Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

Mon. Not. R. Astron. Soc. **421,** 2904–2916 (2012) doi:10.1111/j.1365-2966.2012.20521.x

Confronting theoretical models with the observed evolution of the galaxy population out to $z = 4$

Bruno M. B. Henriques,[1⋆] Simon D. M. White,[1] Gerard Lemson,[1] Peter A. Thomas,[2] Qi Guo,[3,4] Gabriel-Dominique Marleau[1,5] and Roderik A. Overzier[1]

[1]*Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85741 Garching b. München, Germany*
[2]*Astronomy Centre, University of Sussex, Falmer, Brighton BN1 9QH*
[3]*Partner Group of the Max-Planck-Institut für Astrophysik, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China*
[4]*Department of Physics, Institute for Computational Cosmology, University of Durham, South Road, Durham DH1 3LE*
[5]*Department of Physics, McGill University, 3600 Rue University, Montréal, QC H3A 2T8, Canada*

Accepted 2012 January 7. Received 2012 January 6; in original form 2011 September 16

ABSTRACT

We construct light cones for the semi-analytic galaxy formation simulation of Guo et al. and make mock catalogues for comparison with deep high-redshift surveys. Photometric properties are calculated with two different stellar population synthesis codes in order to study sensitivity to this aspect of the modelling. The catalogues are publicly available and include photometry for a large number of observed bands from 4000 Å to 6 μm, as well as rest-frame photometry and other intrinsic properties of the galaxies (e.g. positions, peculiar velocities, stellar masses, sizes, morphologies, gas fractions, star formation rates, metallicities, halo properties). Guo et al. tuned their model to fit the low-redshift galaxy population but noted that at $z \geq 1$ it overpredicts the abundance of galaxies below the 'knee' of the stellar mass function. Here we extend the comparison to deep galaxy counts in the B, i, J, K and IRAC 3.6, 4.5 and 5.8 μm bands, to the redshift distributions of K and 5.8 μm selected galaxies, the evolution of rest-frame luminosity functions in the B and K bands and the evolution of rest-frame optical versus near-infrared colours. The B, i and J counts are well reproduced, but at longer wavelengths the overabundant high-redshift galaxies produce excess faint counts. At bright magnitudes, counts in the IRAC bands are underpredicted, reflecting overly low stellar metallicities and the neglect of polycyclic aromatic hydrocarbon emission. The predicted redshift distributions for K and 5.8 μm selected samples highlight the effect of emission from thermally pulsing asymptotic giant branch (AGB) stars. The full treatment of the Maraston model predicts three times as many $z \sim 2$ galaxies in faint 5.8 μm selected samples as the model of Bruzual & Charlot, whereas the two models give similar predictions for K-band selected samples. Although luminosity functions are adequately reproduced out to $z \sim 3$ in rest-frame B, the same is true at rest-frame K only if thermally pulsating AGB emission is included, and then only at high luminosity. Fainter than $L_\star$, the two synthesis models agree but overpredict the number of galaxies, another reflection of the overabundance of $\sim 10^{10}\,\mathrm{M}_\odot$ model galaxies at $z \geq 1$. The model predicts that red, passive galaxies should already be in place at $z = 2$ as required by observations.

Key words: methods: numerical – methods: statistical – stars: AGB and post-AGB – galaxies: evolution – galaxies: formation – galaxies: high-redshift.

1 INTRODUCTION

Semi-analytic models of galaxy formation aim to predict the evolution of population properties such as the distributions of stellar mass, luminosity, star formation rate, size, rotation velocity, morphology, gas content and metallicity, as well as the scaling relations linking these properties. They follow astrophysical processes affecting the baryonic components using a series of analytic, physically based models which are embedded either in an analytic representation (White & Frenk 1991; Kauffmann, White & Guiderdoni 1993; Cole et al. 1994; Somerville & Primack 1999) or in a direct

⋆E-mail: bhenriques@mpa-garching.mpg.de

numerical simulation (Kauffmann et al. 1999; Springel et al. 2001, 2005) of the evolution of the underlying dark matter distribution. Uncertain efficiencies and scalings of these astrophysical processes are represented by adjustable parameters. These may be set a priori through a detailed calculation or simulation of specific processes, or they may be determined observationally by matching suitably chosen data (e.g. Bower et al. 2006; Cattaneo et al. 2006; Croton et al. 2006; Menci et al. 2006; De Lucia & Blaizot 2007; Monaco, Fontanot & Taffoni 2007; Somerville et al. 2008; Guo et al. 2011). The extremely broad range of relevant data and the considerable freedom in specifying appropriate recipes complicate the systematic comparison of semi-analytic models with data. New and robust statistical tools have recently been developed to facilitate quantitative comparisons (Kampakoglou, Trotta & Silk 2008; Henriques et al. 2009; Bower et al. 2010; Henriques & Thomas 2010; Lu et al. 2011).

Such comparisons are sensitive to stellar population synthesis models which are required both to derive intrinsic galaxy properties, such as mass, age and star formation rate from observational data, and to calculate luminosities, colours and spectra for model galaxies. Erroneous conversions between physical and observable properties lead to incorrect conclusions about galaxy formation physics, so it is important to check the implications of adopting differing stellar population synthesis models (e.g. Buzzoni 1989; Worthey 1994; Vazdekis et al. 1996; Fioc & Rocca-Volmerange 1997; Leitherer et al. 1999; Bruzual & Charlot 2003; Thomas, Maraston & Bender 2003; Maraston 2005; Conroy, Gunn & White 2009).

For example, the impact of including models for thermally pulsating asymptotic giant branch (TP-AGB) stars has been studied in some detail in recent years. The contribution from these stars can significantly enhance the near-infrared (near-IR) emission of galaxies with Gyr-old populations (e.g. Maraston 1998, 2005; van der Wel et al. 2006, Marigo & Girardi 2007; Conroy et al. 2009; Charlot & Bruzual, in preparation). The data of Conroy et al. (2009), Marchesini et al. (2009, 2010), Zibetti, Charlot & Rix (2009) and Santini et al. (2012) show that inclusion of this additional emission can reduce the masses inferred from K-band light by as much as 0.6 dex. The semi-analytic models of Tonini et al. (2009, 2010), Fontanot & Monaco (2010) and Henriques et al. (2011) suggest that a substantial contribution from TP-AGB stars, as predicted by the model of Maraston (2005), may explain the large number of extremely red objects found at $z \sim 2$. Other examples of how uncertainties in stellar evolution modelling affect physical inferences from data are given by Conroy, White & Gunn (2010). Here we compare predictions of the Guo et al. (2011) semi-analytic model for two different population synthesis models, the one originally used by these authors (from Bruzual & Charlot 2003) and that of Maraston (2005).

Even neglecting uncertainties from stellar population modelling, mass-to-light ratios and other physical properties are often poorly constrained by available data. Estimates rely on fitting theoretical models to observed photometry and spectral energy distributions (SEDs) and approximately equivalent fits can often be obtained for broad ranges of assumed star formation history, chemical enrichment history and obscuration by dust. Additional uncertainties arise from possible variations in the initial mass function (IMF) with which stars form, and from possible spectral contributions from an active galactic nucleus (AGN). It has been argued in the past that a single optical colour (e.g. $g - i$) is, in practice, sufficient to derive light-to-mass ratios accurate to 0.1 dex for most galaxies (Bell et al. 2003; Gallazzi & Bell 2009). However, Zibetti et al. (2009) showed that this is only true for relatively weak obscuration. For heavily obscured young populations, resolved photometry is needed to achieve an accuracy better than 0.2 dex and even that requires an additional near-IR colour (e.g. one may combine $g - i$ and $i - H$).

Galaxy formation models directly predict star formation and enrichment histories, so in the absence of obscuration a well-defined SED can be predicted for each galaxy as a superposition of simple stellar populations (SSPs), each made up of coeval stars of a single metallicity. The 'observed' photometry is then easily obtained by redshifting the SED and integrating over the appropriate photometric filter functions. In practice, however, the conversion to observables is heavily influenced by dust and is sensitive to the details of its distribution within a galaxy (e.g. Cole et al. 2000; Granato et al. 2000). This significantly limits the precision with which observables can be predicted from galaxy formation models. Current semi-analytic models often attempt to handle these uncertainties by using observational data to constrain the dust model (e.g. Cole et al. 2000; Granato et al. 2000; Kitzbichler & White 2007; Guo & White 2009).

Because of such difficulties, it seems wise to compare theory and observation for a broad range of properties, at different redshifts and at different 'conversion levels'. The latter is particularly crucial at high redshift, where very limited data are available and the relations between mass, light and star formation rate are very uncertain. There is no preferred 'comparison frame' and conclusions are convincing only if a consistent picture emerges which matches smoothly on to the lower redshift galaxy populations. Guo et al. (2011) compare their model extensively to low-redshift galaxies but only present limited predictions at high redshift. Specifically, they compare to published estimates of the evolution of the stellar mass function of galaxies out to $z \sim 4$, finding significant discrepancies for stellar masses below $5 \times 10^{10}\,\mathrm{M}_{\odot}$.

In this paper, we extend this comparison considerably, analysing the photometric properties of galaxies from high redshift to the present day, and comparing with observations at a variety of levels from number counts as a function of apparent magnitude, through redshift distributions of magnitude-limited samples, to rest-frame luminosity functions as a function of redshift. In particular, we study predictions for the near-IR bands for which data have recently become available from the *Spitzer* satellite. To facilitate this work, we build light cones using the Mock Map Facility (MOMAF) software (Blaizot et al. 2005). We are then able to reproduce the observational selection criteria of modern surveys such as the UKIRT Infrared Deep Sky Survey (UKIDSS; Lawrence et al. 2007), the VIMOS-VLT Deep Survey (VVDS; Le Fèvre et al. 2005), the Deep Evolutionary Exploratory Probe 2 Galaxy Redshift Survey (DEEP2; Davis et al. 2003) and the Cosmic Evolution Survey (COSMOS; Scoville et al. 2007).

