

Galaxy evolution in the infrared: comparison of a hierarchical galaxy formation model with *Spitzer* data

C. G. Lacey,^{1*} C. M. Baugh,¹ C. S. Frenk,¹ L. Silva,² G. L. Granato³ and A. Bressan³

¹*Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE*

²*INAF, Osservatorio Astronomico di Trieste, Via Tiepolo 11, I-34131 Trieste, Italy*

³*INAF, Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 2, I-35122 Padova, Italy*

Accepted 2008 January 9. Received 2008 January 9; in original form 2007 April 11

ABSTRACT

We present predictions for the evolution of the galaxy luminosity function, number counts and redshift distributions in the infrared (IR) based on the Λ CDM cosmological model. We use the combined GALFORM semi-analytical galaxy formation model and GRASIL spectrophotometric code to compute galaxy spectral energy distributions including the reprocessing of radiation by dust. The model, which is the same as that given by Baugh et al., assumes two different initial mass functions (IMFs): a normal solar neighbourhood IMF for quiescent star formation in discs, and a very top-heavy IMF in starbursts triggered by galaxy mergers. We have shown previously that the top-heavy IMF seems to be necessary to explain the number counts of faint submillimetre galaxies. We compare the model with observational data from the *Spitzer Space Telescope*, with the model parameters fixed at values chosen before *Spitzer* data became available. We find that the model matches the observed evolution in the IR remarkably well over the whole range of wavelengths probed by *Spitzer*. In particular, the *Spitzer* data show that there is strong evolution in the mid-IR galaxy luminosity function over the redshift range $z \sim 0-2$, and this is reproduced by our model without requiring any adjustment of parameters. On the other hand, a model with a normal IMF in starbursts predicts far too little evolution in the mid-IR luminosity function, and is therefore excluded.

Key words: dust, extinction – galaxies: evolution – galaxies: formation – galaxies: high-redshift – infrared: galaxies.

1 INTRODUCTION

In recent years, the evolution of galaxies at mid- (mid-IR) and far-infrared (far-IR) wavelengths has been opened up for direct observational study by IR telescopes in space. Already in the 1980s, the *IRAS* satellite surveyed the local universe in the IR, showing that much of present-day star formation is optically obscured, revealing a population of luminous and ultraluminous IR galaxies (LIRGs with total IR luminosities $L_{\text{IR}} \sim 10^{11}-10^{12} L_{\odot}$ and ULIRGs with $L_{\text{IR}} \gtrsim 10^{12} L_{\odot}$), and providing the first hints of strong evolution in the number density of ULIRGs at recent cosmic epochs (e.g. Wright, Joseph & Meikle 1984; Soifer, Houck & Neugebauer 1987a; Sanders & Mirabel 1996). The next major advance came with the discovery by *COBE* of the cosmic far-IR background which has an energy density comparable to that in the optical/near-IR background (Puget et al. 1996; Hauser et al. 1998). This implies that, over the history of the universe, as much energy has been emitted by dust in galaxies as reaches us directly in starlight, after dust ex-

inction is taken into account. This discovery made apparent the need to understand the IR as much as the optical emission from galaxies in order to have a complete picture of galaxy evolution. In particular, it is essential to understand IR emission from dust in order to understand the cosmic history of star formation, since most of the radiation from young stars must have been absorbed by dust over the history of the universe, in order to account for the far-IR background (e.g. Hauser et al. 1998).

Following these early discoveries, the *ISO* satellite enabled the first deep surveys of galaxies in the mid- and far-IR. The deepest of these surveys were in the mid-IR at 15 μm , and probed the evolution of LIRGs and ULIRGs out to $z \sim 1$, showing strong evolution in these populations, and directly resolving most of the cosmic IR background at that wavelength (Elbaz et al. 1999, 2002; Gruppioni et al. 2002). Deep *ISO* surveys in the far-IR at 170 μm (Dole et al. 2001; Patris et al. 2003) probed lower redshifts, $z \sim 0.5$. Around the same time, submillimetre observations using the Submillimetre Common-User Bolometer Array (*SCUBA*) instrument on the James Clerk Maxwell Telescope revealed a huge population of high- z ULIRGs (Smail, Ivison & Blain 1997; Hughes et al. 1998) which were subsequently found to have a redshift distribution peaking at

*E-mail: cedric.lacey@durham.ac.uk

$z \sim 2$ (Chapman et al. 2005), confirming the dramatic evolution in number density for this population seen at shorter wavelengths and lower redshifts. The submillimetre galaxies (SMGs) have been studied in more detail in subsequent *SCUBA* surveys (e.g. SHADES, Mortier et al. 2005).

Now observations using the *Spitzer* satellite (Werner et al. 2004), with its hugely increased sensitivity and mapping speed are revolutionizing our knowledge of galaxy evolution at IR wavelengths from 3.6 to 160 μm . *Spitzer* surveys have allowed direct determinations of the evolution of the galaxy luminosity function (LF) out to $z \sim 1$ in the rest-frame near-IR and to $z \sim 2$ in the mid-IR (Le Flocc'h et al. 2005; Perez-Gonzalez et al. 2005; Babbedge et al. 2006; Franceschini et al. 2006). Individual galaxies have been detected by *Spitzer* out to $z \sim 6$ (Eyles et al. 2005). In the near future, the *Herschel* satellite (Pilbratt 2003) should make it possible to measure the far-IR LF out to $z \sim 2$, and thus directly measure the total IR luminosities of galaxies over most of the history of the universe.

Accompanying these observational advances, various types of theoretical models have been developed to interpret or explain the observational data on galaxy evolution in the IR. We can distinguish three main classes of model.

(i) *Purely phenomenological models*: In these models, the galaxy LF and its evolution are described by a purely empirical expression, and this is combined with observationally based templates for the IR spectral energy distribution (SED). The free parameters in the expression for the LF are then chosen to obtain the best match to some set of observational data, such as number counts and redshift distributions in different IR bands. These parameters are purely descriptive and provide little insight into the physical processes which control galaxy evolution. Examples of these models are (Pearson & Rowan-Robinson 1996; Xu et al. 1998; Blain et al. 1999; Chary & Elbaz 2001; Franceschini et al. 2001; Rowan-Robinson 2001; Lagache, Dole & Puget 2003; Gruppioni et al. 2005).

(ii) *Hierarchical galaxy formation models with phenomenological SEDs*: In these models, the evolution of the LFs of the stellar and total dust emission are calculated from a detailed model of galaxy formation based on the cold dark matter (CDM) model of structure formation, including physical modelling of processes such as gas cooling and galaxy mergers. The stellar luminosity of a model galaxy is computed from its star formation history, and the stellar luminosity absorbed by dust, which equals the total IR luminosity emitted by dust, is calculated from this based on some treatment of dust extinction. However, the SED shapes required to calculate the distribution of the dust emission over wavelength from the total IR dust emission are either observationally based templates (e.g. Guiderdoni et al. 1998; Devriendt & Guiderdoni 2000) or are purely phenomenological, e.g. a modified Planck function with an empirically chosen dust temperature (e.g. Kaviani, Haehnelt & Kauffmann 2003). In this approach, the shape of the IR SED assumed for a model galaxy may be incompatible with its other predicted properties, such as its dust mass and radius.

(iii) *Hierarchical galaxy formation models with theoretical SEDs*: These models are similar to those of type (ii), in that the evolution of the galaxy population is calculated from a detailed physical model of galaxy formation based on CDM, but instead of using phenomenological SEDs for the dust emission, the complete SED of each model galaxy, from the far-UV to the radio, is calculated by combining a theoretical stellar population synthesis model for the stellar emission with a theoretical radiative transfer and dust heating model to predict both the extinction of starlight by dust and the IR/submillimetre SED of the dust emission. The advantages of this type of model are that it is completely *ab initio*, with the maximum possible theoretic-

cal self-consistency, and all of the model parameters relate directly to physical processes. For example, the typical dust temperature and the shape of the SED of dust emission depend on the stellar luminosity and the dust mass, and evolution in all of these quantities is computed self-consistently in this type of model. Following this modelling approach thus allows more rigorous testing of the predictions of physical models for galaxy formation against observational data at IR wavelengths, as well as shrinking the parameter space of the predictions. Examples of such models are Granato et al. (2000) and Baugh et al. (2005). (An alternative modelling approach also based on theoretical IR SEDs but with a simplified treatment of the assembly of galaxies and haloes has been presented by Granato et al. (2004) and Silva et al. (2005).)

In this paper, we follow the third approach, with physical modelling both of galaxy formation and of the galaxy SEDs, including the effects of dust. This paper is the third in a series, where we combine the GALFORM semi-analytical model of galaxy formation (Cole et al. 2000) with the GRASIL model for stellar and dust emission from galaxies (Silva et al. 1998). The GALFORM model computes the evolution of galaxies in the framework of the Λ CDM model for structure formation, based on physical treatments of the assembly of dark matter haloes by merging, gas cooling in haloes, star formation and supernova feedback, galaxy mergers and chemical enrichment. The GRASIL model computes the SED of a model galaxy from the far-UV to the radio, based on theoretical models of stellar evolution and stellar atmospheres, radiative transfer through a two-phase dust medium to calculate both the dust extinction and dust heating, and a distribution of dust temperatures in each galaxy calculated from a detailed dust grain model. In the first paper in the series (Granato et al. 2000), we modelled the IR properties of galaxies in the local universe. While this model was very successful in explaining observations of the local universe, we found subsequently that it failed when confronted with observations of star-forming galaxies at high redshifts, predicting far too few SMGs at $z \sim 2$ and Lyman-break galaxies (LBGs) at $z \sim 3$. Therefore, in the second paper (Baugh et al. 2005), we proposed a new version of the model which assumes a top-heavy IMF in starbursts (with slope $x = 0$, compared to Salpeter slope $x = 1.35$), but a normal solar neighbourhood IMF for quiescent star formation. In this new model, the star formation parameters were also changed to force more star formation to happen in bursts. This revised model agreed well with both the number counts and redshift distributions of SMGs detected at 850 μm , and with the rest-frame far-UV LF of LBGs at $z \sim 3$, while still maintaining consistency with galaxy properties in the local universe such as the optical, near-IR and far-IR LFs, and gas fractions, metallicities, morphologies and sizes.

This same model of Baugh et al. (2005) was found by Le Delliou et al. (2005, 2006) to provide a good match to the observed evolution of the population of Ly α -emitting galaxies over the redshift range $z \sim 3$ –6. Support for the controversial assumption of a top-heavy IMF in bursts came from the studies of chemical enrichment in this model by Nagashima et al. (2005a,b), who found that the metallicities of both the intergalactic gas in galaxy clusters and the stars in elliptical galaxies were predicted to be significantly lower than observed values if a normal IMF was assumed for all star formation, but agreed much better if a top-heavy IMF in bursts was assumed, as in Baugh et al. In this third paper in the series, we extend the Baugh et al. (2005) model to make predictions for galaxy evolution in the IR, and compare these predictions with observational data from *Spitzer*. We emphasize that all of the model parameters for the predictions presented in this paper were fixed by Baugh et al. prior to the publication of any results from *Spitzer*, and we have not tried

to obtain a better fit to any of the *Spitzer* data by adjusting these parameters.¹

Our goals in this paper are to test our model of galaxy evolution with a top-heavy IMF in starbursts against observations of dust-obscured star-forming galaxies over the redshift range $z \sim 0-2$, and also to test our predictions for the evolution of the stellar populations of galaxies against observational data in the rest-frame near- and mid-IR. The plan of the paper is as follows. In Section 2, we give an overview of the GALFORM and GRASIL models, focusing on how the predictions we present later on are calculated. In Section 3, we compare the galaxy number counts predicted by our model with observational data in all seven *Spitzer* bands, from 3.6 to 160 μm . In Section 4, we investigate galaxy evolution in the IR in more detail, by comparing model predictions directly with galaxy LFs constructed from *Spitzer* data. In Section 5, we present the predictions of our model for the evolution of the galaxy stellar mass function (GSMF) and galaxy star formation rate distribution (GSFRD), and investigate the insight our model offers on how well stellar masses and SFRs can be estimated from *Spitzer* data. We present our conclusions in Section 6. In Appendix A, we present model predictions for galaxy redshift distributions in the different *Spitzer* bands, to assist in interpreting data from different surveys.

