo E., Tortora C., 2005, *MNRAS*, 358, 1325
 Catlett C., 2007, *Advances in Parallel Computing*. IOS Press, Amsterdam
 Cazaux S., Spaans M., 2004, *ApJ*, 611, 40
 Chabrier G., 2003, *PASP*, 115, 763
 Christodoulou D. M., Shlosman I., Tohline J. E., 1995, *ApJ*, 443, 551
 Cole S., Lacey C., 1996, *MNRAS*, 281, 716

³¹We intend to maintain a ‘living document’ describing any such alterations at www.galform.org, where we will also make available results from the model via an online data base.

- Cole S., Aragon-Salamanca A., Frenk C. S., Navarro J. F., Zepf S. E., 1994, *MNRAS*, 271, 781
- Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, *MNRAS*, 319, 168
- Cole S. et al., 2001, *MNRAS*, 326, 255
- Cole S., Helly J., Frenk C. S., Parkinson H., 2008, *MNRAS*, 383, 546
- Conroy C., Gunn J. E., White M., 2009, *ApJ*, 699, 486
- Cooray A., Sheth R., 2002, *Phys. Rep.*, 372, 1
- Covington M., Dekel A., Cox T. J., Jonsson P., Primack J. R., 2008, *MNRAS*, 384, 94
- Croton D. J. et al., 2006, *MNRAS*, 365, 11
- de Jong R. S., Lacey C., 2000, *ApJ*, 545, 781
- De Lucia G., Kauffmann G., White S. D. M., 2004, *MNRAS*, 349, 1101
- De Lucia G., Springel V., White S. D. M., Croton D., Kauffmann G., 2006, *MNRAS*, 366, 499
- Dehnen W., McLaughlin D. E., Sachania J., 2006, *MNRAS*, 369, 1688
- Devereux N., Hriljac P., Willner S. P., Ashby M. L. N., Willmer C. N. A., 2010, *ApJ*, in press
- Devriendt J. E. G., Guiderdoni B., 2000, *A&A*, 363, 851
- Devriendt J. E. G., Sethi S. K., Guiderdoni B., Nath B. B., 1998, *MNRAS*, 298, 708
- Diaferio A., Kauffmann G., Colberg J. M., White S. D. M., 1999, *MNRAS*, 307, 537
- Dickinson M., 1998, *Proceedings of the Space Telescope Science Institute Symposium*. Cambridge University Press, Baltimore, Maryland, p. 219
- Doroshkevich A. G., 1970, *Astrofizika*, 6, 581
- Dove J. B., Shull J. M., 1994, *ApJ*, 430, 222
- Draine B. T., Lee H. M., 1984, *ApJ*, 285, 89
- Dunkley J. et al., 2009, *ApJS*, 180, 306
- Dutton A. A., van den Bosch F. C., Dekel A., Courteau S., 2007, *ApJ*, 654, 27
- Dutton A. A., van den Bosch F. C., Courteau S., 2008, in *Formation and Evolution of Galaxy Discs*, Vol. 396, *Astronomical Society of the Pacific, Centro Convegno Matteo Ricci*, Rome, Italy, p. 467
- Efstathiou G., Lake G., Negroponte J., 1982, *MNRAS*, 199, 1069
- Efstathiou G., Frenk C. S., White S. D. M., Davis M., 1988, *MNRAS*, 235, 715
- Einasto J., 1965, *Trudy Inst. Astrofiz. Alma-Ata*, 51, 87
- Eisenstein D. J., Hu W., 1999, *ApJ*, 511, 5
- Eke V. R., Cole S., Frenk C. S., 1996, *MNRAS*, 282, 263
- Elmegreen B. G., Elmegreen D. M., Fernandez M. X., Lomonias J. J., 2009, *ApJ*, 692, 12
- Fan X. et al., 2006, *AJ*, 132, 117
- Ferland G. J., Korista K. T., Verner D. A., Ferguson J. W., Kingdon J. B., Verner E. M., 1998, *PASP*, 110, 761
- Ferrara A., Bianchi S., Cimatti A., Giovanardi C., 1999, *ApJS*, 123, 437
- Fontanot F., Monaco P., Silva L., Grazian A., 2007, *MNRAS*, 382, 903