Similar studies were performed by Kitzbichler & White (2007), Guo & White (2009) and de la Torre et al. (2011) for earlier versions of the Munich semi-analytic model. The results here are based on the model of Guo et al. (2011), which is implemented simultaneously on the Millennium Simulation (MS) and Millennium-II Simulation (MS-II) and was retuned to fit a broad range of 'high-precision' data on the low-redshift galaxy population, primarily from the Sloan Digital Sky Survey (SDSS). We also expand the photometric coverage from the ultraviolet (UV) to IRAC bands and test the dependence on stellar population synthesis modelling over this wavelength range. In a recent paper, Somerville et al. (2011) compared a different semi-analytic model with photometry extending to even longer wavelengths (the far-IR). This required modelling the re-emission of starlight by heated dust, as also considered by Granato et al. (2000), Cole et al. (2000), Baugh et al. (2005) and

Lacey et al. (2010). Here we avoid this complication and compare only to directly observed starlight.

This paper is organized as follows. In Section 2, we summarize the characteristics of the semi-analytic model we use and we describe how we construct light cones for it. Section 3 then presents results for number counts and redshift distributions as a function of apparent magnitude, and for rest-frame *B*- and *K*-band luminosity functions as a function of redshift. In Section 4, we present our conclusions.

2 THE SEMI-ANALYTIC MODEL

Modern semi-analytic models are built on merger trees from high-resolution dark matter simulations. These provide a description of the evolution of the mass and number density of dark matter haloes and the subhaloes within them, as well as of their spatial and kinematic distributions. The evolution of the baryonic components hosted by these (sub)haloes is then followed using a set of simplified formulae describing each of the relevant astrophysical processes. The latest version of the Munich model (Guo et al. 2011) is implemented on two very large dark matter simulations, the MS (Springel et al. 2005) and the MS-II (Boylan-Kolchin et al. 2009). The MS follows the evolution of structure within a cube of side $500\,h^{-1}$ Mpc (comoving) and its merger trees are complete for subhaloes above a mass resolution limit of $1.7 \times 10^{10}\,h^{-1}\mathrm{M}_{\odot}$. The MS-II follows a cube of side $100\,h^{-1}$ Mpc but with 125 times better mass resolution (subhalo masses greater than $1.4 \times 10^{8}\,h^{-1}\,\mathrm{M}_{\odot}$). Both adopt the same first-year *WMAP* based cosmology (Spergel et al. 2003) with parameters $h = 0.73$, $\Omega_{\mathrm{m}} = 0.25$, $\Omega_{\Lambda} = 0.75$, $n = 1$ and $\sigma_8 = 0.9$. These are outside the region preferred by more recent analyses (in particular, σ_8 appears too high) but this is of no consequence for the issues we study in this paper. For consistency, we will use this cosmology whenever it is necessary to derive the physical properties of galaxies from observed fluxes and redshifts. The distributions of physical properties converge in the two simulations for galaxies with $10^{9.5} < M_{\star} < 10^{11.5}\,\mathrm{M}_{\odot}$. In this study, we focus only on results from the MS, since its resolution limit is well below the stellar masses covered by the data sets with which we compare.

2.1 The model of Guo et al. (2011)

For a full description of the semi-analytic model used in this work, we refer the reader to Guo et al. (2011). Here we briefly describe changes from earlier versions of the Munich semi-analytic model that significantly affect our results.

Following Kitzbichler & White (2007) and Guo & White (2009), the model of Guo et al. (2011) includes a redshift-dependent model for internal extinction which assumes that the dust-to-gas ratio increases with metallicity but decreases with redshift. The effective optical depth is given by

$$\tau_{\lambda} = \left(\frac{A_{\lambda}}{A_{\mathrm{v}}}\right)_{\mathrm{Z}_{\odot}} (1+z)^{-0.4} \left(\frac{Z_{\mathrm{gas}}}{\mathrm{Z}_{\odot}}\right)^{s} \left(\frac{\langle N_{\mathrm{H}}\rangle}{2.1 \times 10^{21}\ \mathrm{atoms\,cm^{-2}}}\right), \quad (1)$$

where $\langle N_{\mathrm{H}}\rangle$ represents the mean column density of hydrogen, $(A_{\lambda}/A_{\mathrm{v}})_{\mathrm{Z}_{\odot}}$ is the extinction curve for the solar metallicity taken from Mathis, Mezger & Panagia (1983) and $s = 1.35$ for $\lambda < 2000$ Å and $s = 1.6$ for $\lambda > 2000$ Å.

When they implemented the De Lucia & Blaizot (2007) version of the Munich model on the high-resolution MS-II, Guo et al. (2011) found it to overproduce dwarf galaxies. The authors therefore increased the efficiency of supernova feedback by introducing a direct dependence of the amount of gas reheated and ejected on the virial mass of the host halo. However, although the resulting model fits the stellar mass function of galaxies well at low redshift, it still produces more low-mass galaxies than are observed at $z > 1$. This deficiency is reflected in our results below.

Finally, Guo et al. (2011) introduced a more realistic treatment of satellite galaxy evolution and of mergers. The hot gas content of satellite galaxies is gradually stripped instead of being instantaneously removed at infall, as suggested by the simulations of McCarthy et al. (2008). This allows satellites to continue forming stars for a longer period and reduces the excessively rapid reddening of these objects. In addition, satellites of satellites remain connected to their parent galaxies and can merge with them, rather than being automatically reassigned to the central galaxy of the group or cluster. The model also includes a treatment of the tidal disruption of satellite galaxies.[1]

2.2 Light-cone construction

At high redshift, the observed fluxes at a limited number of wavelengths are often the only data available for a galaxy, so that its redshift must be inferred through comparison of the observed colours to model templates. Even rest-frame magnitudes, colours and luminosities can then be subject to substantial uncertainties, and the conversion to intrinsic properties such as masses and star formation rates is problematic. Results not only depend on the accuracy of the photometric redshift, but are also (almost) degenerate with respect to the star formation history, metallicity and dust content of the galaxy. These quantities are direct predictions of a semi-analytic model, so that the conversion from intrinsic to observed properties is, in principle, well defined, given a stellar population synthesis model, an assumed IMF and a specific model for intrinsic obscuration. It is thus often convenient to consider conversion uncertainties as part of the model and to compare theoretical predictions directly with observables. To do this, we construct light cones which allow the models to be 'observed' in a way that mimics real surveys as directly as possible. We use two different population synthesis models to predict observables in order to assess the impact of the differing mass-to-light conversions they imply. We compute observed- and rest-frame fluxes from the UV to the near-IR so that our theoretical data sets resemble those of modern observational surveys not only in volume, but also in wavelength coverage.

Our light cones are built using the MOMAF developed by Blaizot et al. (2005). We refer the reader to the original paper for a full description of the method. Here we briefly summarize the complications that arise when building light cones from simulations of limited size and resolution.

The MS has side of $500\,h^{-1}$ Mpc (comoving). This is considerably smaller than, for example, the comoving distance to a galaxy observed at $z \sim 2$. Periodic replication of the simulation can lead to multiple appearances of the same object within the light cone, although typically at different redshifts and so with different properties and at offset positions (due to large-scale motions). Blaizot et al. (2005) suggested applying a series of transformations (rotations, translations and inversions) when tiling space with periodic replications. This does not, of course, prevent multiple appearances

[1] See Henriques & Thomas (2010) for an alternative extension of the Munich semi-analytic model modifying supernova feedback and including tidal disruption of satellites.

of a given object within the light cone, but these duplicates are then viewed from different directions and no longer fall on a (nearly) regular lattice.

Unfortunately, this technique also introduces discontinuities in large-scale structure at the boundaries between replications, affecting clustering statistics in a way which is at least as difficult to model as that of the original periodicity. Kitzbichler & White (2007) showed that for light cones of relatively small solid angle, the central line of sight can be chosen to pass through the lattice of periodic replications in such a direction that multiple images of the same object are minimized or eliminated altogether. The latter is not possible if the comoving volume of the light cone exceeds that of the simulation, but this technique can still be used to ensure that multiple appearances occur as far apart as possible both on the sky and in redshift. We therefore use the method of Kitzbichler & White (2007) in this paper. Space is filled with periodic replications of the simulation, a position is chosen for the observer, and the central line of sight of the survey field is given a previously chosen orientation. Galaxies whose positions intercept the light cone are selected and their comoving distance is converted into a redshift.

As explained in Kitzbichler & White (2007), the time between stored snapshots for the MS varies between 100 and 380 Myr. This means that the intrinsic properties of galaxies are not generally available at the time corresponding to their comoving distance. Rather, they must be taken from the stored snapshot which is closest to their light-cone position. Hence, galaxies with redshift $(z_i + z_{i-1})/2 < z < (z_i + z_{i+1})/2$ are assigned the physical properties stored at z_i. The resulting discontinuity in galaxy population properties, at the boundaries between snapshots, could be reduced by interpolating, but this works poorly for positions and velocities since the output separation is comparable to orbital times within groups and clusters. Moreover, it is not straightforward for other galaxy properties either, since these change discontinuously on time-scales shorter than the output spacing, for example through mergers and starbursts. We thus follow Kitzbichler & White (2007) and do not attempt any interpolation. The semi-analytic calculations are performed on these intermediate time-steps that vary between 5 and 15 Myr. This means, for example, that a burst of star formation will have this duration and can happen anywhere between (or at) output snapshots, with the corresponding increase in flux being reflected in galaxy properties at the snapshot.

The *apparent* luminosities and colours of galaxies depend strongly on their redshifts through the conversion between rest- and observed-frame photometric bands and through the inverse square dependence of apparent luminosity on distance. The final redshift of the galaxy in the light cone is not available at the time observed-frame luminosities are computed in the semi-analytic model. However, there will be two extreme redshifts that bracket it. We compute apparent observed-frame luminosities (for fixed intrinsic properties) using these upper and lower limits, and once the galaxy is placed in the light cone, we interpolate to obtain final observed-frame quantities.

For this paper, we construct light cones for square areas of $1.4 \times 1.4\,\mathrm{deg}^2$ out to high redshift with no faint magnitude cut. They are, however, limited by the mass resolution of the dark matter simulation ($1.7 \times 10^{10}\,h^{-1}\,\mathrm{M}_\odot$ in halo mass) corresponding to stellar masses of $\sim 10^{9.5}\,\mathrm{M}_\odot$ at $z = 0$. While this does not matter for the questions we study in this paper, it should be borne in mind if the light cones are used for other purposes.[2]

[2] The light cones are publicly available at http://www.mpa-garching.mpg.de/millennium.