2 MODEL

In this paper we use the GALFORM semi-analytical model to predict the physical properties of the galaxy population at different redshifts, and combine it with the GRASIL spectrophotometric model to predict the detailed SEDs of model galaxies. Both GALFORM and GRASIL have been described in detail in various previous papers, but since the descriptions of the different model components, as well as of our particular choice of parameters, are spread among different papers, we give an overview of both of these here. GALFORM is described in Section 2.1 and GRASIL in Section 2.2. Particularly important features of our model are the triggering of starbursts by mergers (discussed in Section 2.1.4) and the assumption of a top-heavy IMF in starbursts (discussed in Section 2.1.7). We further discuss the choice of model parameters in Section 2.3. Readers who are already familiar with the Baugh et al. (2005) model can skip straight to the results, starting in Section 3.

2.1 GALFORM galaxy formation model

We compute the formation and evolution of galaxies within the framework of the Λ CDM model of structure formation using the semi-analytical galaxy formation model GALFORM. The general methodology and approximations behind the GALFORM model are set out in detail in Cole et al. (2000) [see also the review by Baugh (2006)]. In summary, the GALFORM model follows the main processes which shape the formation and evolution of galaxies. These include: (i) the collapse and merging of dark matter haloes; (ii) the shock-heating and radiative cooling of gas inside dark haloes, leading to the formation of galaxy discs; (iii) quiescent star formation in galaxy discs; (iv) feedback both from supernova explosions

¹ A closely related model of galaxy formation obtained by applying GALFORM principles to the Millennium Simulation of Springel et al. (2005) has recently been published by Bower et al. (2006). This model differs from the current one primarily in that it includes feedback from active galactic nucleus (AGN) activity, but does not have a top-heavy IMF in bursts. We plan to investigate the IR predictions of this alternative model in a subsequent paper.

and from photoionization of the IGM; (v) chemical enrichment of the stars and gas; (vi) galaxy mergers driven by dynamical friction within common dark matter haloes, leading to the formation of stellar spheroids, and also triggering bursts of star formation. The end product of the calculations is a prediction of the numbers and properties of galaxies that reside within dark matter haloes of different masses. The model predicts the stellar and cold gas masses of the galaxies, along with their star formation and merger histories, their sizes and metallicities.

The prescriptions and parameters for the different processes which we use in this paper are identical to those adopted by Baugh et al. (2005), but differ in several important respects from Cole et al. (2000). All of these parameters were chosen by comparison with pre-*Spitzer* observational data. The background cosmology is a spatially flat CDM universe with a cosmological constant, with ‘concordance’ parameters $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $\Omega_b = 0.04$ and $h \equiv H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1}) = 0.7$. The amplitude of the initial spectrum of density fluctuations is set by the rms linear fluctuation in a sphere of radius $8 h^{-1} \text{ Mpc}$, $\sigma_8 = 0.93$. For completeness, we now summarize the prescriptions and parameters used, but give details mainly where they differ from those in Cole et al. (2000), or where they are particularly relevant to predicting IR emission from dust.

2.1.1 Halo assembly histories

As in Cole et al. (2000), we describe the assembly histories of dark matter haloes through halo merger trees which are calculated using a Monte Carlo method based on the extended Press–Schechter approach (e.g. Lacey & Cole 1993). The process of galaxy formation is then calculated separately for each halo merger tree, following the baryonic physics in all of the separate branches of the tree. As has been shown by Helly et al. (2003), the statistical properties of galaxies calculated in semi-analytical models using these Monte Carlo merger trees are very similar to those computed using merger trees extracted directly from N -body simulations.

2.1.2 Gas cooling in haloes

The cooling of gas in haloes is calculated using the same simple spherical model as in Cole et al. (2000). The diffuse gas in haloes (consisting of all of the gas which has not previously condensed into galaxies) is assumed to be shock-heated to the halo virial temperature when the halo is assembled, and then to cool radiatively by atomic processes. The cooling time depends on radius through the gas density profile, which is assumed to have a core radius which grows as gas is removed from the diffuse phase by condensing into galaxies. The gas at some radius r in the halo then cools and collapses to the halo centre on a time-scale which is the larger of the cooling time t_{cool} and the free-fall time t_{ff} at that radius. Thus, for $t_{\text{cool}}(r) > t_{\text{ff}}(r)$, we have *hot accretion*, and for $t_{\text{cool}}(r) < t_{\text{ff}}(r)$, we have *cold accretion*.² In our model, gas only accretes on to the *central* galaxy in a halo, not on to any *satellite* galaxies which share that halo. We denote all of the diffuse gas in haloes as ‘hot’, and all of the gas which has condensed into galaxies as ‘cold’.

² Note that contrary to claims by Birnboim & Dekel (2003), the process of ‘cold accretion’, if not the name, has always been part of semi-analytical models [see Croton et al. (2006) for a detailed discussion].

2.1.3 Star formation time-scale in discs

The global rate of star formation ψ in galaxy discs is assumed to be related to the cold gas mass, M_{gas} , by $\psi = M_{\text{gas}}/\tau_{*,\text{disc}}$, where the star formation time-scale is taken to be

$$\tau_{*,\text{disc}} = \tau_{*0}(V_c/200 \text{ km s}^{-1})^{\alpha_*}, \quad (1)$$

where V_c is the circular velocity of the galaxy disc (at its half-mass radius) and τ_{*0} is a constant. We adopt values $\tau_{*0} = 8$ Gyr and $\alpha_* = -3$, chosen to reproduce the observed relation between gas mass and *B*-band luminosity for present-day disc galaxies. As discussed in Baugh et al. (2005), this assumption means that the disc star formation time-scale is independent of redshift (at a given V_c), resulting in discs at high redshift that are much more gas-rich than at low redshift, and have more gas available for star formation in bursts triggered by galaxy mergers at high redshift.

2.1.4 Galaxy mergers and triggering of starbursts

In the model, all galaxies originate as central galaxies in some halo, but can then become satellite galaxies if their host halo merges into another halo. Mergers can then occur between satellite and central galaxies within the same halo, after dynamical friction has caused the satellite galaxy to sink to the centre of the halo. Galaxy mergers can produce changes in galaxy morphology and trigger bursts. We classify galaxy mergers according to the ratio of masses (including stars and gas) $M_2/M_1 \leq 1$ of the secondary to primary galaxy involved. We define mergers to be *major* or *minor* according to whether $M_2/M_1 > f_{\text{ellip}}$ or $M_2/M_1 < f_{\text{ellip}}$ (Kauffmann, White & Guiderdoni 1993). In major mergers, any stellar discs in either the primary or secondary are assumed to be disrupted, and the stars rearranged into a spheroid. In minor mergers, the stellar disc in the primary galaxy is assumed to remain intact, while all of the stars in the secondary are assumed to be added to the spheroid of the primary. We adopt a threshold $f_{\text{ellip}} = 0.3$ for major mergers, consistent with the results of numerical simulations (e.g. Barnes 1998), which reproduces the observed present-day fraction of spheroidal galaxies. We assume that major mergers always trigger a starburst if any gas is present. We also assume that minor mergers can trigger bursts, if they satisfy both $M_2/M_1 > f_{\text{burst}}$ and the gas fraction in the disc of the primary galaxy exceeds $f_{\text{gas,crit}}$. Following Baugh et al. (2005), we adopt $f_{\text{burst}} = 0.05$ and $f_{\text{gas,crit}} = 0.75$. The parameters for bursts in minor mergers were motivated by trying to explain the number of SMGs. An important consequence of assuming equation (1) for the star formation time-scale in discs, combined with the triggering of starbursts in minor mergers, is that the global SFR at high redshifts is dominated by bursts, while that at low redshifts it is dominated by quiescent discs [see Baugh et al. (2005) for a detailed discussion of these points].

In either kind of starburst, we assume that the burst consumes all of the cold gas in the two galaxies involved in the merger, and that the stars produced are added to the spheroid of the merger remnant. During the burst, we assume that star formation proceeds according to the relation $\psi = M_{\text{gas}}/\tau_{*,\text{burst}}$. For the burst time-scale, we assume

$$\tau_{*,\text{burst}} = \max[f_{\text{dyn}} \tau_{\text{dyn,sph}}; \tau_{*,\text{burst,min}}], \quad (2)$$

where $\tau_{\text{dyn,sph}}$ is the dynamical time in the newly formed spheroid. We adopt $f_{\text{dyn}} = 50$ and $\tau_{*,\text{burst,min}} = 0.2$ Gyr (these parameters were chosen by Baugh et al. (2005) to allow a simultaneous match to the submillimetre number counts and to the local 60- μm LF). The SFR in a burst thus decays exponentially with time after the galaxy merger. It is assumed to be truncated after three e-folding times

[where the e-folding time takes account of stellar recycling and feedback – see Granato et al. (2000) for details], with the remaining gas being ejected into the galaxy halo at that time.

2.1.5 Feedback from photoionization

After the intergalactic medium (IGM) has been reionized at redshift z_{reion} , the formation of low-mass galaxies is inhibited, both by the effect of the IGM pressure inhibiting collapse of gas into haloes, and by the reduction of gas cooling in haloes due to the photoionizing background. We model this in a simple way, by assuming that for $z < z_{\text{reion}}$, cooling of gas is completely suppressed in haloes with circular velocities $V_c < V_{\text{crit}}$. We adopt $V_{\text{crit}} = 60 \text{ km s}^{-1}$, based on the detailed modelling by Benson et al. (2002). We assume in this paper that reionization occurs at $z_{\text{reion}} = 6$, for consistency with Baugh et al. (2005), but increasing this to $z_{\text{reion}} \sim 10$ in line with the *Wilkinson Microwave Anisotropy Probe* 3-yr estimate of the polarization of the microwave background (Spergel et al. 2006) has no significant effect on the model results presented in this paper.

2.1.6 Feedback from supernovae

Photoionization feedback only affects very low-mass galaxies. More important for most galaxies is feedback from supernova explosions. We assume that energy input from supernovae causes gas to be ejected from galaxies at a rate

$$\dot{M}_{\text{ej}} = \beta(V_c) \psi = [\beta_{\text{reh}}(V_c) + \beta_{\text{sw}}(V_c)] \psi. \quad (3)$$

The supernova feedback is assumed to operate for both quiescent star formation in discs and for starbursts triggered by galaxy mergers, with the only difference being that we take V_c to be the circular velocity at the half-mass radius of the disc in the former case, and at the half-mass radius of the spheroid in the latter case. For simplicity, we keep the same feedback parameters for starbursts as for quiescent star formation.

The supernova feedback has two components: the *reheating* term $\beta_{\text{reh}} \psi$ describes gas which is reheated and ejected into the galaxy halo, from where it is allowed to cool again after the halo mass has doubled through hierarchical mass build-up. For this, we use the parametrization of Cole et al. (2000):

$$\beta_{\text{reh}} = (V_c/V_{\text{hot}})^{-\alpha_{\text{hot}}}, \quad (4)$$

where we adopt parameter values $V_{\text{hot}} = 300 \text{ km s}^{-1}$ and $\alpha_{\text{hot}} = 2$. The *reheating* term has the largest effect on low-mass galaxies, for which ejection of gas from galaxies flattens the faint-end slope of the galaxy LF.

The second term $\beta_{\text{sw}} \psi$ in equation (3) is the *superwind* term, which describes ejection of gas out of the halo rather than just the galaxy. Once ejected, this gas is assumed never to re-accrete on to any halo. We model the superwind ejection efficiency as

$$\beta_{\text{sw}} = f_{\text{sw}} \min[1, (V_c/V_{\text{sw}})^{-2}] \quad (5)$$

based on Benson et al. (2003). We adopt parameter values $f_{\text{sw}} = 2$ and $V_{\text{sw}} = 200 \text{ km s}^{-1}$, as in Baugh et al. (2005). The *superwind* term mainly affects higher mass galaxies, where the ejection of gas from haloes causes an increase in the cooling time of gas in haloes by reducing the gas densities. This brings the predicted break at the bright end of the local galaxy LF into agreement with observations, as discussed in Benson et al. (2003). The various parameters for supernova feedback are thus chosen in order to match the observed present-day optical and near-IR galaxy LFs, as well as the galaxy metallicity–luminosity relation.