- Fontanot F., Lucia G. D., Monaco P., Somerville R. S., Santini P., 2009a, *MNRAS*, 397, 1776
- Fontanot F., Somerville R. S., Silva L., Monaco P., Skibba R., 2009b, *MNRAS*, 392, 553
- Font A. S. et al., 2008, *MNRAS*, 389, 1619
- Galli D., Palla F., 1998, *A&A*, 335, 403
- Gao L., Navarro J. F., Cole S., Frenk C. S., White S. D. M., Springel V., Jenkins A., Neto A. F., 2008, *MNRAS*, 387, 536
- Gnedin N. Y., Kravtsov A. V., 2006, *ApJ*, 645, 1054
- Gnedin O. Y., Kravtsov A. V., Klypin A. A., Nagai D., 2004, *ApJ*, 616, 16
- González J. E., Lacey C. G., Baugh C. M., Frenk C. S., Benson A. J., 2009, *MNRAS*, 397, 1254
- González-Perez V., Baugh C. M., Lacey C. G., Almeida C., 2009, *MNRAS*, 398, 497
- Governato F., Baugh C. M., Frenk C. S., Cole S., Lacey C. G., Quinn T., Stadel J., 1998, *Nat*, 392, 359
- Granato G. L., Lacey C. G., Silva L., Bressan A., Baugh C. M., Cole S., Frenk C. S., 2000, *ApJ*, 542, 710
- Guiderdoni B., Hivon E., Bouchet F. R., Maffei B., 1998, *MNRAS*, 295, 877
- Gunn J. E., Peterson B. A., 1965, *ApJ*, 142, 1633
- Gustafsson M., Fairbairn M., Sommer-Larsen J., 2006, *Phys. Rev. D*, 74, 123522
- Haardt F., Madau P., 1996, *ApJ*, 461, 20
- Häring N., Rix H., 2004, *ApJ*, 604, L89
- Hatton S., Devriendt J. E. G., Ninin S., Bouchet F. R., Guiderdoni B., Vibert D., 2003, *MNRAS*, 343, 75
- Heger A., Woosley S. E., 2002, *ApJ*, 567, 532
- Henriques B. M. B., Thomas P. A., Oliver S., Roseboom I., 2009, *MNRAS*, 396, 535
- Hollenbach D., McKee C. F., 1979, *ApJS*, 41, 555
- Hopkins A. M., 2004, *ApJ*, 615, 209
- Hopkins A. M., Beacom J. F., 2006, *ApJ*, 651, 142
- Hoyle F., 1949, *Problems in Cosmical Aerodynamics*. *Proceedings of the Symposium on the Motion of Gaseous Masses of Cosmical Dimensions*. Central Air Documents Office, Ohio
- Huchtmeier W. K., Richter O., 1988, *A&A*, 203, 237
- Jiang C. Y., Jing Y. P., Faltenbacher A., Lin W. P., Li C., 2008, *ApJ*, 675, 1095
- Kauffmann G., 1996, *MNRAS*, 281, 487
- Kauffmann G., Charlot S., 1998, *MNRAS*, 294, 705
- Kauffmann G., Haehnelt M., 2000, *MNRAS*, 311, 576
- Kauffmann G., White S. D. M., Guiderdoni B., 1993, *MNRAS*, 264, 201
- Kauffmann G., Guiderdoni B., White S. D. M., 1994, *MNRAS*, 267, 981
- Kauffmann G., Colberg J. M., Diaferio A., White S. D. M., 1999a, *MNRAS*, 303, 188
- Kauffmann G., Colberg J. M., Diaferio A., White S. D. M., 1999b, *MNRAS*, 307, 529
- Kim H. S., Baugh C. M., Cole S., Frenk C. S., Benson A. J., 2009, *MNRAS*, in press
- Kirby E. N., Simon J. D., Geha M., Guhathakurta P., Frebel A., 2008, *ApJ*, 685, L43
- Kistler M. D., Yuksel H., Beacom J. F., Hopkins A. M., Wyithe J. S. B., 2009, *ApJ*, 705, 104
- Koposov S. et al., 2008, *ApJ*, 686, 279
- Kuhlen M., Diemand J., Madau P., Zemp M., 2008, *J. Phys. Conf. Ser.*, 125, 2008
- Lacey C., Cole S., 1993, *MNRAS*, 262, 627
- Lacey C. G., Baugh C. M., Frenk C. S., Silva L., Granato G. L., Bressan A., 2008, *MNRAS*, 385, 1155