2.3 Stellar populations and photometry

2.3.1 Stellar population synthesis models

Semi-analytic models predict intrinsic properties of galaxies, such as stellar mass, star formation history, gas and dust content and metallicity. In order to convert these into observed SEDs or broad-band photometry, evolutionary population synthesis and dust models are required. The former predict the evolution of the light associated with a single short burst of star formation of given metallicity and with an assumed IMF, the so-called SSP. The intrinsic stellar emission from a model galaxy is then represented as a superposition of SSPs weighted according to its star formation history. This emission must be processed through a dust model in order to predict the observable stellar emission. Uncertainties in the conversion between mass and light can jeopardize any comparison between theory and observations. There are still significant differences between published evolutionary population synthesis models and these should be considered as part of the systematic uncertainties when comparing semi-analytic model predictions to data. Throughout this paper, we present results from two distinct stellar population synthesis codes: one that has been traditionally used in the Munich model (Bruzual & Charlot 2003) and the Maraston (2005) model implemented in the semi-analytic code by Henriques et al. (2011). In both cases, we adopt the same Chabrier (2003) IMF and a similar metallicity grid. We hope that the differences we find will give some indication of the impact of mass-to-light conversion uncertainties on galaxy formation modelling.

2.3.2 Photometry

In order to increase the predictive power of the model and allow it to be tested against a wider range of observations, we also expand the number of photometric bands for which fluxes are computed, covering all wavelengths dominated by direct emission from stars, from the UV to the near-IR IRAC bands. In Fig. 1, we plot the relevant filter transmission curves. In the top panel, we show the *GALEX* far-UV (FUV) and near-UV (NUV), the Johnson $U, B, V, R_\mathrm{C}, I_\mathrm{C}, Z, Y, J, H, K_\mathrm{s}, K$ and the IRAC 3.6, 4.5, 5.8 and 8.0 μm bands; in the second panel, we show the SDSS u, g, r, i, z bands; in the third panel, bands from *Hubble Space Telescope* (*HST*) instruments, three UV bands from the WFC3 UVIS (0.225, 0.275 and 0.336 μm), seven optical bands from the ACS WFC (0.435, 0.475, 0.606, 0.625, 0.775, 0.814 and 0.850 μm) and three near-IR bands from the WFC3 IR (1.05, 1.25 and 1.60 μm) are shown; and in the bottom panel, we show the VIMOS U band, the two NICMOS near-IR bands (1.1 and 1.6 μm) and two *HST* WFPC2 bands (0.30 and 0.45 μm). While we will not give results for all these bands in this paper, we will include the relevant apparent magnitudes in our light-cone catalogues in order to enhance their utility to others.

All magnitudes are in the AB system. In order to be as close as possible to observations, we use the K_s band when presenting results for number counts and redshift distributions and the K band when discussing the evolution of the rest-frame luminosity function.

The light cones constructed and made public in this work provide a useful tool to test observational derivations of intrinsic galaxy properties. The wide wavelength coverage of observed- and rest-frame photometry, together with the two stellar population synthesis models considered, can be used to check derivations of rest-frame magnitudes from observed photometry, as well as the reliability of

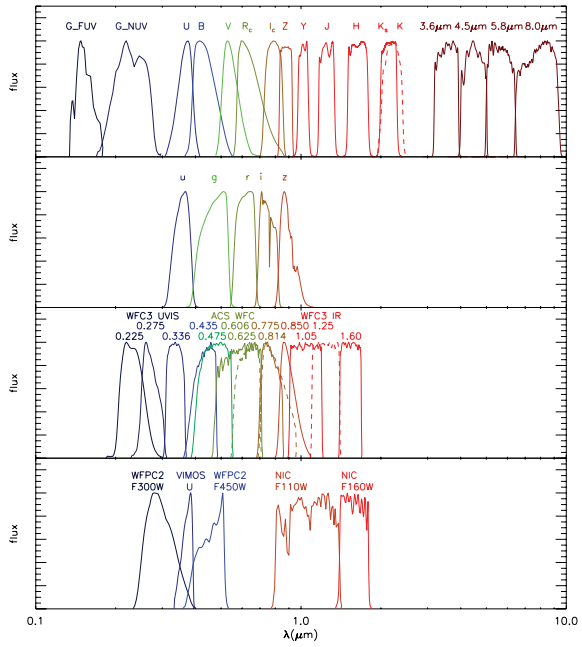

Figure 1. The response functions of the filters for which fluxes are computed for the light cones produced in this paper. These extend from the FUV to the IRAC bands and include *GALEX* FUV and NUV, Johnson $UBVR_CI_CJHK_sK$ and IRAC bands (in the top panel), SDSS *ugriz* (in the second panel), *HST* WFC3 UV and IR and ACS WFC (in the third panel) and *HST* WFPC2, VIMOS *U* and NICMOS (in the bottom panel).

properties obtained from SED fitting, such as stellar masses, ages and star formation histories.

3 RESULTS

In this section, we compare predictions of our models to observational data. We start with number counts as a function of apparent magnitude in a wide range of photometric bands (from the optical blue to the IRAC bands) and move on to redshift distributions for *K* and IRAC 5.8 μm selected galaxies. Finally, we investigate the evolution of the rest-frame optical and near-IR luminosity functions and colours which, although further from the directly observed quantities, allow a better understanding of galaxy evolution.

Kitzbichler & White (2007) and de la Torre et al. (2011) presented similar tests for an earlier version of the Munich semi-analytic model. Here, we take advantage of recent advances in the available observations and extend these comparisons to higher redshift and to a wider range of wavelengths. We also test the impact of population synthesis models by comparing results for the Bruzual & Charlot (2003) and the Maraston (2005) models. This follows up work by Tonini et al. (2009, 2010), Fontanot & Monaco (2010) and Henriques et al. (2011), who showed that the inclusion of near-IR emission from TP-AGB stars increases the predicted number of massive and extremely red objects at $z \sim 2$, as seems to be required by observation. Our comparison is based on a large number of light-cone realizations with areas and selection effects matching the relevant observational surveys.

3.1 Number counts

Galaxy counts in a given observed band can be difficult to interpret. At each apparent magnitude, they consist of galaxies at a wide range of redshifts and thus with correspondingly wide ranges of absolute magnitude and of emitted wavelength. Nevertheless, such counts provide an important test of models because they are directly observed and so are independent of uncertainties in redshift, *k*-correction, obscuration correction, etc.

Fig. 2 shows galaxy number counts for the *B* and *i*, *J*, K_s, IRAC 3.6, 4.5 and 5.8 μm bands (from top left to bottom right). Solid red and dashed blue lines are model predictions for the Maraston (2005) and Bruzual & Charlot (2003) stellar population synthesis models, respectively. Filled regions show the 1σ field-to-field scatter expected among surveys of area 2 deg^2, except that 100 arcmin^2 fields are assumed for the IRAC bands at faint magnitudes ($M > 18.5$).

The optical *B*- and *i*-band number counts are compared with data from the SSDS (Yasuda et al. 2001) and the VVDS (McCracken et al. 2003) at brighter magnitudes. At fainter magnitudes, we use the *Hubble Deep Field*-North (HDF-N; Capak et al. 2004) for the *B* band and the COSMOS sample (Capak et al. 2007) for the *i* band. Both population synthesis models match the data for bright galaxies (seen at low redshift in the rest-frame optical), while the Bruzual & Charlot (2003) model predicts more galaxies at faint apparent magnitudes, in better agreement with observations. Similar trends were found for previous versions of the Munich semi-analytic model (Kitzbichler & White 2007; de la Torre et al. 2011). It is difficult to draw firm conclusions from this result, however, since these faint counts correspond to rest-frame UV emission from galaxies at high redshift where uncertainties affect not only the stellar population modelling, but also the simplistic treatments of starbursts and of dust obscuration in the semi-analytic model. Conroy et al. (2010) showed, for example, that increasing the number of blue stragglers or blue horizontal branch stars increases the predicted UV emission from passive galaxies.

We compare the *J*- and K_s-band counts to observations from the *Chandra Deep Field* and the *Hubble Deep Field*-South (CDF and HDF-S; Saracco et al. 2001) from the DEEP2 and Palomar surveys (Conselice et al. 2008), and from the MOIRCS sample (Keenan et al. 2010). In addition, we show *K*-band counts based on the UKIDSS Ultra Deep Field data (UKIDSS-UDF; Cirasuolo et al. 2010). The two stellar population synthesis models give similar predictions for the *J*-band number counts which agree with the data. Both predict too many faint objects in the K_s and K bands. As shown by Guo et al. (2011), and as this paper will clarify, this is because the semi-analytic model overpredicts the abundance of low-mass galaxies at high redshift. The two stellar population synthesis models predict similar counts at both bright (low redshift, near-IR emission) and faint (high redshift, red optical emission) apparent magnitudes, but they disagree at intermediate apparent magnitudes. As we will see in more detail below, the difference is a consequence of TB-AGB emission from stars with ages of 1 or 2 Gyr which is fully included in the Maraston (2005) but not in the Bruzual & Charlot (2003) stellar population model.