We note that the galaxy formation model in this paper, unlike some other recent semi-analytical models, does not include AGN feedback. Instead, the role of AGN feedback in reducing the amount of gas cooling to form massive galaxies is taken by superwinds driven by supernova explosions. The first semi-analytical model to include AGN feedback was that of Granato et al. (2004), who introduced a detailed model of feedback from QSO winds during the formation phase of supermassive black holes (SMBHs), with the aim of explaining the co-evolution of the spheroidal components of galaxies and their SMBHs. The predictions of the Granato et al. model for number counts and redshift distributions in the IR have been computed by Silva et al. (2005) using the GRASIL spectrophotometric model, and compared to *ISO* and *Spitzer* data. However, the Granato et al. (2004) model has the limitations that it does not include the merging of galaxies or of dark haloes, and does not treat the formation and evolution of galactic discs. More recently, several semi-analytical models have been published which propose that heating of halo gas by relativistic jets from an AGN in an optically inconspicuous or ‘radio’ mode can balance radiative cooling of gas in high-mass haloes, thus suppressing *hot accretion* of gas on to galaxies (Bower et al. 2006; Cattaneo et al. 2006; Croton et al. 2006; Monaco, Fontanot & Taffoni 2007). However, these AGN feedback models differ in detail, and all are fairly schematic. None of these models has been shown to reproduce the observed number counts and redshifts of the faint SMGs.

The effects of our superwind feedback are qualitatively quite similar to those of the radio-mode AGN feedback. Both superwind and AGN feedback models contain free parameters, which are adjusted in order to make the model fit the bright end of the observed present-day galaxy LF at optical and near-IR wavelengths. However, since the physical mechanisms are different, they make different predictions for how the galaxy LF should evolve with redshift. Current models for the radio-mode AGN feedback are very schematic, but they have the advantage over the superwind model that the energetic constraints are greatly relaxed, since accretion on to black holes can convert mass into energy with a much higher efficiency than can supernova explosions. We will investigate the predictions of models with AGN feedback for the IR and submillimetre evolution of galaxies in a future paper.

2.1.7 The stellar initial mass function and chemical evolution

Stars in our model are assumed to form with different initial mass functions (IMFs), depending on whether they form in discs or in bursts. Both IMFs are taken to be piecewise power laws, with slopes x defined by $dN/d \ln m \propto m^{-x}$, with N the number of stars and m the stellar mass (so the Salpeter slope is $x = 1.35$), and covering a stellar mass range $0.15 < m < 120 M_{\odot}$. Quiescent star formation in galaxy discs is assumed to have a solar neighbourhood IMF, for which we use the Kennicutt (1983) parametrization, with slope $x = 0.4$ for $m < M_{\odot}$ and $x = 1.5$ for $m > M_{\odot}$. (The Kennicutt 1983 IMF is similar to other popular parametrizations of the solar neighbourhood IMF, such as that of Kroupa (2001).) Bursts of star formation triggered by galaxy mergers are assumed to form stars with a top-heavy IMF with slope $x = 0$. As discussed in detail in Baugh et al. (2005), the top-heavy IMF in bursts was found to be required in order to reproduce the observed number counts and redshift distributions of the faint SMGs. Furthermore, as shown by Nagashima et al. (2005a,b), the predicted chemical abundances of the X-ray emitting gas in galaxy clusters and of the stars in elliptical galaxies also agree better with observational data in a model with the top-

heavy IMF in bursts, rather than a universal solar neighbourhood IMF.

A variety of other observational evidence has accumulated which suggests that the IMF in some environments may be top-heavy compared to the solar neighbourhood IMF. Rieke et al. (1993) argued for a top-heavy IMF in the nearby starburst M82, based on modelling its integrated properties, while Parra et al. (2007) found possible evidence for a top-heavy IMF in the ultraluminous starburst Arp220 from the relative numbers of supernovae of different types observed at radio wavelengths. Evidence has been found for a top-heavy IMF in some star clusters in intensely star-forming regions, both in M82 (e.g. McCrady, Gilbert & Graham 2003), and in our own Galaxy (e.g. Figer et al. 1999; Stolte et al. 2005; Harayama, Eisenhauer & Martins 2008). Observations of both the old and young stellar populations in the central 1 pc of our Galaxy also favour a top-heavy IMF (Paumard et al. 2006; Maness et al. 2007). Fardal et al. (2007) found that reconciling measurements of the optical and IR extragalactic background with measurements of the cosmic star formation history also seemed to require an average IMF that was somewhat top-heavy. Finally, van Dokkum (2007) found that reconciling the colour and luminosity evolution of early-type galaxies in clusters also favoured a top-heavy IMF. Larson (1998) summarized other evidence for a top-heavy IMF during the earlier phases of galaxy evolution, and argued that this could be a natural consequence of the temperature dependence of the Jeans mass for gravitational instability in gas clouds. Larson (2005) extended this to argue that a top-heavy IMF might also be expected in starburst regions, where there is strong heating of the dust by the young stars.

In our model, the fraction of star formation occurring in the burst mode increases with redshift (see Baugh et al. 2005), so the average IMF with which stars are being formed shifts from being close to a solar neighbourhood IMF at the present-day to being very top-heavy at high redshift. In this model, 30 per cent of star formation occurred in the burst mode when integrated over the past history of the universe, but only 7 per cent of the current stellar mass was formed in bursts, because of the much larger fraction of mass recycled by dying stars for the top-heavy IMF. We note that our predictions for the IR and submillimetre luminosities of starbursts are not sensitive to the precise form of the top-heavy IMF, but simply require a larger fraction of $m \sim 5\text{--}20 M_{\odot}$ stars relative to a solar neighbourhood IMF.

In this paper, we calculate chemical evolution using the instantaneous recycling approximation, which depends on the total fraction of mass recycled from dying stars (R), and the total yield of heavy elements (p). Both of these parameters depend on the IMF. We use the results of stellar evolution computations to calculate values of R and p consistent with each IMF [see Nagashima et al. (2005a) for details of the stellar evolution data used]. Thus, we use $R = 0.41$ and $p = 0.023$ for the quiescent IMF, and $R = 0.91$ and $p = 0.15$ for the burst IMF. Our chemical evolution model then predicts the masses and total metallicities of the gas and stars in each galaxy as a function of time.

2.1.8 Galaxy sizes and dust masses

For calculating the extinction and emission by dust, it is essential to have an accurate calculation of the dust optical depths in the model galaxies, which in turn depends on the mass of dust and the size of the galaxy. The dust mass is calculated from the gas mass and metallicity predicted by the chemical enrichment model, assuming that the dust-to-gas ratio is proportional to metallicity, normalized to match the

local interstellar medium (ISM) value at solar metallicity. The sizes of galaxies are computed exactly as in Cole et al. (2000): gas which cools in a halo is assumed to conserve its angular momentum as it collapses, forming a rotationally supported galaxy disc; the radius of this disc is then calculated from its angular momentum, including the gravity of the disc, spheroid (if any) and dark halo. Galaxy spheroids are built up both from pre-existing stars in galaxy mergers, and from the stars formed in bursts triggered by these mergers; the radii of spheroids formed in mergers are computed using an energy conservation argument. In calculating the sizes of discs and spheroids, we include the adiabatic contraction of the dark halo due to the gravity of the baryonic components. This model was tested for discs by Cole et al. (2000) and for spheroids by Almeida, Baugh & Lacey (2007) (see also Coenda et al., in preparation and Gonzalez et al., in preparation). During a burst, we assume that the gas and stars involved in the burst have a distribution with the same half-mass radius as the spheroid (i.e. $\eta = 1$ in the notation of Granato et al. (2000), who used a value $\eta = 0.1$).

2.2 GRASIL model for stellar and dust emission

For each galaxy in our model, we compute the SED using the spectrophotometric model GRASIL (Silva et al. 1998; Granato et al. 2000). GRASIL computes the emission from the stellar population, the absorption and emission of radiation by dust, and also radio emission (thermal and synchrotron) powered by massive stars (Bressan, Silva & Granato 2002).

2.2.1 SED model

The main features of the GRASIL model are as follows.

(i) The stars are assumed to have an axisymmetric distribution in a disc and a bulge. Given the distribution of stars in age and metallicity (obtained from the star formation and chemical enrichment history), the SED of the stellar population is calculated using a population synthesis model based on the Padova stellar evolution tracks and Kurucz model atmospheres (Bressan, Granato & Silva 1998). This is done separately for the disc and bulge.

(ii) The cold gas and dust in a galaxy are assumed to be in a two-phase medium, consisting of dense gas in giant molecular clouds embedded in a lower density diffuse component. In a quiescent galaxy, the dust and gas are assumed to be confined to the disc, while for a galaxy undergoing a burst, the dust and gas are confined to the spheroidal burst component.

(iii) Stars are assumed to be born inside molecular clouds, and then to leak out into the diffuse medium on a time-scale t_{esc} . As a result, the youngest and most massive stars are concentrated in the dustiest regions, so they experience larger dust extinctions than older, typically lower mass stars, and dust in the clouds is also much more strongly heated than dust in the diffuse medium.

(iv) The extinction of the starlight by dust is computed using a radiative transfer code; this is used also to compute the intensity of the stellar radiation field heating the dust at each point in a galaxy.

(v) The dust is modelled as a mixture of graphite and silicate grains with a continuous distribution of grain sizes (varying between 8 Å and 0.25 μm), and also polycyclic aromatic hydrocarbon (PAH) molecules with a distribution of sizes. The equilibrium temperature in the local interstellar radiation field is calculated for each type and size of grain, at each point in the galaxy, and this information is then used to calculate the emission from each grain. In the case of very small grains and PAH molecules, temperature fluctuations

are important, and the probability distribution of the temperature is calculated. The detailed spectrum of the PAH emission is obtained using the PAH cross-sections from Li & Draine (2001), as described in Vega et al. (2005). The grain size distribution is chosen to match the mean dust extinction curve and emissivity in the local ISM, and is not varied, except that the PAH abundance in molecular clouds is assumed to be 10^{-3} of that in the diffuse medium (Vega et al. 2005).

(vi) Radio emission from ionized gas in H II regions and from synchrotron radiation from relativistic electrons accelerated in supernova remnant shocks are calculated as described in Bressan et al. (2002).

The output from GRASIL is then the complete SED of a galaxy from the far-UV to the radio (wavelengths $100 \text{ Å} \lesssim \lambda \lesssim 1 \text{ m}$). The SED of the dust emission is computed as a sum over the different types of grains, having different temperatures depending on their size and their position in the galaxy. The dust SED is thus intrinsically multitemperature. GRASIL has been shown to give an excellent match to the measured SEDs of both quiescent (e.g. M51) and starburst (e.g. M82) galaxies (Silva et al. 1998; Bressan et al. 2002).

The assumption of axisymmetry in GRASIL is a limitation when considering starbursts triggered by galaxy mergers. However, observations of local ULIRGs imply that most of the star formation happens in a single burst component after the galaxy merger is substantially complete, so the assumption of axisymmetry for the burst component may not be so bad.

2.2.2 GRASIL parameters

The main parameters in the GRASIL dust model are the fraction f_{mc} of the cold gas which is in molecular clouds, the time-scale t_{esc} for newly formed stars to escape from their parent molecular cloud, and the cloud masses M_c and radii r_c in the combination M_c/r_c^2 , which determines the dust optical depth of the clouds. We assume $f_{\text{mc}} = 0.25$, $M_c = 10^6 M_{\odot}$ and $r_c = 16 \text{ pc}$ as in Granato et al. (2000), and also adopt the same geometrical parameters as in that paper. We make the following two changes in GRASIL parameters relative to Granato et al., as discussed in Baugh et al. (2005). (i) We assume $t_{\text{esc}} = 1 \text{ Myr}$ in both discs and bursts (instead of the Granato et al. values $t_{\text{esc}} = 2$ and 10 Myr , respectively). This value was chosen in order to obtain a better match of the predicted rest-frame far-UV LF of galaxies at $z \sim 3$ to that measured for LBGs. (b) The dust emissivity law in bursts at long wavelengths is modified from $\epsilon_{\nu} \propto \nu^{-2}$ to $\epsilon_{\nu} \propto \nu^{-1.5}$ for $\lambda > 100 \text{ μm}$. This was done in order to improve slightly the fit of the model to the observed submillimetre number counts. In applying GRASIL to model the SEDs of a sample of nearby galaxies, Silva et al. (1998) found that a similar modification (to $\epsilon_{\nu} \propto \nu^{-1.6}$) seemed to be required in the case of Arp220 (the only ultraluminous starburst in their sample), in order to reproduce the observed submillimetre data for that galaxy. This modification in fact has little effect on the IR predictions presented in the present paper, but we retain it for consistency with Baugh et al. (2005).