- Lacey C. G., Baugh C. M., Frenk C. S., Benson A. J., Orsi A., Silva L., Granato G. L., Bressan A., 2010, *MNRAS*, in press
- Lanzoni B., Guiderdoni B., Mamon G. A., Devriendt J., Hatton S., 2005, *MNRAS*, 361, 369
- Lehnert M. D., Nesvadba N. P. H., Tiran L. L., Matteo P. D., van Driel W., Douglas L. S., Chemin L., Bournaud F., 2009, *ApJ*, 699, 1660
- Lemson G., Kauffmann G., 1999, *MNRAS*, 302, 111
- Macciò A. V., Kang X., Fontanot F., Somerville R. S., Koposov S., Monaco P., 2010, *MNRAS*, 402, 1995
- Madau P., Diemand J., Kuhlen M., 2008a, *ApJ*, 679, 1260
- Madau P., Kuhlen M., Diemand J., Moore B., Zemp M., Potter D., Stadel J., 2008b, *ApJ*, 689, L41
- Malbon R. K., Baugh C. M., Frenk C. S., Lacey C. G., 2007, *MNRAS*, 382, 1394
- Maller A. H., Prochaska J. X., Somerville R. S., Primack J. R., 2001, *MNRAS*, 326, 1475
- Maller A. H., Dekel A., Somerville R., 2002, *MNRAS*, 329, 423
- Maller A. H., Prochaska J. X., Somerville R. S., Primack J. R., 2003, *MNRAS*, 343, 268
- Marchesini D., van Dokkum P. G., 2007, *ApJ*, 663, L89
- Marigo P., 2001, *A&A*, 370, 194
- Martínez-Serrano F. J., Serna A., Domínguez-Tenreiro R., Mollá M., 2008, *MNRAS*, 388, 39
- Mateo M. L., 1998, *ARA&A*, 36, 435
- McCarthy I. G., Frenk C. S., Font A. S., Lacey C. G., Bower R. G., Mitchell N. L., Balogh M. L., Theuns T., 2008, *MNRAS*, 383, 593
- McGee S. L., Balogh M. L., 2010, *MNRAS*, 403, 79
- McKay M. D., Beckman R. J., Conover W. J., 1979, *Technometrics*, 21, 239

- McLure R. J., Cirasuolo M., Dunlop J. S., Foucaud S., Almaini O., 2009, *MNRAS*, 395, 2196
- Meiksin A., 2006, *MNRAS*, 365, 807
- Merritt D., Navarro J. F., Ludlow A., Jenkins A., 2005, *ApJ*, 624, L85
- Monaco P., Fontanot F., Taffoni G., 2007, *MNRAS*, 375, 1189
- Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, *ApJ*, 524, L19
- Muñoz J. M., Madau P., Loeb A., Diemand J., 2009, *MNRAS*, 400, 1593
- Nagashima M., Lacey C. G., Baugh C. M., Frenk C. S., Cole S., 2005a, *MNRAS*, 358, 1247
- Nagashima M., Lacey C. G., Okamoto T., Baugh C. M., Frenk C. S., Cole S., 2005b, *MNRAS*, 363, L31
- Navarro J. F., Frenk C. S., White S. D. M., 1997, *ApJ*, 490, 493
- Navarro J. F. et al., 2004, *MNRAS*, 349, 1039
- Norberg P. et al., 2002, *MNRAS*, 332, 827
- Okamoto T., Gao L., Theuns T., 2008, *MNRAS*, 390, 920
- Parkinson H., Cole S., Helly J., 2008, *MNRAS*, 383, 557
- Parry O. H., Eke V. R., Frenk C. S., 2009, *MNRAS*, 396, 1972
- Peebles P. J. E., 1968, *ApJ*, 153, 1
- Peebles P. J. E., 1969, *ApJ*, 155, 393
- Pizagno J. et al., 2007, *AJ*, 134, 945
- Portinari L., Chiosi C., Bressan A., 1998, *A&A*, 334, 505
- Pozzetti L. et al., 2003, *A&A*, 402, 837
- Prada F., Klypin A. A., Simonneau E., Betancort-Rijo J., Patiri S., Gottlber S., Sanchez-Conde M. A., 2006, *ApJ*, 645, 1001
- Press W. H., Schechter P., 1974, *ApJ*, 187, 425
- Reed D. S., Bower R., Frenk C. S., Jenkins A., Theuns T., 2007, *MNRAS*, 374, 2
- Rees M. J., Ostriker J. P., 1977, *MNRAS*, 179, 541
- Sage L. J., 1993, *A&A*, 272, 123