Predicted number counts for the IRAC 3.6, 4.5 and 5.8 μm bands are plotted against *Spitzer* (Fazio et al. 2004), FIREWORKS (Wuyts et al. 2008) and NEWFIRM (Whitaker et al. 2011) observations. Both the models and the observations show a pronounced change in slope at an apparent magnitude near 20, but the break is stronger in the observations than in the model and occurs at slightly brighter apparent magnitudes. As a result, the models underpredict the number of bright objects (low redshift, emission longwards of the rest-frame *K* band) and overpredict the number of faint objects (high redshift, emission in the rest-frame *JHK* region). The latter underprediction is even more pronounced here than in the *K* band and again is likely due to the overabundance of lower mass galaxies

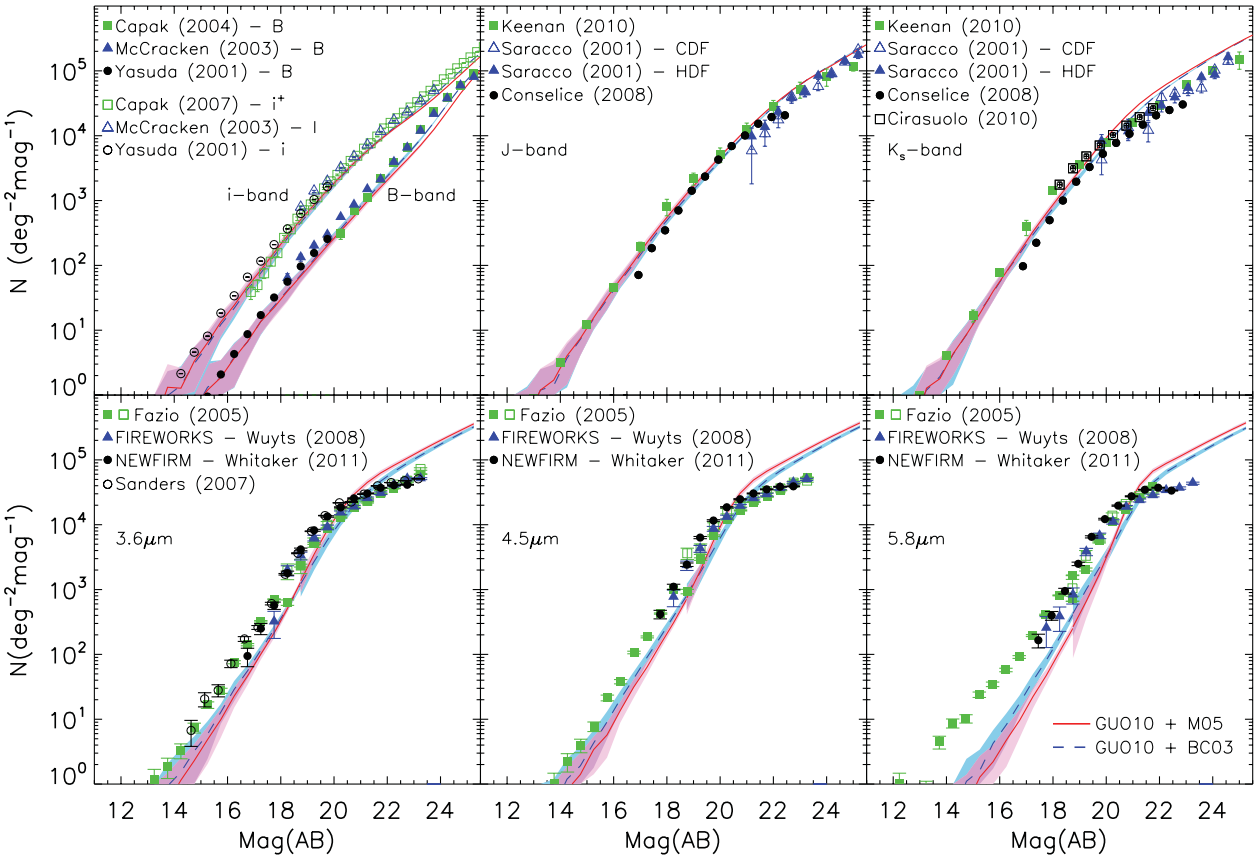

Figure 2. Galaxy number counts as a function of apparent magnitude. From top left to the bottom right, the panels show number counts in the *B* and *i*, *J*, K_s, IRAC 3.6, 4.5 and 5.8 μm bands. Theoretical predictions for the Maraston (2005) and Bruzual & Charlot (2003) population synthesis models are shown as solid red and dashed blue lines, respectively. The filled regions represent the 1σ field-to-field scatter for surveys of area 2 deg^2, except for the IRAC bands at faint magnitudes ($M > 18.5$), where 100 arcmin^2 fields are assumed. The *B*-band number counts are compared with data from the SDSS (Yasuda et al. 2001), VVDS (McCracken et al. 2003) and HDF-N (Capak et al. 2004); *i*-band counts are also compared with data from the SDSS and VVDS and with the COSMOS sample (Capak et al. 2007); for the *J* and K_s bands, we show observations from the CDF and HDF-S (Saracco et al. 2001), DEEP2 and Palomar (Conselice et al. 2008) and MOIRCS (Keenan et al. 2010) with UKIDSS-UDF data (Cirasuolo et al. 2010) plotted for the *K* band; *Spitzer* (Fazio et al. 2004), FIREWORKS (Wuyts et al. 2008) and NEWFIRM (Whitaker et al. 2011) data are shown for the $\mathrm{IRAC}_{3.6\,\mu m}$, $\mathrm{IRAC}_{4.5\,\mu m}$ and $\mathrm{IRAC}_{5.8\,\mu m}$ bands.

at $z \geq 1$ in the model. The deficit of bright galaxies is visible also in the $z = 0$ rest-frame *K*-band luminosity function (Fig. 5). Since Guo et al. (2011) tuned their semi-analytic model to match the observed low-redshift stellar mass function, this deficit implies overly large mass-to-near-IR light ratios which might be explained by overly small stellar metallicities. Indeed, Henriques & Thomas (2010) showed that the most massive low-redshift galaxies in the De Lucia & Blaizot (2007) version of the model have stellar metallicities which are too low by about a factor of 2 (the dashed red lines in their figs 4 and 10). An increase in metallicity could remove the discrepancy by reducing the mass-to-near-IR light ratios in model.[3] Another important factor, particularly for the IRAC 5.8 μm band, is the possible contamination by emission from hot dust, specifically emission in the 3.3, 6.2 and 7.7 μm polycyclic aromatic hydrocarbon (PAH) features (Draine & Li 2007; Draine et al. 2007; da Cunha, Charlot & Elbaz 2008). Such emission is not included in the models but may well be significant in the real low-redshift galaxies.

[3] The luminosity correction depends on the actual deficit in metallicity, which in turn depends strongly on which stellar population model is used to derive masses and metallicities for the observed galaxies.

3.2 Redshift distributions for *K* and $\mathrm{IRAC}_{5.8\,\mu m}$ selected samples

In Fig. 3, we compare predictions from our models to the observed photometric redshift distributions of galaxy samples selected above K_s and IRAC 5.8 μm apparent magnitude limits. As for the number counts, the solid red and dashed blue lines represent predictions based on the Maraston (2005) and Bruzual & Charlot (2003) stellar population synthesis models. Filled regions show the expected 1σ scatter among fields with an area of 100 arcmin^2. In the left-hand panel, the number of galaxies per unit area and redshift is plotted for samples with $K_s < 21.8$ and <23.3, while the right-hand panel gives similar results but for samples with IRAC 5.8 μm apparent magnitude brighter than 21.8 (and $K_s < 23.0$). Our theoretical predictions are compared with data from two public catalogues: the FIREWORKS data for the Great Observatories Origins Deep Survey–Chandra Deep Field (GOODS-CDF; Wuyts et al. 2008) and the NEWFIRM Medium-Band Survey data for the COSMOS and All-Wavelength Extended Groth Strip International Survey (AEGIS) fields (Whitaker et al. 2011). The wide photometric coverage of these two data sets results in robust and relatively precise photometric redshift measurements. Our theoretical samples are selected using photometric criteria very similar to those defining

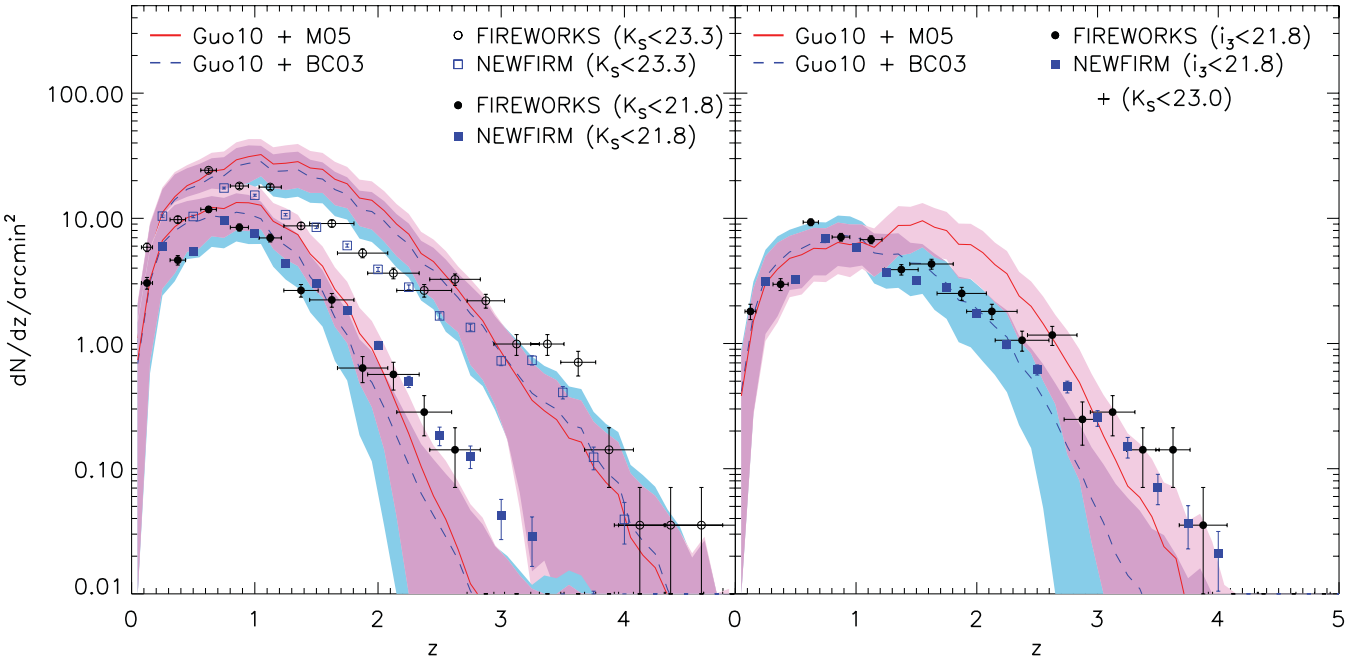

Figure 3. The redshift distribution of galaxies selected above observed-frame K_s-band (left-hand panel) and IRAC 5.8 μm band (right-hand panel) apparent magnitude limits. The solid red and dashed blue lines represent the mean predictions of our semi-analytic model for the Maraston (2005) and Bruzual & Charlot (2003) stellar population synthesis models. The filled regions show the 1σ scatter among fields with area 100 arcmin^2. The left-hand panel shows the total number of galaxies per unit area and redshift for samples with $K_s < 21.8$ and <23.3. Similar curves are shown in the right-hand panel but for samples with IRAC$_{5.8\,\mu m} < 21.8$ and $K_s < 23.0$. Theoretical distributions are compared with data from FIREWORKS (Wuyts et al. 2008) and from the NEWFIRM Medium-Band Survey (Whitaker et al. 2011).

the observed samples, although we plot the distribution of their true redshifts rather than attempting to reproduce the observational redshift estimation procedure.