2.2.3 Interface with GALFORM

For calculating the statistical properties of the galaxy population from the combined GALFORM + GRASIL model, we follow the same strategy as described in Granato et al. (2000). We first run the GALFORM code to generate a large catalogue of model galaxies at any redshift, and then run the GRASIL code on subsamples of these. For the quiescent galaxies, we select a subsample which has equal numbers of galaxies in equal logarithmic bins of stellar mass, while

for the bursting galaxies, we select a subsample with equal numbers of galaxies in equal logarithmic bins of burst mass. For the burst sample, we compute SEDs at several different representative stages in the burst evolution, while for the quiescent sample, we only compute SEDs at a single epoch. Using this sampling strategy, we obtain a good coverage of all the different masses, types and evolutionary stages of galaxies, while minimizing the computational cost of running the GRASIL code. The statistical properties of the galaxy population are then obtained by assigning the model galaxies appropriate weights depending on their predicted number density in a representative cosmological volume.

The outputs from the GALFORM galaxy formation model required by GRASIL to calculate the galaxy SEDs are: the combined star formation history and metallicity distribution for the disc and bulge, the radii of both components, and the total mass of dust. The dust mass is calculated from the mass and metallicity of the cold gas in the galaxy, assuming that the dust-to-gas ratio is proportional to the metallicity. Since the gas mass and metallicity both evolve, so does the dust mass, and this evolution is fully taken into account in GRASIL. For simplicity, we assume that the size distribution of the dust grains and PAH molecules does not evolve, apart from the normalization.

Once we have calculated the SEDs for the model galaxies, we compute luminosities in different observed bands (e.g. the optical B band or the *Spitzer* 24- μm band) by convolving the SED with the filter + detector response function for that band. For computing the predicted fluxes from galaxies in a fixed observer-frame band, we redshift the SED before doing the convolution.

The GRASIL code is quite CPU-intensive, requiring several minutes of CPU time per galaxy. Consequently, we are limited to running samples of a few thousand galaxies at each redshift. As a result, quantities such as LFs and redshift distributions still show some small amount of noise, rather than being completely smooth curves, as can be seen in many of the figures in this paper.

2.3 Choice of parameters in the GALFORM + GRASIL model

The combined GALFORM + GRASIL model has a significant number of parameters, but this is inevitable given the very wide range of physical processes which are included. The parameters are constrained by requiring the model predictions to reproduce a limited set of observational data – once this is done, there is rather little freedom in the choice of parameters. We have described above how the main parameters are fixed, and more details can be found in Cole et al. (2000) and Baugh et al. (2005). For both of these papers, large grids of \sc galform models were run with different parameters, in order to decide which set of parameters gave the best overall fit to the set of calibrating observational data. These papers also show the effects of varying some of the main model parameters around their best-fitting values. The parameters in the standard model for which we present results in this paper were chosen to reproduce the following properties for present-day galaxies: the LFs in the B and K bands and at 60 μm , the relations between gas mass and luminosity and metallicity and luminosity, the size-luminosity relation for galaxy discs, and the fraction of spheroidal galaxies. In addition, the model was required to reproduce the observed rest-frame far-UV (1500 \AA) LF at $z = 3$, and the observed submillimetre number counts and redshift distribution at 850 μm (Baugh et al. 2005). The submillimetre number counts are the main factor driving the need to include a top-heavy IMF in bursts.

The parameters for our standard model are exactly the same as in Baugh et al. (2005), which were chosen before *Spitzer* data be-

came available. Since these parameters were not adjusted to match any data obtained with *Spitzer*, the predictions of our model in the *Spitzer* bands are genuine predictions. We could obviously have fine-tuned our parameters in order to match better the observational data we considered in this paper, but this would have conflicted with our main goal, which is to present predictions for a wide set of observable properties based on a single physical model in a series of papers.

Since our assumption of a top-heavy IMF in bursts is a controversial one, we will also show some predictions from a variant model, which is identical to the standard model, except that we assume the same solar neighbourhood (Kennicutt) IMF in bursts and in discs. Comparing the predictions for the standard and variant models then shows directly the effects of changing the IMF in bursts. We note that the variant model matches the present-day optical and near-IR LFs almost as well as the standard model, though it is a poorer fit to the local 60- μm LF for the brightest galaxies (see Fig. 9). The variant model underpredicts the 850 μm counts by a factor of 10–30.

3 NUMBER COUNTS

We begin our comparison of the predictions of our galaxy formation model against *Spitzer* data with the galaxy number counts. Fig. 1 shows number counts in the four IRAC bands (3.6, 4.5, 5.8 and 8.0 μm), and Fig. 2 does the same for the three MIPS bands (24, 70 and 160 μm). Each panel is split in two: the upper subpanel plots the counts per logarithmic flux interval, $dN/d \ln S_\nu$, while the lower subpanel instead plots $S_\nu dN/d \ln S_\nu$. The latter is designed to take out much of the trend with flux, in order to show more clearly the differences between the model and the observational data. In each case we plot three curves for our standard model: the solid blue line shows the total number counts including both extinction and emission by dust, the solid red line shows the contribution to this from galaxies currently forming stars in a burst, and the solid green line shows the contribution from all other galaxies (star-forming or not), which we denote as ‘quiescent’. In Fig. 1, we also plot a dashed blue line which shows the predicted total counts if we ignore absorption and emission from interstellar dust (emission from dust in the envelopes of asymptotic giant branch (AGB) stars is still included in the stellar contribution, however). In the MIPS bands, the predicted counts are negligible in the absence of interstellar dust, so we do not plot them in Fig. 2. In the lower subpanels, we also show by a dashed magenta line the prediction from a variant model which assumes a normal (Kennicutt) IMF for all star formation, but is otherwise identical to our standard model (which has a top-heavy IMF in bursts). This variant model fits the local B and K bands and 60- μm LFs about as well as our standard model, but dramatically underpredicts the 850 μm number counts. The observed number counts are shown by black symbols with error bars.

Overall, the agreement between the predictions of our standard model and the observed counts is remarkably good, when one takes account of the fact that no parameters of the model were adjusted to improve the fit to any data from *Spitzer*. Consider first the results for the IRAC bands, shown in Fig. 1. Here, the agreement of the model with observations seems best at 3.6 and 8.0 μm , and somewhat poorer at 5.8 μm . The model predicts somewhat too few objects at fainter fluxes in all of the IRAC bands. Comparing the red and green curves, we see that quiescent galaxies rather than bursts dominate the counts at all observed fluxes in all of the IRAC bands, but especially at the shorter wavelengths, consistent with the expectation that at 3.6 and 4.5 μm , we are seeing mostly light from old stellar populations. Comparing the solid and dashed blue

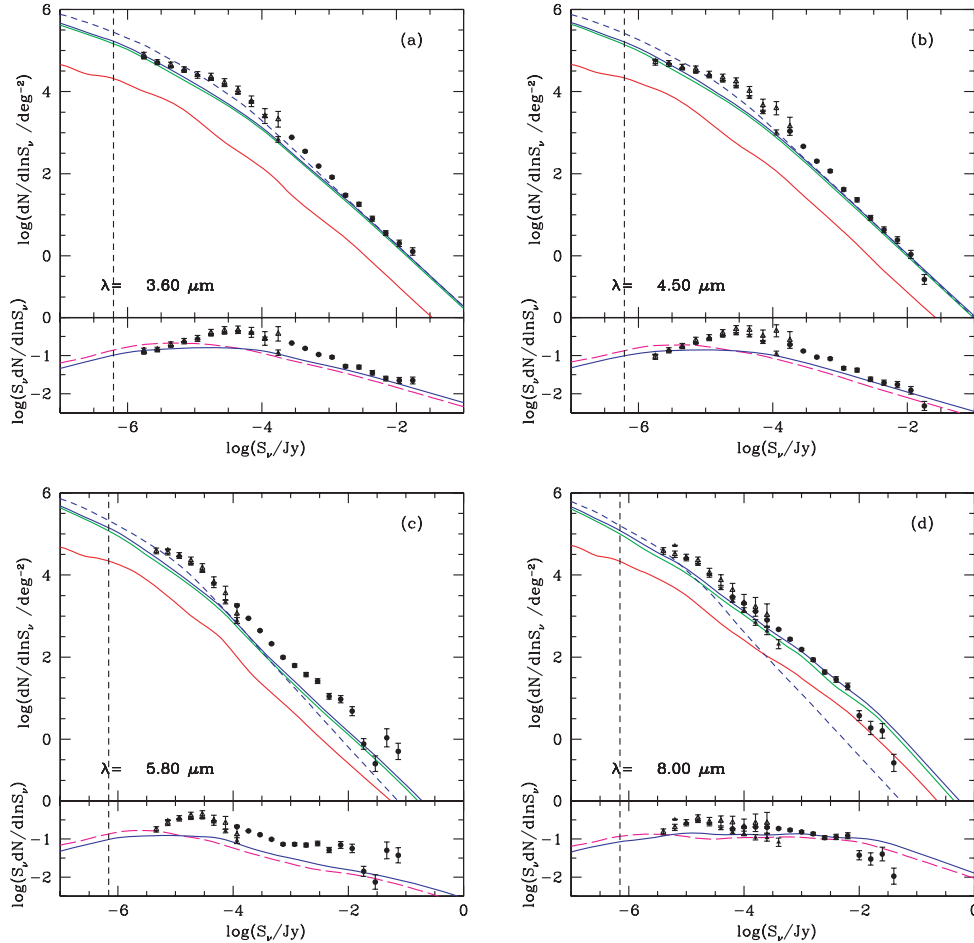


Figure 1. Galaxy differential number counts in the four IRAC bands. The curves show model predictions, while the symbols with error bars show observational data from Fazio et al. (2004a) (with different symbols for data from different survey fields). Each panel is split in two: the upper subpanel plots the counts as $dN/d \ln S_v$ versus S_v , while the lower subpanel plots $S_v dN/d \ln S_v$ (in units mJy deg^{-2}) on the same horizontal scale. The upper subpanels show four different curves for our standard model – solid blue: total counts including dust extinction and emission; dashed blue: total counts excluding interstellar dust; solid red: ongoing bursts (including dust); solid green: quiescent galaxies (including dust). The lower subpanels compare the total counts including dust for the standard model (solid blue line) with those for a variant model with a normal IMF for all stars (dashed magenta line). The vertical dashed line shows the estimated confusion limit for the model. (a) 3.6 μm ; (b) 4.5 μm ; (c) 5.8 μm ; (d) 8.0 μm .

lines, we see that the effects of dust are small at 3.6 and 4.5 μm , with a small amount of extinction at faint fluxes (and thus higher average redshifts), but negligible extinction for brighter fluxes (and thus lower redshifts). On the other hand, dust has large effects at 8.0 μm , with dust emission (due to strong PAH features at $\lambda \sim 6\text{--}9 \mu\text{m}$) becoming very important at bright fluxes (which correspond to low average redshifts – see Fig. A1b in Appendix A). The 8.0 μm counts thus are predicted to be dominated by dust emission from quiescently star-forming galaxies, except at the faintest fluxes. The counts at 5.8 μm show behaviour which is intermediate, with mild emission effects at bright fluxes and mild extinction at faint fluxes. Comparing the solid blue and dashed magenta lines, we see that the predicted number counts in the IRAC bands are almost the same whether or not we assume a top-heavy IMF in bursts, consistent with the counts being dominated by quiescent galaxies.

Consider next the results for the MIPS bands, shown in Fig. 2. We again see remarkably good agreement of the standard model with the observational data. The agreement is especially good at faint fluxes (corresponding to higher redshifts). In particular, the model matches well the observed 24 μm counts at the ‘bump’ around fluxes $S_v \sim 0.1\text{--}1 \text{ mJy}$. Accurate modelling of the PAH

emission features is obviously crucial for modelling the 24 μm number counts, since the PAH features dominate the flux in the 24- μm band as they are redshifted into the band at $z \gtrsim 0.5$. On the other hand, the standard model overpredicts the number counts at bright fluxes (corresponding to low redshifts) in all three MIPS bands. The evolution at these wavelengths predicted by our ΛCDM -based model thus seems to be not quite as strong as indicated by observations.