- Saunders W., Rowan-Robinson M., Lawrence A., Efstathiou G., Kaiser N., Ellis R. S., Frenk C. S., 1990, *MNRAS*, 242, 318
- Sharma S., Steinmetz M., 2005, *ApJ*, 628, 21
- Shen S., Mo H. J., White S. D. M., Blanton M. R., Kauffmann G., Voges W., Brinkmann J., Csabai I., 2003, *MNRAS*, 343, 978
- Shull J. M., van Steenberg M. E., 1985, *ApJ*, 298, 268
- Silva L., Granato G. L., Bressan A., Danese L., 1998, *ApJ*, 509, 103
- Sobol' I. M., 1967, *U.S. Comput. Math. Math. Phys.*, 7, 86
- Somerville R. S., 2002, *ApJ*, 572, L23
- Somerville R. S., Livio M., 2003, *ApJ*, 593, 611
- Somerville R. S., Primack J. R., 1999, *MNRAS*, 310, 1087
- Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008b, *MNRAS*, 391, 481
- Somerville R. S. et al., 2008a, *ApJ*, 672, 776
- Songaila A., 2004, *AJ*, 127, 2598
- Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, *MNRAS*, 328, 726
- Springel V. et al., 2008, *MNRAS*, 391, 1685
- Stadel J., Potter D., Moore B., Diemand J., Madau P., Zemp M., Kuhlen M., Quilis V., 2009, *MNRAS*, 398, L21
- Stark D. P., Swinbank A. M., Ellis R. S., Dye S., Smail I. R., Richard J., 2008, *Nat*, 455, 775
- Steidel C. C., Adelberger K. L., Giavalisco M., Dickinson M., Pettini M., 1999, *ApJ*, 519, 1
- Stringer M. J., Benson A. J., 2007, *MNRAS*, 382, 641
- Sun M., Voit G. M., Donahue M., Jones C., Forman W., Vikhlinin A., 2009, *ApJ*, 693, 1142
- Sutherland R. S., Dopita M. A., 1993, *ApJS*, 88, 253
- Swinbank M. et al., 2009, *MNRAS*, 400, 1121
- Taylor J. E., Babul A., 2004, *MNRAS*, 348, 811
- Taylor J. E., Babul A., 2005, *MNRAS*, 364, 515
- Taylor J. E., Navarro J. F., 2001, *ApJ*, 563, 483
- Tegmark M., Silk J., Rees M. J., Blanchard A., Abel T., Palla F., 1997, *ApJ*, 474, 1
- Tollerud E. J., Bullock J. S., Strigari L. E., Willman B., 2008, *ApJ*, 688, 277
- Tormen G., 1997, *MNRAS*, 290, 411
- Tremonti C. A. et al., 2004, *ApJ*, 613, 898
- Tumlinson J., 2006, *ApJ*, 641, 1
- Tumlinson J., Shull J. M., Venkatesan A., 2003, *ApJ*, 584, 608
- Venkatesan A., Giroux M. L., Shull J. M., 2001, *ApJ*, 563, 1
- Verner D. A., Yakovlev D. G., 1995, *A&AS*, 109, 125
- Vikhlinin A. et al., 2009, *ApJ*, 692, 1033
- Vitvitska M., Klypin A. A., Kravtsov A. V., Wechsler R. H., Primack J. R., Bullock J. S., 2002, *ApJ*, 581, 799
- Warren M. S., Quinn P. J., Salmon J. K., Zurek W. H., 1992, *ApJ*, 399, 405
- Wechsler R. H., Somerville R. S., Bullock J. S., Kolatt T. S., Primack J. R., Blumenthal G. R., Dekel A., 2001, *ApJ*, 554, 85
- Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., 2006, *MNRAS*, 366, 2
- White S. D. M., 1984, *ApJ*, 286, 38
- White S. D. M., Frenk C. S., 1991, *ApJ*, 379, 52
- White S. D. M., Rees M. J., 1978, *MNRAS*, 183, 341
- Xue X. X. et al., 2008, *ApJ*, 684, 1143
- Zaritsky D., Kennicutt R. C., Huchra J. P., 1994, *ApJ*, 420, 87
- Zibetti S., White S. D. M., Schneider D. P., Brinkmann J., 2005, *MNRAS*, 358, 949

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