Selection by observed-frame K_s-band magnitude picks galaxies on the basis of their rest-frame K-band emission at low redshift, their rest-frame J-band emission at $z \sim 1$ and their rest-frame optical emission at $z > 1.5$. For both magnitude limits, the samples are dominated by intrinsically faint objects at low redshift but by intrinsically bright galaxies beyond $z \sim 1$ (for $K_s < 21.8$) or $z \sim 1.5$ (for $K_s < 23.3$)). For both apparent magnitude limits, the predictions for the two population synthesis models agree, and for the brighter limit they are consistent with the observations out to a redshift of almost 2. Both underpredict the counts at higher redshift, with the effect being slightly larger for the Bruzual & Charlot (2003) model. This may reflect an underabundance of intrinsically bright objects (in the optical) in the model at these redshifts, but could also be due to magnitude and photometric redshift errors in the data which primarily affect the tails of the distribution. For the fainter apparent magnitude limit, the model clearly overestimates the number of objects over the redshift range $1 < z < 2.5$. These galaxies typically have stellar masses of the order of a few 10^{10} M$_\odot$ and this discrepancy reflects the overabundance of objects of this mass and redshift unity that was flagged by Guo et al. (2011). The same problem was identified in earlier versions of the model by Kitzbichler & White (2007) and de la Torre et al. (2011). At higher and lower redshifts, the abundances agree quite well in model and data, reflecting the fact that the semi-analytic model was tuned to fit galaxy abundances at low redshift, and predict an abundance of high-mass galaxies which fits observed estimates quite well at high redshift.

For the IRAC 5.8 μm selected samples shown in the right-hand panel of Fig. 3, there is a significant difference between the predictions of the two population synthesis models. While the Bruzual & Charlot (2003) model predicts a distribution with similar shape to those in the left-hand panel, the Maraston (2005) model makes a concordant prediction only at $z < 1.2$. Beyond this point there is a 'bump' and at higher redshift it predicts roughly three times as many galaxies as the Bruzual & Charlot (2003) model. A corresponding bump is not present in the observational data which are better described by the Bruzual & Charlot (2003) model, at least out to $z \sim 2.5$. The bump in the Maraston (2005) model is caused by strong rest-frame *JHK* emission from TP-AGB stars associated with intermediate-age stellar populations. While this emission brings the predicted numbers of galaxies into rough agreement with the data at the highest redshifts, it results in an overabundance at $z \sim 2$ where the observational data sets appear most robust. Since this effect is also present for data in other wavebands, for which TP-AGB emission is not an issue, it suggests that it results from the semi-analytic model overpredicting the abundance of the relevant moderate-mass galaxies by a substantial factor at this redshift. In view of this, the agreement achieved by the Bruzual & Charlot (2003) model is probably coincidental, resulting from the overestimated abundance of moderate-mass galaxies being compensated by an overestimate of their rest-frame *JHK* mass-to-light ratios.

3.3 The rest-frame *B*-band luminosity function

Rest-frame luminosity functions and colour distributions as a function of redshift provide direct estimates of the abundance evolution of various galaxy types (e.g. star-forming/passive, high/low mass). However, they require accurate redshifts and appropriate photometry if they are to be determined reliably from observed-frame fluxes. The wide wavelength coverage of modern surveys produces robust photometric redshifts, and, in addition, allows rest-frame optical

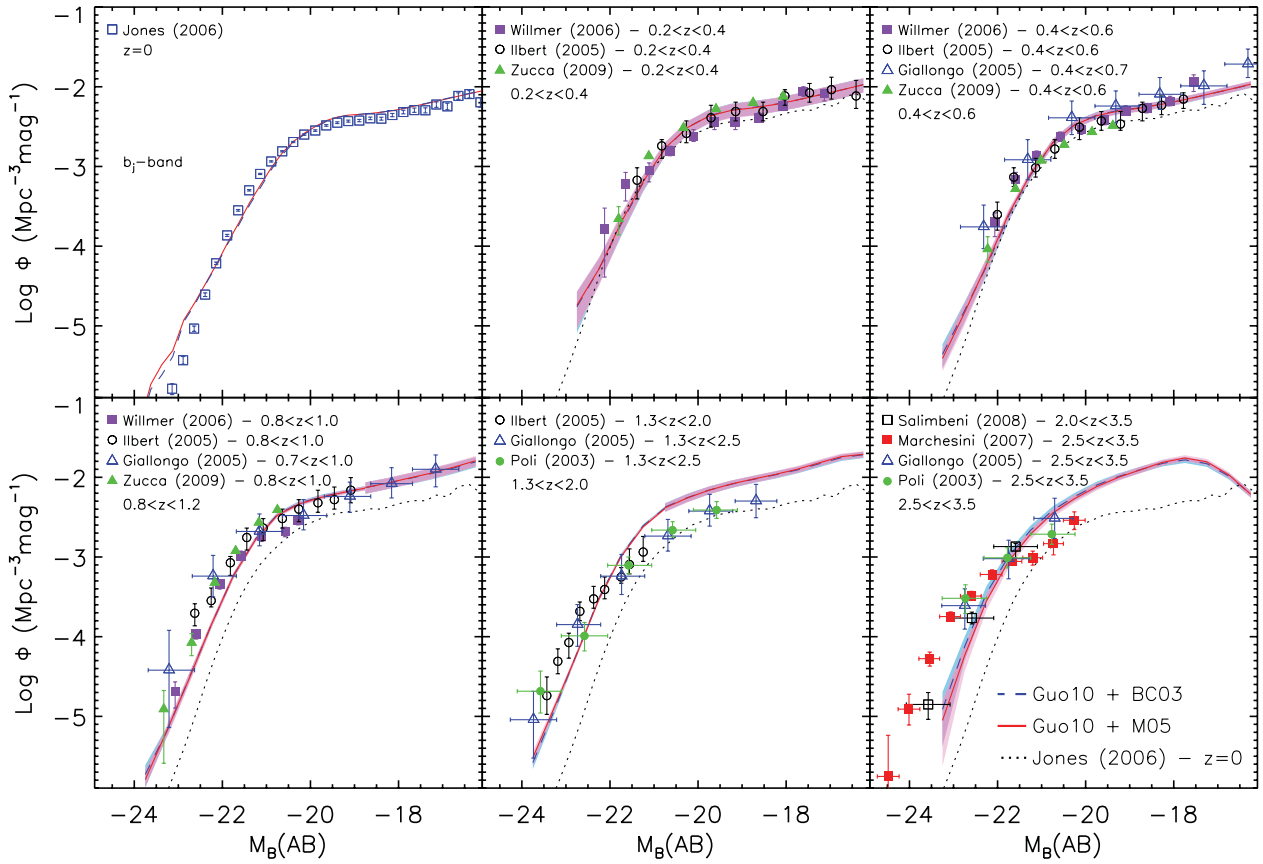

Figure 4. Evolution of the rest-frame B-band luminosity function from $z = 3$ to 0. Theoretical predictions for the Maraston (2005) and Bruzual & Charlot (2003) stellar population models are shown as solid red and dashed blue lines, respectively. Filled regions represent the 1σ field-to-field scatter expected for surveys of area 1.4 deg^2, except in the highest redshift panel, where 150 arcmin^2 fields are assumed. Fields of this size are also assumed for the intrinsically fainter galaxies in the $0.8 < z < 1.2$ and $1.3 < z < 2.0$ panels (for galaxies with $M_B > -19.0$ and > -21.0, respectively). At $z = 0$, the model b_j-band luminosity function [$b_j = B - 0.267(B - V)$; Norberg et al. 2002] is compared with observations from the 6DFGRS (Jones et al. 2006), repeated at all redshifts as a dotted black line. At higher redshifts, we show observational estimates from VVDS (Ilbert et al. 2005), DEEP2 (Willmer et al. 2006), zCOSMOS (Zucca et al. 2009), HDF-S (Poli et al. 2003), HDF-N (Giallongo et al. 2005), GOODS-MUSYC plus FIRES (Marchesini et al. 2007) and GOODS-MUSYC (Salimbeni et al. 2008).

and near-IR magnitudes to be determined by interpolation over the full range $0 < z < 4$, rather than requiring an uncertain extrapolation based on an SED fit.

Guo et al. (2011) showed that their semi-analytic model reproduces observed $z \sim 0.1$ luminosity functions in the SDSS g, r, i and z bands. At higher redshift, they implemented the redshift-dependent dust model of Kitzbichler & White (2007). This reproduces the observed abundance of colour-selected galaxies at $z \sim 2$ and 3 for the previous version of the semi-analytic model (see Guo & White 2009). In Fig. 4, we show the evolution of the B-band luminosity function from $z = 0$ to 3 for our current semi-analytic model. Solid red and dashed blue lines represent versions with the Maraston (2005) and Bruzual & Charlot (2003) stellar population models, respectively. Filled regions give the expected 1σ field-to-field scatter for surveys of area 1.4 deg^2, except that smaller fields (with area 150 arcmin^2) were assumed for the $2.5 < z < 3.5$ panel, for the $1.3 < z < 2.0$ panel fainter than -21.0 and for the $0.8 < z < 1.2$ panel fainter than -19.0.

At $z \sim 0$, the model b_j-band luminosity function[4] is compared with the 6 Degree Field Galaxy Redshift Survey (6DFGRS) result of Jones et al. (2006), repeated for reference as a black dotted line in all panels. For the $z \sim 0$ panel, we use the final snapshot of the simulation rather than the light cone, finding excellent agreement with the 6DFGRS result, just as was the case for the corresponding SDSS luminosity function (in the g band) in Guo et al. (2011). For the other $z \leq 1$ panels, we compare with data from the relatively wide VVDS (Ilbert et al. 2005), DEEP2 (Willmer et al. 2006) and zCOSMOS (Zucca et al. 2009) surveys. At higher redshift, only data for smaller fields are available. We show results from Poli et al. (2003, HDF-S), Giallongo et al. (2005, HDF-N), Marchesini et al. [2007, GOODS-Multiwavelength Survey by Yale–Chile (MUSYC)+Faint Extragalactic Infrared Survey (FIRES)] and Salimbeni et al. (2008, GOODS-MUSYC).