In the MIPS bands, emission from galaxies is completely dominated by dust, which is why no dashed blue lines are shown in Fig. 2. Comparing the red and green curves, we see that quiescent (but star-forming) galaxies tend to dominate the number counts in these bands at brighter fluxes, and bursts at fainter fluxes. This reflects the increasing dominance of bursts in the mid- and far-IR LF at higher redshifts. Comparing the solid blue and dashed magenta curves, we see that our standard model with a top-heavy IMF in bursts provides a significantly better overall fit to the observed 24 μm counts than the variant model with a normal IMF in bursts (although at the brightest fluxes, the variant model fits better). The faint number counts at 70 μm also favour the top-heavy IMF model, while the number counts at 160 μm cover a smaller flux range, and

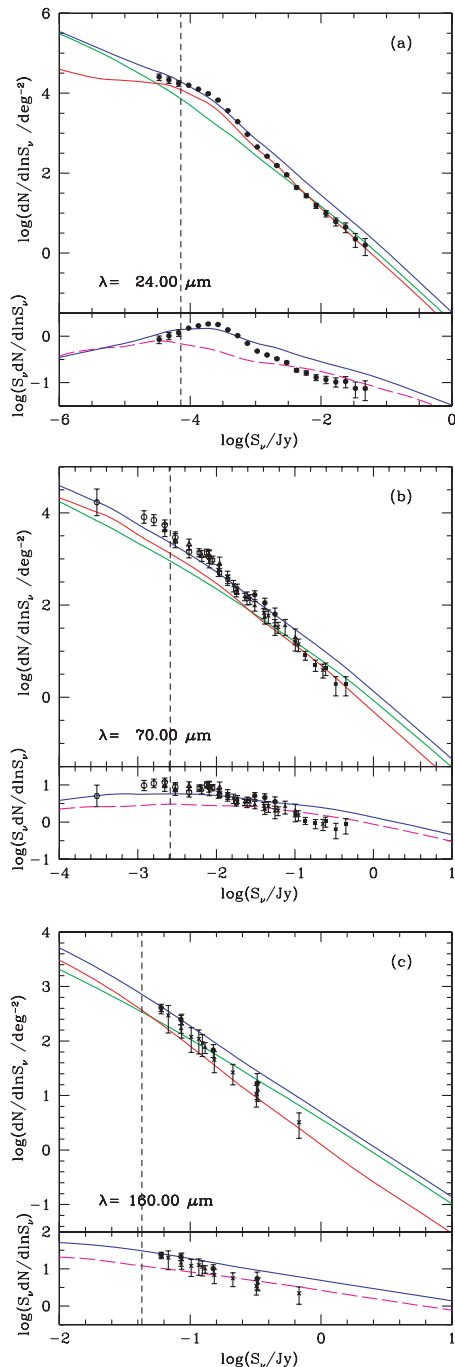


Figure 2. Galaxy differential number counts in the three MIPS bands. The curves show model predictions while the symbols with error bars show observational data. The meaning of the different model lines is the same as in Fig. 1. (a) 24 μm , with observational data from Papovich et al. (2004). (b) 70 μm , with observational data from Dole et al. (2004a) (filled symbols), Frayer et al. (2006a) (crosses) and Frayer et al. (2006b) (open symbols). (c) 160 μm (bottom panel), with observational data from Dole et al. (2004a) (filled symbols) and Frayer et al. (2006a) (crosses).

do not usefully distinguish between the two variants of our model with different burst IMFs.

We can use our model to predict the flux levels at which sources should become confused in the different *Spitzer* bands. We estimate the confusion limit using the *source density criterion* (e.g. Vaisanen, Tollestrup & Fazio 2001; Dole, Lagache & Puget 2003): if the

telescope has a full width at half-maximum (FWHM) beamwidth of θ_{FWHM} , we define the effective beam solid angle as $\omega_{\text{beam}} = [\pi/(4\ln 2)] \theta_{\text{FWHM}}^2 = 1.13\theta_{\text{FWHM}}^2$, and then define the confusion limited flux S_{conf} to be such that $N(>S_{\text{conf}}) = 1/(N_{\text{beam}}\omega_{\text{beam}})$, where $N(>S)$ is the number per solid angle of sources brighter than flux S . We choose $N_{\text{beam}} = 20$ for the number of beams per source, which gives similar results to more detailed analyses (e.g. Vaisanen et al. 2001; Dole et al. 2004b). We use values of the beam size $\theta_{\text{FWHM}} = (1.66, 1.72, 1.88, 1.98)$ arcsec for the four IRAC bands (Fazio et al. 2004b) and $(5.6, 16.7, 35.2)$ arcsec for the three MIPS bands (Dole et al. 2003). Our standard model then predicts confusion-limited fluxes of $S_{\text{conf}} = (0.62, 0.62, 0.69, 0.70)$ μJy in the $(3.6, 4.5, 5.8, 8.0)$ μm IRAC bands, and $S_{\text{conf}} = (0.072, 2.6, 43)$ mJy in the $(24, 70, 160)$ μm MIPS bands. These confusion estimates for the MIPS bands are similar to those of Dole et al. (2004b), which were based on extrapolating from the observed counts. These values for the confusion limits are indicated in Figs 1 and 2 by vertical dashed lines.

Our galaxy evolution model does not compute the contribution of AGN to the IR luminosities of galaxies. On the other hand, the observed number counts to which we compare include both normal galaxies, in which the IR emission is powered by stellar populations, and AGN, in which there is also IR emission from a dust torus, which is expected to be most prominent in the mid-IR. However, multi-wavelength studies using optical, IR and X-ray data indicate that even at 24 μm , the fraction of sources dominated at that wavelength by AGN is only 10–20 per cent (e.g. Franceschini et al. 2005), and the contribution of AGN-dominated sources in the other *Spitzer* bands is likely to be smaller. Therefore we should not make any serious error by comparing our model predictions directly with the total number counts, as we have done here.

4 EVOLUTION OF THE GALAXY LUMINOSITY FUNCTION

While galaxy number counts provide interesting constraints on theoretical models, it is more physically revealing to compare with galaxy LFs, since these isolate behaviour at particular redshifts, luminosities and rest-frame wavelengths. In the following subsections, we compare our model predictions with recent estimates of LF evolution based on *Spitzer* data.

4.1 Evolution of the galaxy luminosity function at 3–8 μm

We consider first the evolution of the LF in the wavelength range covered by the IRAC bands, i.e. 3.6–8.0 μm . Fig. 3 shows what our standard model with a top-heavy IMF in bursts predicts for LF evolution at rest-frame wavelengths of 3.6 and 8.0 μm for redshifts $z = 0$ –3.³ We see that at a rest-frame wavelength of 3.6 μm , the model LF hardly evolves at all over the whole redshift range $z = 0$ –3. This lack of evolution appears to be somewhat fortuitous. Galaxy luminosities at a rest-frame wavelength of 3.6 μm are dominated by the emission from moderately old stars, but the stellar mass function in the model evolves quite strongly over the range $z = 0$ –3 (as we show in Section 5). The weak evolution in the 3.6- μm LF results from a cancellation between a declining luminosity-to-stellar-mass ratio with increasing time and increasing stellar masses (see Figs 13a and e). On the other hand, at a rest-frame wavelength

³ In this figure, and in Figs 4, 5, 8 and 10, the luminosities L_v are calculated through the corresponding *Spitzer* passbands.

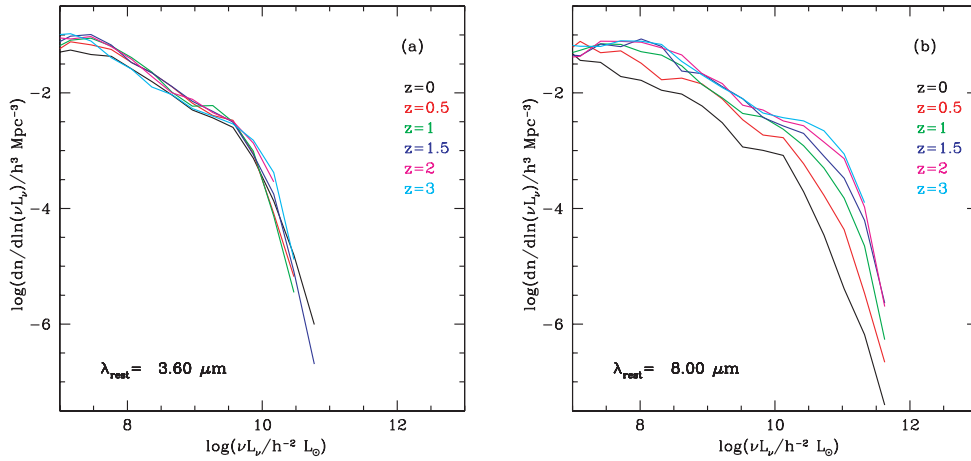


Figure 3. Predicted evolution of the galaxy LF in our standard model (including dust) at rest-frame wavelengths of (a) 3.6 and (b) 8.0 μm for redshifts $z = 0, 0.5, 1, 1.5, 2$ and 3 , as shown in the key.

of 8.0 μm, the model LF becomes significantly brighter in going from $z = 0$ to 3 . Galaxy luminosities at a rest-frame wavelength of 8.0 μm are dominated by emission from dust heated by young stars, so this evolution reflects the increase in star formation activity with increasing redshift (see Fig. 13b in Section 5).

In Fig. 4, we compare the model predictions for evolution of the LF at 3.6 μm with observational estimates from Babbedge et al. (2006) and Franceschini et al. (2006).⁴ The model predictions are given for redshifts $z = 0, 0.5$ and 1 . For the observational data, the mean redshifts for the different redshift bins used do not exactly coincide with the model redshifts, so we plot them with the model output closest in redshift.⁵ The observational estimates of the 3.6-μm LF rely on the measured redshifts. In the case of Babbedge et al. (2006), these are mostly photometric, using optical and near-IR (including 3.6 and 4.5 μm) fluxes, while for the Franceschini et al. sample, about 50 per cent of the redshifts are spectroscopic and the remainder photometric. In both samples, the measured 3.6-μm fluxes were k -corrected to estimate the rest-frame 3.6-μm luminosities.

We see from comparing the blue curve with the observational data in Fig. 4 that the 3.6-μm LF predicted by our standard model is in very good agreement with the observations. In particular, the observational data show very little evolution in the 3.6-μm LF over the redshift range $z = 0$ – 1 . The largest difference seen is at $z = 1$, where the Babbedge et al. data show a tail of objects to very high luminosities, which is not seen in the model predictions. However, this tail is not seen in the Franceschini et al. data at the same redshift, and is also not present in the observational data at the lower redshifts. More spectroscopic redshifts are needed for the Babbedge et al. sample to clarify whether this high-luminosity tail is real. Comparing the red, green and blue lines for the standard model shows that the model LF is dominated by quiescent galaxies at low luminosity, but the contribution of bursts becomes comparable to that of

quiescent galaxies at high luminosities. We have not shown model LFs excluding dust extinction in this figure, since they are almost identical to the predictions including dust. The dashed magenta lines show the predicted LFs for the variant model with a normal IMF in bursts. We see that these differ only slightly from our standard model, but are a somewhat poorer fit to the observational data at higher luminosities.

In Fig. 5 we show a similar comparison for the LF evolution at a rest-frame wavelength of 8 μm. The model predictions are given for redshifts $z = 0, 1$ and 2 , and are compared with observational estimates by Huang et al. (2007) (for $z \sim 0$), Babbedge et al. (2006) (for $z \sim 0$ and 1) and Caputi et al. (2007) (for $z \sim 1$ and 2). These papers all classified objects in their samples as either galaxies or AGN, and then computed separate LFs for the two types of objects.⁶ Our model does not make any predictions for AGN, so we compare our model predictions with the observed LFs for objects classified as galaxies only. We see that for redshifts around $z = 1$, the observed LFs from Babbedge et al. and Caputi et al. are in very poor agreement with each other, with the Caputi et al. LF being around 10 times higher in number density at the same luminosity. This difference presumably results from some combination of (i) different methods of classifying objects as galaxies or AGN (Babbedge et al. used only optical and IR fluxes to do this, while Caputi et al. also used X-ray data); (ii) different photometric redshift estimators and (iii) different methods for k -correcting luminosities to a rest-frame wavelength of 8 μm. There are smaller differences between the Huang et al. and Babbedge et al. LFs at $z \sim 0$. Further observational investigation appears to be necessary to resolve these issues. Our standard model is in reasonable agreement with the Babbedge et al. observed LF at $z \sim 0$, and with the Caputi et al. observed LFs at $z \sim 1$ and 2 , but not with the Babbedge et al. observed LF at $z \sim 1$. The comparison with Caputi et al. favours our standard model with a top-heavy IMF in starbursts over the variant model with a normal IMF.

⁴ Babbedge et al. (2006) also compared their measured LFs at 3.6, 8.0 and 24 μm with predictions from a preliminary version of the model described in this paper.

⁵ Specifically, for $z = 0$, we compare with the $z = 0.1$ data from Babbedge et al., for $z = 0.5$ we compare with the $z = 0.5$ data from Babbedge et al. and $z = 0.3$ data from Franceschini et al., and for $z = 1$, we compare with the $z = 0.75$ (open symbols) and $z = 1.25$ (filled symbols) data from Babbedge et al. and $z = 1.15$ data from Franceschini et al.

⁶ Note that a variety of criteria have been used for classifying observed IR sources as AGN or normal galaxies, and these do not all give equivalent results. Even if an object is classified as an AGN, it is also not clear that in all cases the AGN luminosity dominates over that of the host galaxy in all *Spitzer* bands.

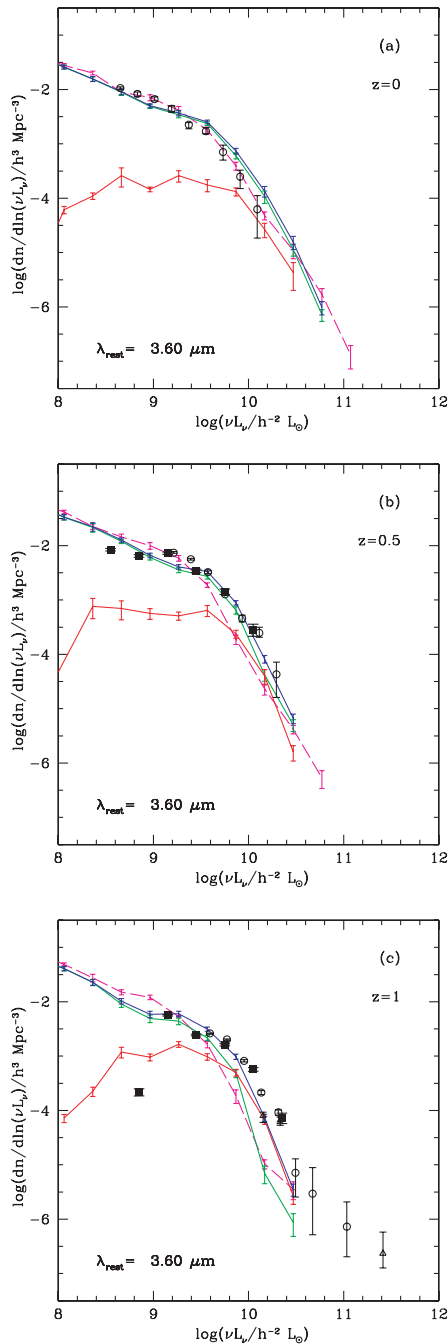


Figure 4. Predicted evolution of the galaxy LF at rest-frame 3.6 μm compared to observational data. The different panels show redshifts (a) $z = 0$, (b) $z = 0.5$ and (c) $z = 1$. The predictions for our standard model are shown by the blue line, with the red and green lines showing the separate contributions from ongoing bursts and quiescent galaxies. The dashed magenta line shows the prediction for a variant model with a normal IMF for all stars. The error bars on the model lines indicate the Poisson uncertainties due to the finite number of galaxies simulated. The black symbols with error bars show observational data from Babbedge et al. (2006) (open circles and triangles, for $z = 0, 0.5$ and 1) and Franceschini et al. (2006) (filled squares, for $z = 0.5$ and 1).

4.2 Evolution of the galaxy luminosity function at 12–24 μm

In this subsection, we consider the evolution of the galaxy LF at mid-IR wavelengths, and compare with data obtained using mainly the MIPS 24- μm band.

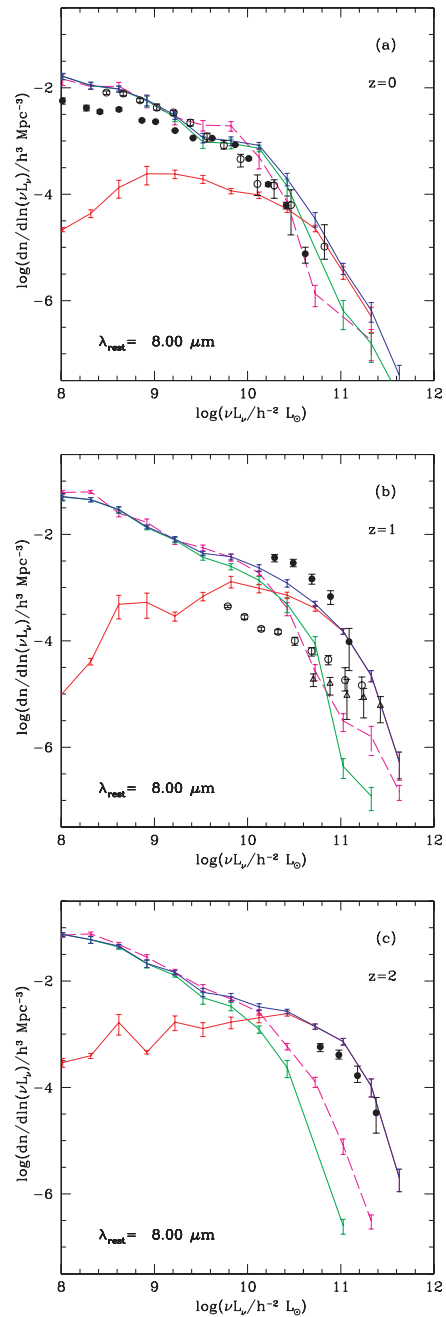


Figure 5. Predicted evolution of the galaxy LF at rest-frame 8.0 μm compared to observational data. The different panels show redshifts (a) $z = 0$, (b) $z = 1$ and (c) $z = 2$. The coloured lines showing the model predictions have the same meaning as in Fig. 4. The black symbols with error bars show observational data from Babbedge et al. (2006) (open circles for $z = 0$ and 0.7, triangles for $z = 1.2$), Huang et al. (2007) (filled circles for $z = 0$) and Caputi et al. (2007) (filled circles for $z = 1$ and 2). The observed LFs are for normal galaxies and exclude AGN.

Fig. 6 shows what our standard model with a top-heavy IMF in bursts predicts for the evolution of the galaxy LF at rest-frame wavelengths of 15 and 24 μm for redshifts $z = 0$ –3.⁷ At rest-frame

⁷ In this figure, and in Figs 7 and 8, the 24- μm luminosities are calculated through the corresponding MIPS passband, while the 15- μm luminosities are calculated through a top-hat filter with a fractional width of 10 per cent in wavelength.

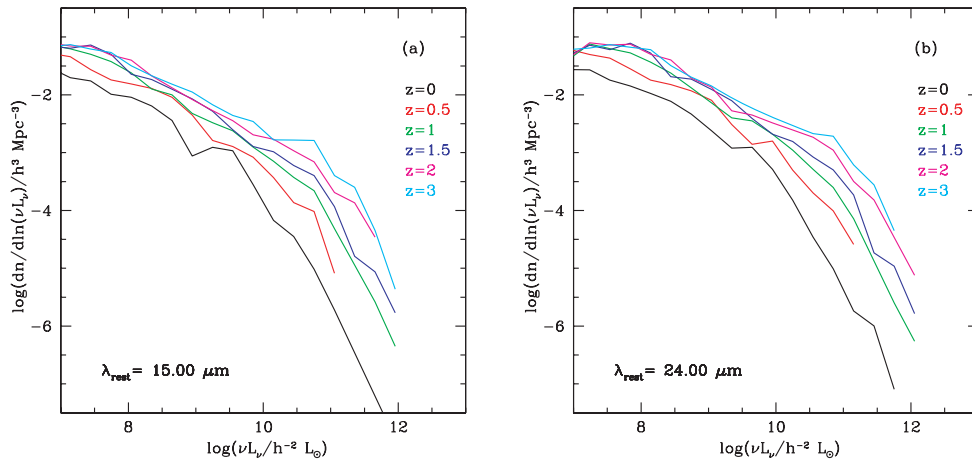


Figure 6. Predicted evolution of the galaxy LF in our standard model at rest-frame wavelengths (a) 15 μm (left-hand panel) and (b) 24 μm (right-hand panel) for redshifts $z = 0, 0.5, 1, 1.5, 2$ and 3 , as shown in the key.

wavelengths of 15 and 24 μm , galaxy luminosities are typically dominated by the continuum emission from warm dust grains heated by young stars (although PAH emission is also significant at some nearby wavelengths). Fig. 6 shows strong evolution in the model LFs over the redshift range $z = 0$ – 3 at both wavelengths, reflecting both the increase in star formation activity with increasing redshift (see Fig. 13b) and the increasing dominance of the burst mode of star formation, for which the top-heavy IMF further boosts the mid- and far-IR luminosities compared to a normal IMF. Comparing Fig. 6 with Fig. 3(a), we also see a difference in the shape of the bright end of the LF: at 3.6 μm , where the LF is dominated by emission from stars, the bright end cuts off roughly exponentially, while at 15 and 24 μm , where the LF is dominated by emission from warm dust, the bright end declines more gradually, roughly as a power law. This difference reflects the difference in shape of the GSMF and GSFRD. The GSMF shows an exponential-like cut-off at high masses, while the GSFRD shows a more gradual cut-off at high SFRs because of starbursts triggered by galaxy mergers (see Figs 13a and b in Section 5). This difference was noticed earlier by observers comparing optical and far-IR LFs of galaxies, but its origin was not understood (Lawrence et al. 1986; Soifer et al. 1987b).

In Fig. 7, we compare the model LFs at rest-frame wavelengths 12 and 15 μm with observational estimates. For $z = 0$, we plot the observational estimates from Soifer & Neugebauer (1991) and Rush, Malkan & Spinoglio (1993), based on *IRAS* 12- μm data (with AGN removed). For $z = 0.5$ – 1 and 1.5 – 2.5 , we plot the data of Le Floc’h et al. (2005) and Perez-Gonzalez et al. (2005), respectively, which were obtained from galaxy samples selected on *Spitzer* 24- μm flux. Le Floc’h et al. k -corrected their measured 24- μm fluxes to 15- μm rest-frame luminosities, while Perez-Gonzalez et al. k -corrected to 12- μm rest-frame.⁸ Le Floc’h et al. obtained most of their redshifts from photometric redshifts based on optical data, while Perez-Gonzalez et al. used a new photometric redshift technique based on fitting empirical SEDs to all of the available broadband data from the far-UV to 24 μm , and also removed ‘extreme’ AGN from their observed LF. Note that the redshifts for the observed

⁸ The exact passband used for the model LF in each panel depends on which observational data we are comparing with. For $z = 0$, we use the *IRAS* 12 μm passband; at $z = 0.5$ and 1 we use a top-hat passband centred at 15 μm and at $z = 1.5, 2$ and 2.5 , we use a top-hat passband centred at 12 μm (both top-hat passbands having fractional width 10 per cent in wavelength).

LFs do not exactly coincide with model redshifts in all cases, but are close.

We see from comparing the blue line to the observational data points in Fig. 7 that our standard model with a top-heavy IMF in bursts fits the observations remarkably well up to $z = 2$. In particular, the model matches the strong evolution in the mid-IR LF seen in the observational data. The model falls below the observational data at $z = 2.5$, but here both the photometric redshifts and the k -corrections are probably the most uncertain. The standard model also does not provide a perfect fit to the $z = 0$ data, predicting somewhat too many very bright galaxies and somewhat too few very faint galaxies (though the latter discrepancy might be affected by local galaxy clustering in the *IRAS* data). Comparing the red, green and blue lines for the standard model in the figure, we see that the bright end of the 12- or 15- μm LF is dominated by bursts at all redshifts. The figure also shows by a dashed magenta line the predictions for the variant model with a normal IMF in bursts. This latter model predicts much less evolution in the bright end of the LF than is observed. This comparison thus strongly favours the model with the top-heavy IMF in bursts.