The model reproduces the evolution of the rest-frame B-band luminosity function reasonably well out to $z = 3$. It overpredicts the abundance of faint objects at $z \sim 2$ and underpredicts the abundance of bright objects at $z \sim 3$, although it may still be compatible with the data given the relatively large error bars quoted by the observers and the substantial scatter between the observational determinations. The two stellar population synthesis models give very similar results in this band. We note that the predicted fluxes are strongly affected by dust, and so are dependent on the adopted dust model. Further testing of the simplistic and relatively poorly motivated model of Kitzbichler & White (2007) is clearly needed.

[4] We assume $b_j = B - 0.267(B - V)$ (Norberg et al. 2002).

3.4 The *K*-band luminosity function

The *K*-band luminosity function has long been thought of as a proxy for the stellar mass function. Recent results have shown, however, that this assumption, while moderately accurate at low redshift, can break down badly at early times. Notably, the fact that the characteristic luminosity $L_\star$ increases with increasing redshift just as for the optical bands (Cirasuolo et al. 2010) is inconsistent with a time-independent *K*-band mass-to-light ratio, which would imply the 'evaporation' of material from the most massive galaxies. This luminosity function behaviour is easily understood in the context of recent stellar population synthesis models. In particular, a significant amount of *K*-band emission comes not from the old populations which dominate the stellar mass, but rather from intermediate-age stars (∼1 Gyr) passing through the TB-AGB phase (Maraston 2005; Charlot & Bruzual, in preparation). At high redshifts, these relatively young populations can dominate the rest-frame luminosity in the *K* band and they are only later replaced by predominantly old populations (Henriques et al. 2011).

In Fig. 5, we plot the evolution of the *K*-band luminosity function out to $z = 3$. As in previous figures, the solid red and dashed blue lines represent predictions based on the Maraston (2005) and Bruzual & Charlot (2003) stellar population models, and filled regions outline the expected 1σ field-to-field scatter among surveys of area 0.7 deg^2. At $z \sim 0$, the model K_s-band luminosity function is compared with observational data from the Two Micron All Sky Survy (2MASS; Bell et al. 2003) and 6DFGRS+2MASS (Jones et al. 2006). As a reference, we repeat the latter at all redshifts as a black dotted line. For the $z \sim 0$ panel, we use the final snapshot of the simulation rather than the light cone to obtain the theoretical prediction. At higher redshifts, the model is compared with data from the Munich Near-Infrared Cluster Survey (MUNICS; Drory et al. 2003), K20 (Pozzetti et al. 2003), HDF-S (Saracco et al. 2006) and UKIDSS-UDF (Cirasuolo et al. 2010). Only the last of these surveys estimates the rest-frame *K*-band flux directly by interpolating between observed-frame magnitudes at corresponding wavelengths (from *Spitzer*/IRAC). The other surveys extrapolate the *observed-frame K* flux to longer wavelength using an uncertain SED fit and thus may be subject to substantial systematic errors.

Our two population synthesis models give very similar predictions for the rest-frame *K*-band luminosity function out to $z \sim 0.5$. At higher redshifts, their shapes and normalizations remain similar but their characteristic luminosities diverge with the Maraston (2005) prediction being brighter by about 0.4, 0.5, 0.6 and 0.75 mag at redshifts of 1.0, 1.5, 2.0 and 3.0, respectively. This reflects the increasing contribution from TP-AGB stars as the mean age of the galaxies gets younger. The predictions of both models are strongly at variance with observation at these redshifts. While the Maraston

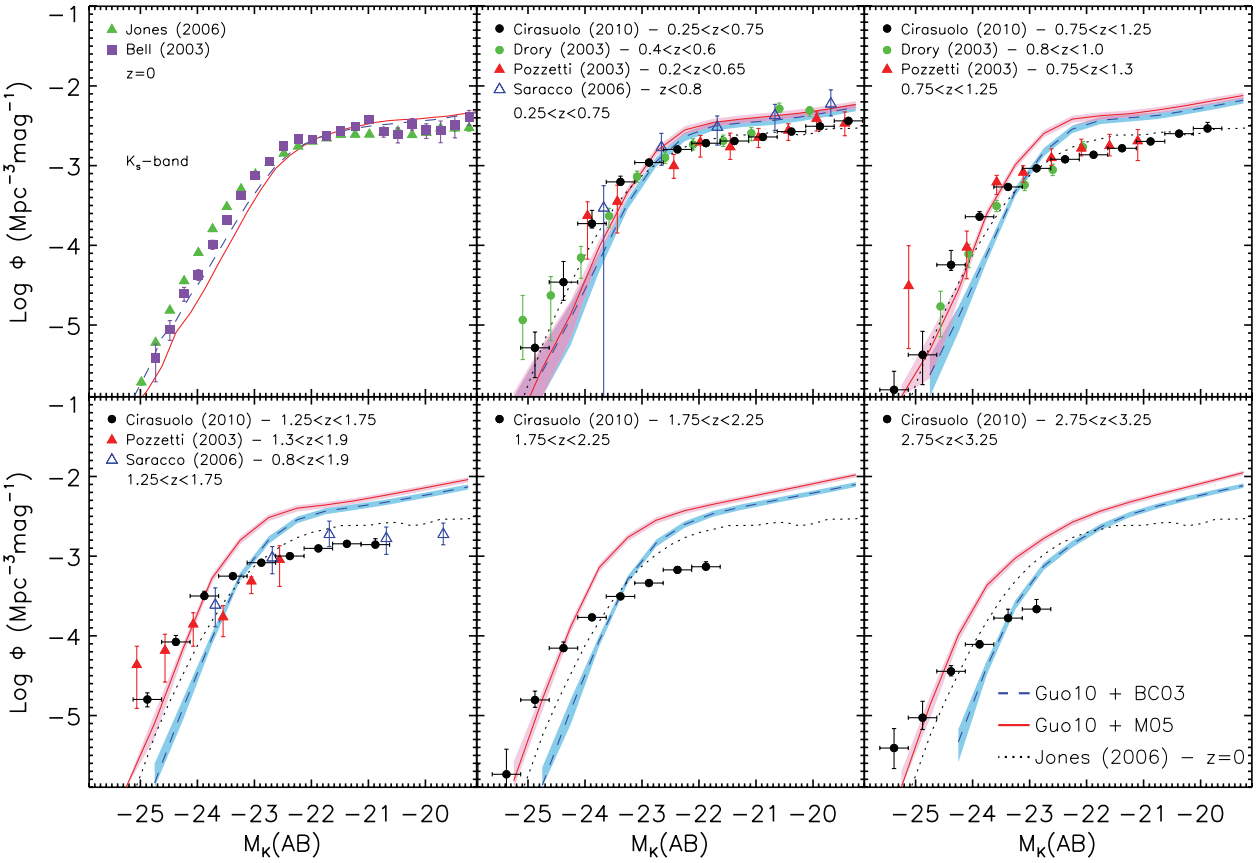

Figure 5. Evolution of the rest-frame *K*-band luminosity function from $z = 3$ to 0. Predictions of our semi-analytic model for the Maraston (2005) and Bruzual & Charlot (2003) population synthesis models are shown by solid red and dashed blue lines, respectively. Filled regions represent the expected 1σ field-to-field scatter for surveys of area 0.7 deg^2. At $z \sim 0$, the model K_s-band luminosity function is compared with data from 2MASS (Bell et al. 2003) and 6DFGRS+2MASS (Jones et al. 2006). We repeat the latter at all redshifts as a black dotted line. At higher redshifts, we show observational estimates based on MUNICS (Drory et al. 2003), the UKIDSS-UDF (Cirasuolo et al. 2010), the K20 Survey (Pozzetti et al. 2003) and the HDF-S (Saracco et al. 2006). Note that in all surveys other than the UKIDSS-UDF, the rest-frame *K* luminosities are not directly measured but are rather estimated by extrapolating from the *observed-frame K*-band fluxes using an SED model.

(2005) model agrees with the high-mass tail of the observed luminosity functions at all redshifts, it seriously overpredicts the abundance of less massive galaxies at $z = 1$ and earlier. For the lowest luminosity bin of the UKIDSS-UDF data set, the overprediction is by factors of 2, 4, 6 and almost 8 at redshifts of 1.0, 1.5, 2.0 and 3.0, respectively. The Bruzual & Charlot (2003) model fails to reproduce the rest-frame K luminosities of the most massive systems (by about 0.7 mag by $z \sim 3$) but nevertheless overpredicts the abundance of less luminous systems almost as badly as the Maraston (2005) model. This is the substantial problem already pointed out by Guo et al. (2011); their galaxy formation assumptions produce moderate-mass galaxies ($M_\star \sim 10^{10}\,\mathrm{M}_\odot$) too early to be compatible with current data on populations at $z \geq 1$.

As noted above when discussing Fig. 2, the model also slightly underestimates the K-band luminosities of massive low-redshift galaxies. This more subtle problem is due to these massive galaxies being too blue (see the colour distributions in fig. 12 of Guo et al. 2011). This is likely caused by an underabundance of heavy elements (Henriques & Thomas 2010).

Similar results for the evolution of the rest-frame K-band luminosity function have been obtained for an earlier version of the Munich semi-analytic model (De Lucia & Blaizot 2007, as well as for the semi-analytic models of Menci et al. 2006, Monaco et al. 2007 and Fontanot et al. 2009). In a recent paper, Somerville et al. (2011) compared another independent semi-analytic model to observational data on the evolution of the rest-frame 1500 Å, B- and K-band luminosity functions. The authors were able to get a reasonable match to the bright tail without including TP-AGB emission, but they do obtain a much weaker evolution of the characteristic $L_\star$ than observed, in concordance with our Bruzual & Charlot (2003) results. They also overpredict the abundance of lower luminosity galaxies by very similar factors to those that we find here. It seems that whatever is causing the overly early formation of lower mass galaxies is common to all recent semi-analytic models.