Finally, in Fig. 8, we carry out a similar comparison of the evolution of predicted and observed LFs at a rest-frame wavelength of 24 μm over the redshift range $z = 0$ – 1 , in this case comparing with observational estimates from Shupe et al. (1998) (for $z = 0$), based on *IRAS* data, and from Babbedge et al. (2006) (for $z = 0$ – 1), based on *Spitzer* data.⁹ The galaxy redshifts for the Babbedge et al. data were obtained in the same way as for the 3.6- μm LFs shown in Fig. 4, and the luminosities were k -corrected from observer-frame 24 μm to rest-frame 24 μm . The LF plotted from Babbedge et al. is that for normal galaxies, with AGN excluded.

The conclusions from comparing the model with the 24- μm LFs are similar to those from the comparison with the 12- and 15- μm LFs. The data favour our standard model over the variant with a normal IMF in bursts (except possibly for $z = 0.5$), as the latter predicts too little evolution at the bright end. At $z = 0$, the model fits the 24- μm data rather better than for the corresponding comparison at 12 μm . On the other hand, at $z = 0.5$ and 1 , the model LF is a

⁹ The model luminosities are all computed through the *Spitzer* 24 μm passband.

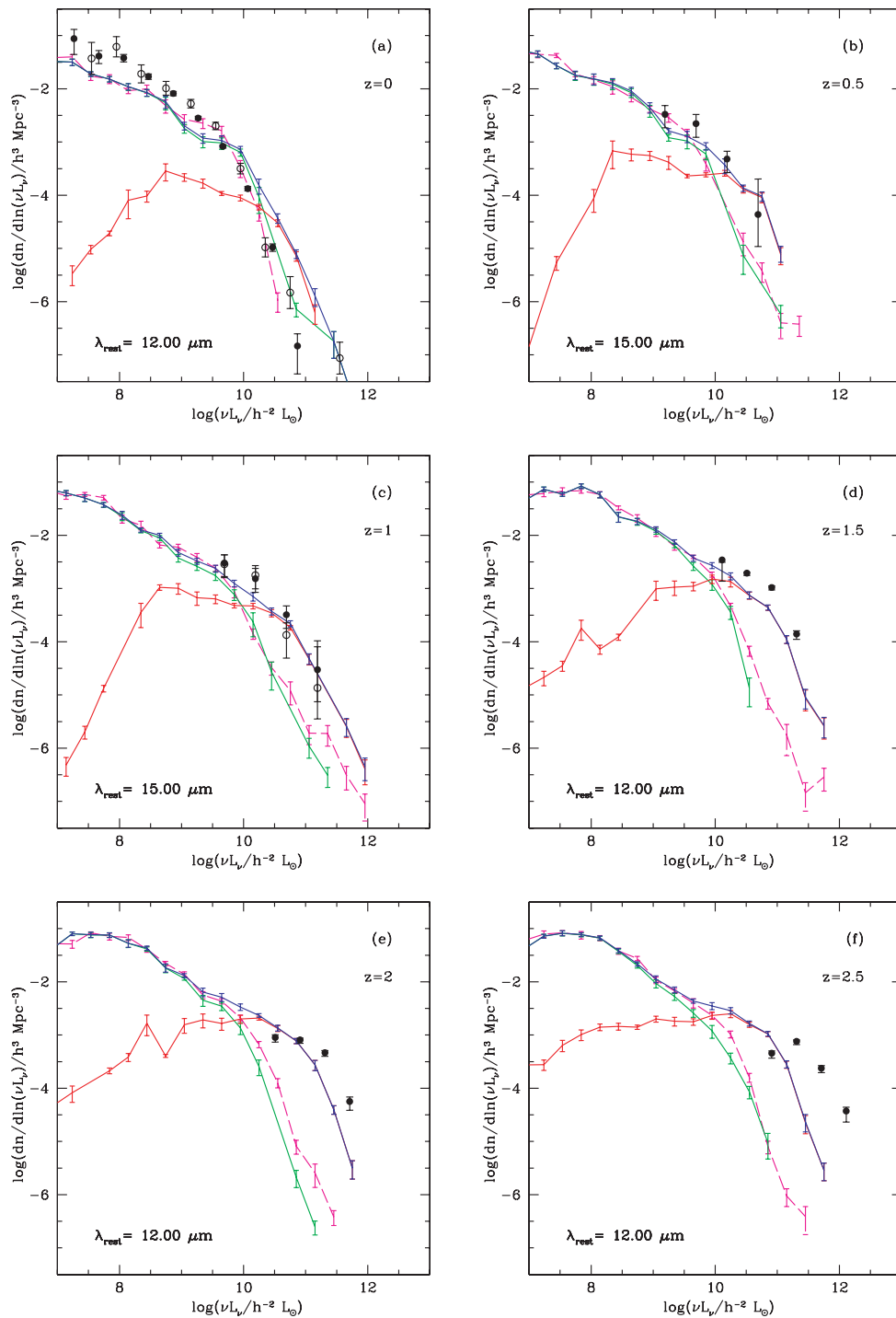


Figure 7. Predicted evolution of the galaxy LF at rest-frame wavelength 12 or 15 μm compared to observational data. The different panels show redshifts: (a) $z = 0$, (b) $z = 0.5$, (c) $z = 1$, (d) $z = 1.5$, (e) $z = 2$ and (f) $z = 2.5$. The meaning of the curves showing the model predictions is the same as in Fig. 4. In panel (a), the predictions at 12 μm are compared to observational determinations from Soifer & Neugebauer (1991) (open symbols) and Rush et al. (1993) (filled symbols) based on *IRAS* data. In panels (b) and (c), the predictions at 15 μm are compared to observational data from Le Floch et al. (2005). In panels (d), (e) and (f), the predictions at 12 μm are compared to observational data from Perez-Gonzalez et al. (2005).

somewhat worse fit to the observational data at 24 μm than at 15 μm . These differences between the 12/15 and 24 μm comparisons might result from the different photometric redshifts and k -corrections used in the observational samples in the two cases. Alternatively, they might result from problems in modelling the dust SEDs in the complex mid-IR range.

4.3 Evolution of the galaxy luminosity function at 70–160 μm

We now briefly consider the evolution of the LF in the far-IR. The far-IR is the wavelength range where most of the luminosity from dust in normal galaxies is emitted. The local 60- μm LF was very well measured by surveys with *IRAS*, and so is commonly used

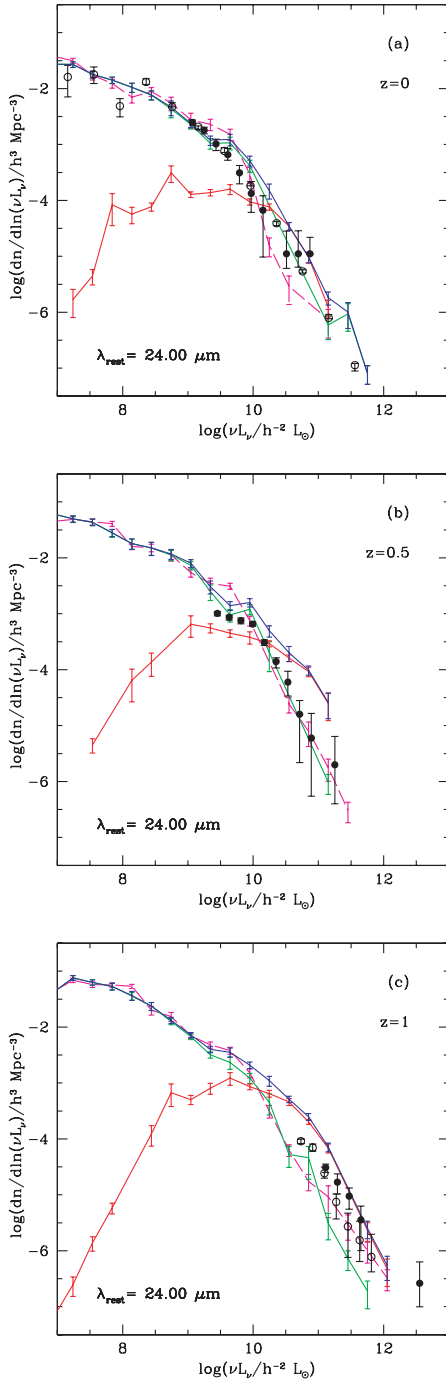


Figure 8. Predicted evolution of the galaxy LF at rest-frame wavelength $24\ \mu\text{m}$ compared to observational data from Shupe et al. (1998) (at $z = 0$, open symbols) and from Babbedge et al. (2006) (for the same redshifts as in Fig. 4). The meaning of the curves showing the model predictions is the same as in Fig. 4. (a) $z = 0$, (b) $z = 0.5$ and (c) $z = 1$.

as a starting point or benchmark for modelling the evolution of the galaxy population in the far-IR. We therefore present in Fig. 9 the model prediction for the $60\text{-}\mu\text{m}$ LF at $z = 0$, compared with observational data from Saunders et al. (1990), Soifer & Neugebauer (1991) and Takeuchi, Yoshikawa & Ishii (2003). As discussed in Baugh et al. (2005), the local $60\text{-}\mu\text{m}$ LF was used as one of the primary constraints in fixing the parameters of our galaxy formation model, and the figure shows that our standard model provides a good

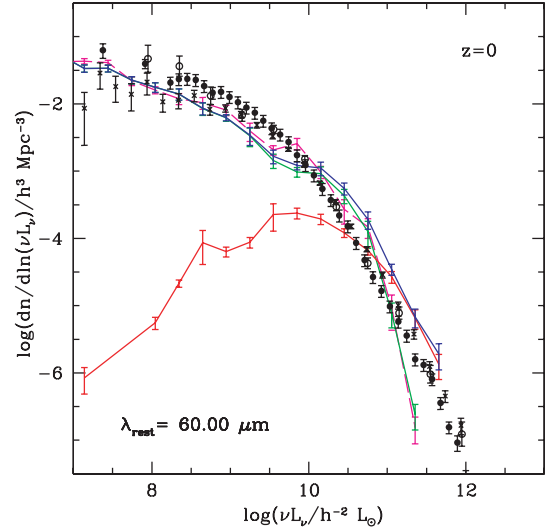


Figure 9. The predicted galaxy LF at $60\ \mu\text{m}$ compared to observational data from *IRAS*. The meaning of the different lines is the same as in Fig. 4. The black symbols show observational data from Saunders et al. (1990) (crosses), Soifer & Neugebauer (1991) (open circles), and Takeuchi et al. (2003) (filled circles).

match to the data. The variant model with a normal IMF in bursts underpredicts the abundance of the brightest $60\text{-}\mu\text{m}$ galaxies.

In Fig. 10, we show the model predictions for the evolution of the LF in the two longer wavelength MIPS bands, at rest-frame wavelengths of 70 and $160\ \mu\text{m}$, from $z = 0$ to 3 . At $70\ \mu\text{m}$, the LF at high luminosities is predicted to brighten by about a factor of 10 going from $z = 0$ to 2 . This is about a factor of 2 less than the brightening predicted in the mid-IR at $15\ \mu\text{m}$ (compare to Fig. 6), but nearly a factor of 2 more evolution than is predicted at $160\ \mu\text{m}$. These differences between the amount of evolution seen at different IR wavelengths reflect evolution in the shapes of the SEDs of the galaxies responsible for the bulk of the IR emission. No observational estimates of the evolution of the LF at 70 and $160\ \mu\text{m}$ have yet been published, but they are expected to be forthcoming from ongoing surveys with *Spitzer*.

4.4 Evolution of the total mid- + far-IR luminosity function

The total mid- + far-IR luminosity of a galaxy, L_{IR} , integrated over the whole wavelength range $8\text{--}1000\ \mu\text{m}$, is a very good approximation to the total luminosity emitted by interstellar dust grains in all galaxies except those with very small dust contents. In galaxies with significant star formation, L_{IR} is mostly powered by dust heated by young stars, and so provides a quantitative indicator of the amount of dust-obscured star formation which is independent of the shape of the IR SED (though still subject to uncertainties about the IMF). The evolution of the LF in L_{IR} is therefore a very interesting quantity to compare between models and observations. We show in Fig. 11 what our standard model predicts for the evolution of the IR LF over the range $z = 0\text{--}6$. We see that the model predicts substantial evolution in this LF, with the high-luminosity end brightening by a factor of ~ 10 from $z = 0$ to 2 , followed by a ‘plateau’ from $z = 2$ to 4 , and a decline from $z = 4$ to 6 .