3.5 Galaxy colours

Recent observations have shown that the local bimodality between blue, star-forming and red, passive galaxies persists at least up to $z = 2$ (e.g. Wuyts et al. 2007; Williams et al. 2009; Ilbert et al. 2010; Whitaker et al. 2011). These authors have used a combination of a rest-frame near-IR colour and an optical colour in order to separate dusty star-forming galaxies from passive objects. At fixed $U - V$, red passive galaxies will have bluer $V - J$ colours than dusty

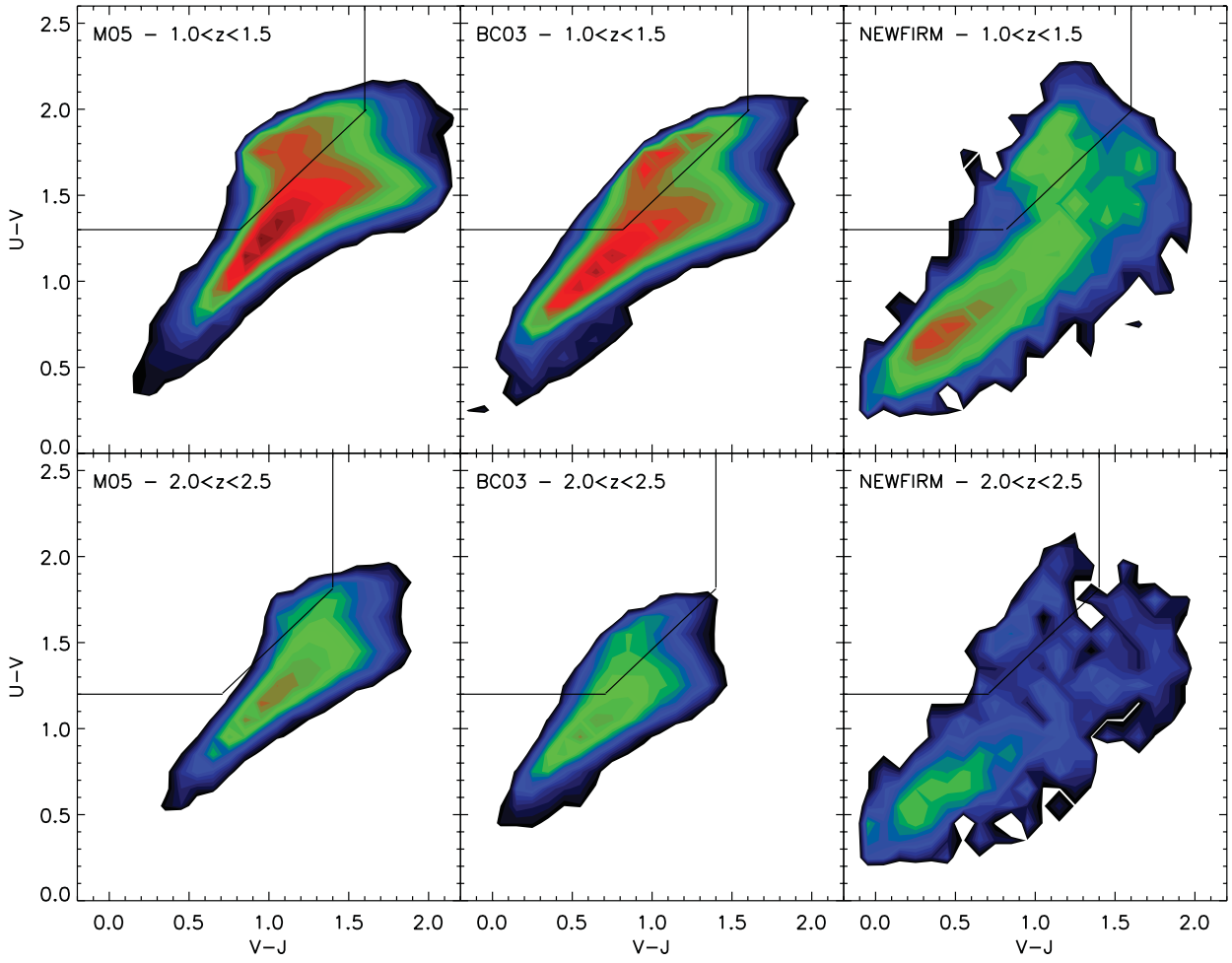

Figure 6. The rest-frame $U - V$ versus $V - J$ colour diagram for $1.0 < z < 1.5$ (top panel) and $2.0 < z < 2.5$ (bottom panel). Model predictions for the Maraston (2005) and Bruzual & Charlot (2003) stellar populations (respectively, left-hand and middle panels) are compared with NEWFIRM data from Whitaker et al. (2011) (right-hand panel). Theoretical galaxies were selected to have observed $K < 23.0$ in order to match the observational selection. The contours represent the density of points with the total number of objects normalized by the area surveyed. The solid black line shows the observational dividing line between active and passive objects.

star-forming objects. In Fig. 6, we plot rest-frame $U - V$ versus $V - J$ diagrams in two redshift bins, $1.0 < z < 1.5$ in the upper panels and $2.0 < z < 2.5$ in the bottom panels. Predictions are shown for two different stellar populations [Maraston (2005) in the left-hand panels and Bruzual & Charlot (2003) in the middle panels] and for NEWFIRM observations (Whitaker et al. 2011). Theoretical galaxies in the light cones were selected to match observations by applying a flux limit at observed K band $= 23.0$, roughly the 90 per cent completion limit quoted for observations. As described in Whitaker et al. (2011), observational galaxies were carefully deblended and only objects with signal-to-noise ratio >8 in the K band were included. The contours represent the density of points with the total number of objects normalized by the area surveyed. The solid black line shows the empirical dividing line between active and passive objects.

For both redshift bins, the models correctly predict the existence of two distinct populations, although they fail to match the exact observational spread in colour. The two populations have less scatter and are closer to each other in the models also covering a smaller range in $V - J$. This might in part result from incorrect physics in the model but it can also be explained by uncertainties in the conversion between aperture and total magnitudes, photometric redshifts and the process of SED fitting when deriving total rest-frame magnitudes from observations. At $z = 2$, the colours of red galaxies seem to be better matched by the Bruzual & Charlot (2003) prescription. The Maraston (2005) predictions are shifted to larger $V - J$ colours. Nevertheless, we note that the position of a galaxy population in this diagram is strongly dependent on the dust model assumed. For both population models, passive galaxies do not form a distinct peak, but rather a cloud of objects departing from the blue sequence towards redder $U - V$ colours [at $1.0 < V - J < 1.5$ for Maraston (2005) and $0.5 < V - J < 1.0$ for Bruzual & Charlot (2003)]. These are in fact passive galaxies in the model with almost no ongoing star formation.

The current model for galaxy formation in a hierarchical Universe predicts the most massive galaxies to grow rapidly at the centres of clusters and large groups. Their history is rich in merger events which can fuel gas into their central black holes. Feedback from these objects can then shut down star formation at early times. The redder objects that can be seen at $z = 2$ in the model have masses between 10^{11} and $10^{11.5}\,\mathrm{M}_\odot$ and black hole masses as big as $10^8\,\mathrm{M}_\odot$.

4 CONCLUSIONS

We have constructed light cones from the latest version of the Munich semi-analytic model (Guo et al. 2011) and used them to compare the model with the high-redshift galaxy population as revealed by recent deep surveys at optical and near-IR wavelengths. We have combined the model with two different stellar population synthesis packages (Bruzual & Charlot 2003; Maraston 2005) in order to understand how differences in the photometric modelling are reflected in inferences about galaxy evolution. We use multiple independent light cones to characterize cosmic variance uncertainties in currently available data sets. Our mock catalogues are made publicly available and provide observer-frame photometry in 40 commonly used photometric bands, in addition to rest-frame photometry and a variety of physical properties of the galaxies (positions, peculiar velocities, stellar masses, halo masses, sizes, morphologies, gas fractions, star formation rates, metallicities, halo properties).

We now summarize the principal conclusions from our comparison of models and data.

(i) For both stellar populations, the model matches the observed-frame B, i and J number counts but overpredicts the counts at faint magnitudes ($m_{\mathrm{AB}} > 20$) in the K_s, IRAC 3.6, 4.5 and 5.8 μm bands. This reflects the overproduction of moderate-mass galaxies (stellar masses $M_\star \sim 10^{10}\,\mathrm{M}_\odot$) at $z \geq 1$ already noted by Guo et al. (2011). The matching of the faint optical counts is fortuitous – this overproduction is masked by overly large mass-to-light ratios in the rest-frame near-UV, perhaps due to problems with the dust modelling.

(ii) At bright magnitudes ($m_{\mathrm{AB}} < 20$), the model underpredicts the counts in the three IRAC bands. This is due to an underestimation of the near-IR luminosities of low-redshift massive galaxies caused in part by the fact that such galaxies are insufficiently metal rich in the model, and in part by the model's neglect of PAH emission from hot dust (which is particularly significant at 5.8 μm).

(iii) At magnitudes where the model K_s-band counts agree with observations, the redshift distribution of K_s-selected samples is also reproduced. At fainter magnitudes where the counts are overpredicted, the excess galaxies occur primarily at $1 < z < 2.5$, again reflecting the overproduction of $M_\star \sim 10^{10}\,\mathrm{M}_\odot$ galaxies at these epochs. The two population synthesis models give similar results in both regimes.

(iv) The two population synthesis models predict different redshift distributions for galaxies selected to $m_{\mathrm{AB}} \sim 22$ in the IRAC 5.8 μm band. Emission from TP-AGB stars enhances the number of galaxies at $z > 1.5$ in the Maraston (2005) model by about a factor of 3 relative to the Bruzual & Charlot (2003) model, causing it to overpredict the observed abundance at $z \sim 2$. Although the Bruzual & Charlot (2003) model agrees with the observed redshift distribution for $0 < z < 3$, this is a result of the overabundance of moderate-mass galaxies being cancelled by an overestimate of their near-IR mass-to-light ratios.

(v) The two population synthesis give similar results for the evolution of the rest-frame B-band luminosity function, agreeing well with observation out to $z = 1.2$. At higher redshift, the agreement is less convincing. The overabundance of lower mass model galaxies starts to become evident, and there is some indication that the models underpredict the abundance of the most luminous objects. Cosmic variance and other uncertainties in the currently available data, together with dust modelling uncertainties in the model, preclude any strong conclusions.

(vi) The Maraston (2005) population model reproduces the bright tail of the rest-frame K-band luminosity function all the way out to $z \sim 3$, whereas the Bruzual & Charlot (2003) model underpredicts the near-IR luminosities of these massive galaxies by an amount which increases from about 0.3 mag at $z \sim 1$ to 0.7 mag at $z \sim 3$. The overproduction of $M_\star \sim 10^{10}\,\mathrm{M}_\odot$ galaxies at these times causes both models to substantially overpredict galaxy abundances below the knee of the luminosity function.

(vii) The model predicts that a population of red, passive galaxies should be in place already at $z = 2$, as seen in observations. These are the most massive galaxies at the centres of clusters and large groups which can rapidly grow a central black hole capable of producing enough feedback to stop star formation at early times.

In the literature, it has often been suggested that semi-analytic models fail to reproduce the rest-frame K-band galaxy luminosities of the brightest high-redshift galaxies (at $z \sim 2$–3), and the failure is usually attributed to insufficiently rapid mass growth at early times (Pozzetti et al. 2003; Cimatti et al. 2004; Kitzbichler & White 2007; Cirasuolo et al. 2010). This study (and that of Henriques et al. 2011) suggests otherwise – current models seem fully capable

of reproducing the data, given realistic assessments of population synthesis uncertainties and of the effects of observational errors. A much more serious problem, as already pointed out in the literature (e.g. Marchesini et al. 2009, 2010; Guo et al. 2011; Somerville et al. 2011), is that the models grow somewhat lower mass galaxies too early. Objects with $M_\star \sim 10^{10}\,\mathrm{M}_\odot$ are already present with a large fraction of their $z = 0$ abundance at redshifts of 2 or 3, whereas the observations indicate a drop in abundance by about an order of magnitude. Cosmic downsizing thus appears much stronger in the real Universe than in the models. Reconciling theory and observation in the context of the Λ cold dark matter cosmology will require star formation efficiencies to scale with mass and redshift in a very different way than current models (and simulations) assume.

ACKNOWLEDGMENTS

The halo/subhalo merger trees for the Millennium and Millennium-II Simulations as well as galaxy catalogues implemented on these simulations are publicly available at http://www.mpa-garching.mpg.de/millennium. This interface was created as part of the activities of the German Astrophysical Virtual Observatory (Lemson & Virgo Consortium 2006). The light-cone mock catalogues constructed for this work are also public at the same location. The work of BH, SW and GL was supported by Advanced Grant 246797 'GALFORMOD' from the European Research Council. PT was supported by the Science and Technology Facilities Council (grant number ST/I000976/1). QG acknowledges support from the National basic research programme of China (programme 973 under grant No. 2009CB24901), the Young Researcher Grant of National Astronomical Observatories, CAS, the NSFC grants programme (No. 11143005) and the Partner Group programme of the Max Planck Society, as well as the hospitality of the Institute for Computational Cosmology in Durham, UK. The authors thank Jeremy Blaizot for providing the MOMAF code, Stijn Wuyts and Katherine Whitaker for the use of their observational data and the anonymous referee for helpful comments.

REFERENCES

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
Bell E. F., McIntosh D. H., Katz N., Weinberg M. D., 2003, ApJS, 149, 289
Blaizot J., Wadadekar Y., Guiderdoni B., Colombi S. T., Bertin E., Bouchet F. R., Devriendt J. E. G., Hatton S., 2005, MNRAS, 360, 159
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., Vernon I., Goldstein M., Benson A. J., Lacey C. G., Baugh C. M., Cole S., Frenk C. S., 2010, MNRAS, 407, 2017
Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, MNRAS, 398, 1150
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Buzzoni A., 1989, ApJS, 71, 817
Capak P. et al., 2004, AJ, 127, 180
Capak P. et al., 2007, ApJS, 172, 99
Cattaneo A., Dekel A., Devriendt J., Guiderdoni B., Blaizot J., 2006, MNRAS, 370, 1651
Chabrier G., 2003, PASP, 115, 763
Cimatti A. et al., 2004, Nat, 430, 184
Cirasuolo M., McLure R. J., Dunlop J. S., Almaini O., Foucaud S., Simpson C., 2010, MNRAS, 401, 1166
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
Conroy C., Gunn J. E., White M., 2009, ApJ, 699, 486
Conroy C., White M., Gunn J. E., 2010, ApJ, 708, 58
Conselice C. J., Bundy K., U V., Eisenhardt P., Lotz J., Newman J., 2008, MNRAS, 383, 1366
Croton D. J. et al., 2006, MNRAS, 365, 11
da Cunha E., Charlot S., Elbaz D., 2008, MNRAS, 388, 1595
Davis M. et al., 2003, in Guhathakurta P., ed., Proc. SPIE Vol. 4834, Discoveries and Research Prospects from 6- to 10-Meter-Class Telescopes II. SPIE, Bellingham, p. 161
de la Torre S. et al., 2011, A&AS, 525, A125
De Lucia G., Blaizot J., 2007, MNRAS, 375, 2
Draine B. T., Li A., 2007, ApJ, 657, 810
Draine B. T. et al., 2007, ApJ, 663, 866
Drory N., Bender R., Feulner G., Hopp U., Maraston C., Snigula J., Hill G. J., 2003, ApJ, 595, 698
Fazio G. G. et al., 2004, ApJS, 154, 39
Fioc M., Rocca-Volmerange B., 1997, A&AS, 326, 950
Fontanot F., Monaco P., 2010, MNRAS, 405, 705
Fontanot F., Somerville R. S., Silva L., Monaco P., Skibba R., 2009, MNRAS, 392, 553
Gallazzi A., Bell E. F., 2009, ApJS, 185, 253
Giallongo E., Salimbeni S., Menci N., Zamorani G., Fontana A., Dickinson M., Cristiani S., Pozzetti L., 2005, ApJ, 622, 116
Granato G. L., Lacey C. G., Silva L., Bressan A., Baugh C. M., Cole S., Frenk C. S., 2000, ApJ, 542, 710
Guo Q., White S. D. M., 2009, MNRAS, 396, 39
Guo Q. et al., 2011, MNRAS, 413, 101
Henriques B. M. B., Thomas P. A., 2010, MNRAS, 403, 768
Henriques B. M. B., Thomas P. A., Oliver S., Roseboom I., 2009, MNRAS, 396, 535
Henriques B., Maraston C., Monaco P., Fontanot F., Menci N., De Lucia G., Tonini C., 2011, MNRAS, 415, 3571
Ilbert O. et al., 2005, A&AS, 439, 863
Ilbert O. et al., 2010, ApJ, 709, 644
Jones D. H., Peterson B. A., Colless M., Saunders W., 2006, MNRAS, 369, 25
Kampakoglou M., Trotta R., Silk J., 2008, MNRAS, 384, 1414
Kauffmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201
Kauffmann G., Colberg J., Diaferio A., White S., 1999, MNRAS, 303, 188
Keenan R. C., Barger A. J., Cowie L. L., Wang W.-H., 2010, ApJ, 723, 40
Kitzbichler M. G., White S. D. M., 2007, MNRAS, 376, 2
Lacey C. G., Baugh C. M., Frenk C. S., Benson A. J., Orsi A., Silva L., Granato G. L., Bressan A., 2010, MNRAS, 405, 2
Lawrence A. et al., 2007, MNRAS, 379, 1599
Le Fèvre O. et al., 2005, A&AS, 439, 845
Leitherer C. et al., 1999, ApJS, 123, 3
Lemson G., Virgo Consortium T., 2006, preprint (arXiv e-prints)
Lu Y., Mo H. J., Weinberg M. D., Katz N. S., 2011, MNRAS, 416, 1949
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
McCracken H. J. et al., 2003, A&AS, 410, 17
Maraston C., 1998, MNRAS, 300, 872
Maraston C., 2005, MNRAS, 362, 799
Marchesini D. et al., 2007, ApJ, 656, 42
Marchesini D., van Dokkum P. G., Förster Schreiber N. M., Franx M., Labbé I., Wuyts S., 2009, ApJ, 701, 1765
Marchesini D. et al., 2010, ApJ, 725, 1277
Marigo P., Girardi L., 2007, A&AS, 469, 239
Mathis J. S., Mezger P. G., Panagia N., 1983, A&AS, 128, 212
Menci N., Fontana A., Giallongo E., Grazian A., Salimbeni S., 2006, ApJ, 647, 753
Monaco P., Fontanot F., Taffoni G., 2007, MNRAS, 375, 1189
Norberg P. et al., 2002, MNRAS, 336, 907
Poli F. et al., 2003, ApJ, 593, L1
Pozzetti L. et al., 2003, A&AS, 402, 837
Salimbeni S. et al., 2008, A&AS, 477, 763
Santini P. et al., 2012, A&AS, 538, 33
Saracco P., Giallongo E., Cristiani S., D'Odorico S., Fontana A., Iovino A., Poli F., Vanzella E., 2001, A&AS, 375, 1

Saracco P. et al., 2006, MNRAS, 367, 349
Scoville N. et al., 2007, ApJS, 172, 1
Somerville R. S., Primack J. R., 1999, MNRAS, 310, 1087
Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008, MNRAS, 391, 481
Somerville R. S., Gilmore R. C., Primack J. R., Dominguez A., 2011, preprint (arXiv e-prints)
Spergel D. N. et al., 2003, ApJS, 148, 175
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Springel V. et al., 2005, Nat, 435, 629
Thomas D., Maraston C., Bender R., 2003, MNRAS, 339, 897
Tonini C., Maraston C., Devriendt J., Thomas D., Silk J., 2009, MNRAS, 396, L36
Tonini C., Maraston C., Thomas D., Devriendt J., Silk J., 2010, MNRAS, 403, 1749
van der Wel A., Franx M., Wuyts S., van Dokkum P. G., Huang J., Rix H., Illingworth G. D., 2006, ApJ, 652, 97
Vazdekis A., Casuso E., Peletier R. F., Beckman J. E., 1996, ApJS, 106, 307
Whitaker K. E. et al., 2011, ApJ, 735, 86
White S. D. M., Frenk C. S., 1991, ApJ, 379, 52
Williams R. J., Quadri R. F., Franx M., van Dokkum P., Labbé I., 2009, ApJ, 691, 1879
Willmer C. N. A. et al., 2006, ApJ, 647, 853
Worthey G., 1994, ApJS, 95, 107
Wuyts S. et al., 2007, ApJ, 655, 51
Wuyts S., Labbé I., Schreiber N. M. F., Franx M., Rudnick G., Brammer G. B., van Dokkum P. G., 2008, ApJ, 682, 985
Yasuda N. et al., 2001, AJ, 122, 1104
Zibetti S., Charlot S., Rix H.-W., 2009, MNRAS, 400, 1181
Zucca E. et al., 2009, A&AS, 508, 1217

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