In Fig. 12, we compare our model predictions with existing observational estimates of the total IR LF for $z = 0\text{--}2$. These observational estimates are only robust for $z = 0$, where they are based on *IRAS* measurements covering the wavelength range $12\text{--}100\ \mu\text{m}$. At all of

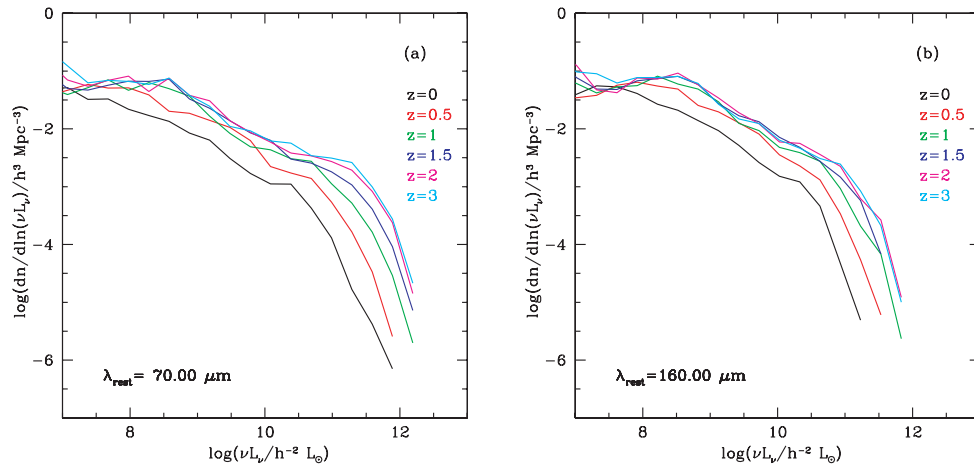


Figure 10. Predicted evolution of the galaxy LF in our standard model (including dust) at rest-frame wavelengths (a) 70 μm and (b) 160 μm , for redshifts $z = 0, 0.5, 1, 1.5, 2$ and 3 , as shown in the key.

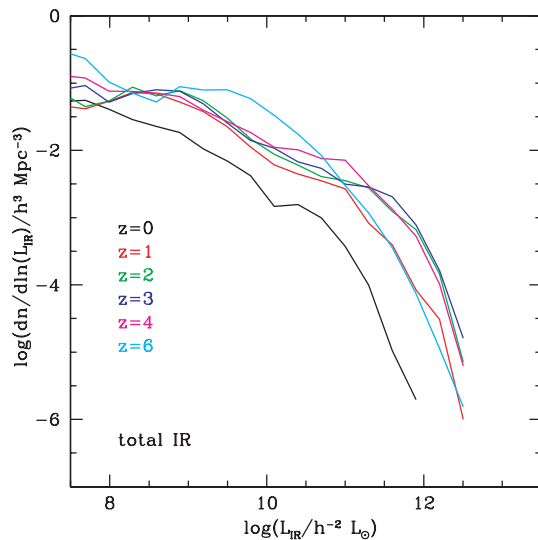


Figure 11. Predicted evolution of the total mid- + far-IR (8–1000 μm) galaxy LF for our standard model, for redshifts $z = 0, 1, 2, 3, 4$ and 6 , as shown in the key.

the higher redshifts plotted, the observational estimates are based on measurements of the mid-IR luminosity derived from *Spitzer* 24- μm fluxes, converted to total IR luminosities by assuming SED shapes for the mid- to far-IR emission. The bolometric correction from the observed mid-IR luminosity to the inferred total IR luminosity is typically a factor of ~ 10 , and is significantly uncertain. Therefore, the most robust way to compare the models with the observations is to compare them at the mid-IR wavelengths where the measurements are actually made, as we have done in Sections 4.1 and 4.2. None the less, if we take the observational determinations at face value, then we see that observed evolution of the total IR LF agrees remarkably well with the predictions of our standard model with a top-heavy IMF. On the other hand, the variant model with a normal IMF predicts far too few high L_{IR} galaxies at higher z , and is strongly disfavoured by the existing data.

5 INFERRING STELLAR MASSES AND SFRS FROM *Spitzer* DATA

In this section, we consider what the models imply about how well we can infer the stellar masses and SFRs in galaxies from measurements of rest-frame IR luminosities. The top two panels of Fig. 13 show the predicted GSMF (left-hand panel) and GSFRD (right-hand panel), for redshifts $z = 0$ – 6 . We see that the predicted stellar mass function shows dramatic evolution over this redshift range, with a monotonic decline in the number of high-mass galaxies with increasing redshift. On the other hand, the SFR distribution shows much less dramatic evolution over this redshift range, with a mild increase in the number of high-SFR objects up to $z \sim 3$, followed by a decline above that. The lower four panels in Fig. 13 show the relation in the models between stellar masses and SFRs and rest-frame luminosities at different IR wavelengths. (Note that in all cases, luminosities are measured in units of the *bolometric* solar luminosity.) The middle and bottom left-hand panels, respectively, show the mean ratio of luminosity in the rest-frame K (2.2 μm) or 3.6- μm bands to stellar mass as a function of stellar mass. The middle and bottom right-hand panels, respectively, show the mean ratio of total mid- + far-IR (8–1000 μm) or rest-frame 15- μm luminosity to SFR as a function of SFR. (The mean L/M_* or L/SFR ratios plotted are computed by dividing the total luminosity by the total mass or SFR, in each bin of mass or SFR.)

The near-IR luminosity is often used as a tracer of stellar mass. The left-hand panels of Fig. 13 show that the L/M_* ratio varies strongly with redshift, reflecting the difference in the ages of the stellar populations. At higher redshifts it also shows a significant dependence on stellar mass, presumably reflecting a trend of age with mass. However, the variation of mean L/M_* with redshift is seen to be much smaller at 2.2 μm than at 3.6 μm , implying that the rest-frame K -band light should provide a more robust estimator of stellar mass than the light at longer wavelengths. The differences between L/M_* values at 2.2 and 3.6 μm reflect the larger contribution from AGB compared to red giant branch (RGB) stars at the longer wavelength. AGB stars have higher masses and younger ages than RGB stars, and so are more sensitive to star formation at recent epochs. The scatter in L/M_* at a given mass is also found in the models to increase with redshift. In the K band, it increases from ~ 40 per cent at $z \sim 0$ to a factor of ~ 3 at $z \sim 6$. The large

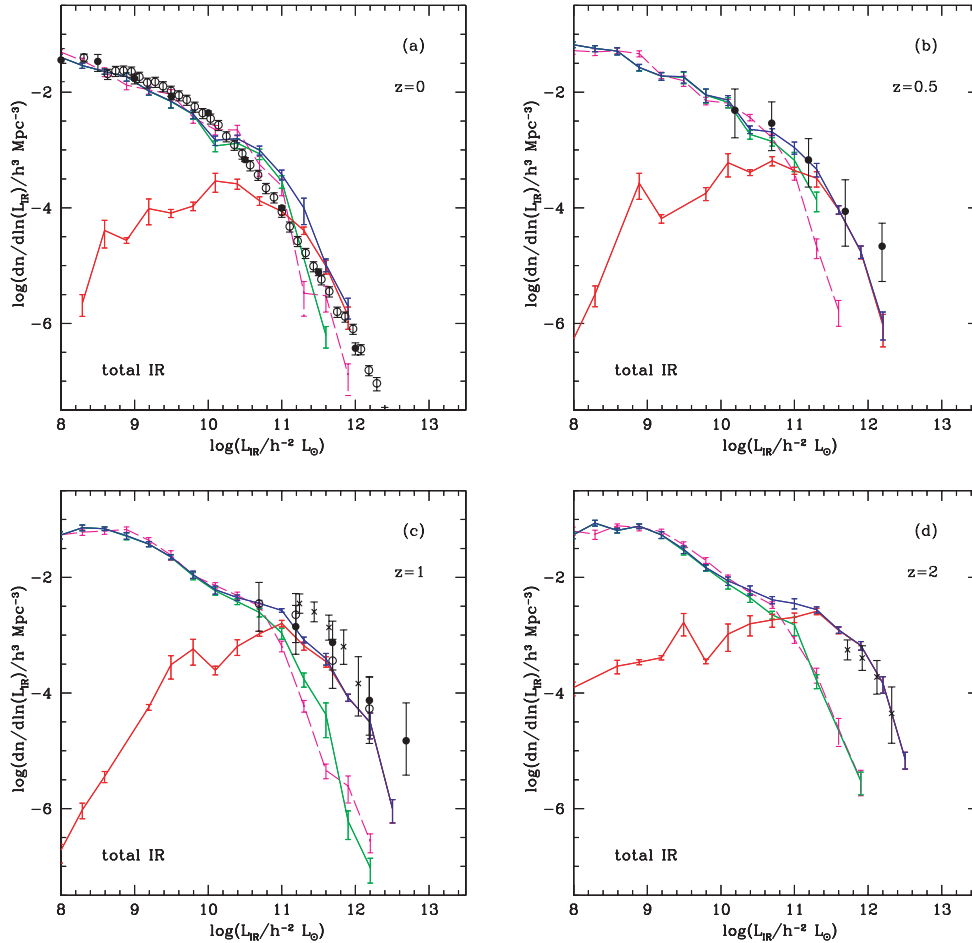


Figure 12. Predicted evolution of the total mid- + far-IR (8–1000 μm) galaxy LF compared to observational data. The different panels show redshifts (a) $z = 0$, (b) $z = 0.5$, (c) $z = 1$ and (d) $z = 2$. For $z = 0$, we compare with observational data from Sanders et al. (2003) (filled symbols) and Takeuchi et al. (2003) [open symbols, converting his 60- μm LF to a total IR LF assuming a constant conversion factor, $L_{\text{IR}}/\nu L_{\nu}(60 \mu\text{m}) = 2.5$]. We compare with data from Le Flocc’h et al. (2005) for $z = 0.5$ and 1 (filled and open symbols), and with Caputi et al. (2007) for $z = 1$ and 2 (crosses).

scatter at high redshifts results in part from having two different IMFs.

The luminosity in the mid- and far-IR is widely used as a tracer of dust-obscured star formation (although in galaxies with very low SFRs, the dust heating can be dominated by older stars). The total mid- + far-IR (rest-frame 8–1000 μm) luminosity is expected to provide a more robust tracer of star formation than the luminosity at any single IR wavelength, since the shape of the SED of dust emission depends on the dust temperature distribution (as well as on the dust grain properties). This is borne out by our model predictions. The middle right-hand panel of Fig. 13 shows that the L_{IR}/SFR ratio depends weakly on both SFR and redshift. This behaviour results mostly from having different IMFs in the model in quiescent and bursting galaxies, with the fractional contribution of the bursts increasing both with SFR and with redshift. If we look at quiescent and bursting galaxies separately, we find roughly constant ratios $L_{\text{IR}}/\text{SFR} \approx 6 \times 10^9 h^{-1} L_{\odot}/M_{\odot}$ and $L_{\text{IR}}/\text{SFR} = 2 \times 10^{10} h^{-1} L_{\odot}/M_{\odot}$, respectively, for galaxies where L_{IR} is powered mostly by young stars. However, there is also a trend at lower redshift for L_{IR}/SFR to be larger at lower SFR – this reflects the larger fraction of dust heating from older stars in galaxies with lower SFRs, which more than compensates for the lower average dust obscuration in these galaxies. The lower right-hand panel of Fig. 13 shows

that the L/SFR ratio in the mid-IR (in this case at 15 μm in the rest-frame) shows more variation with SFR and redshift than the ratio for the total IR luminosity. This reflects the variation in the mid- to far-IR SED shapes in the model. The scatter in the L/SFR ratio is roughly a factor of 2 around the average relation for the total IR luminosity, but is larger for the 15- μm luminosity.

The results of this section illustrate why it is not straightforward to compare theoretical predictions for the evolution of the GSMF and GSFrd (or even the stellar mass and SFR densities) with observational estimates. In addition to assumptions about galaxy star formation histories and metallicities (for stellar mass estimates), and about the SED shapes for dust emission (for SFR estimates from IR and submillimetre data), observational estimates all rest on some assumed form for the IMF. If the IMF assumed in the observational analysis is different from the true IMF, the observational estimates for stellar masses and SFRs can be wrong by large factors. If the IMFs differ only below 1 M_{\odot} , then one can apply a simple rescaling to relate stellar mass and SFR estimates for different IMFs. However, if our current galaxy formation model is correct, stars form with different IMFs in quiescent discs and in merger-driven bursts, and so no observational estimate based on assuming a single IMF can give the correct GSMFs and GSFrd, nor the correct stellar mass and SFR densities. A direct comparison of the GSMF

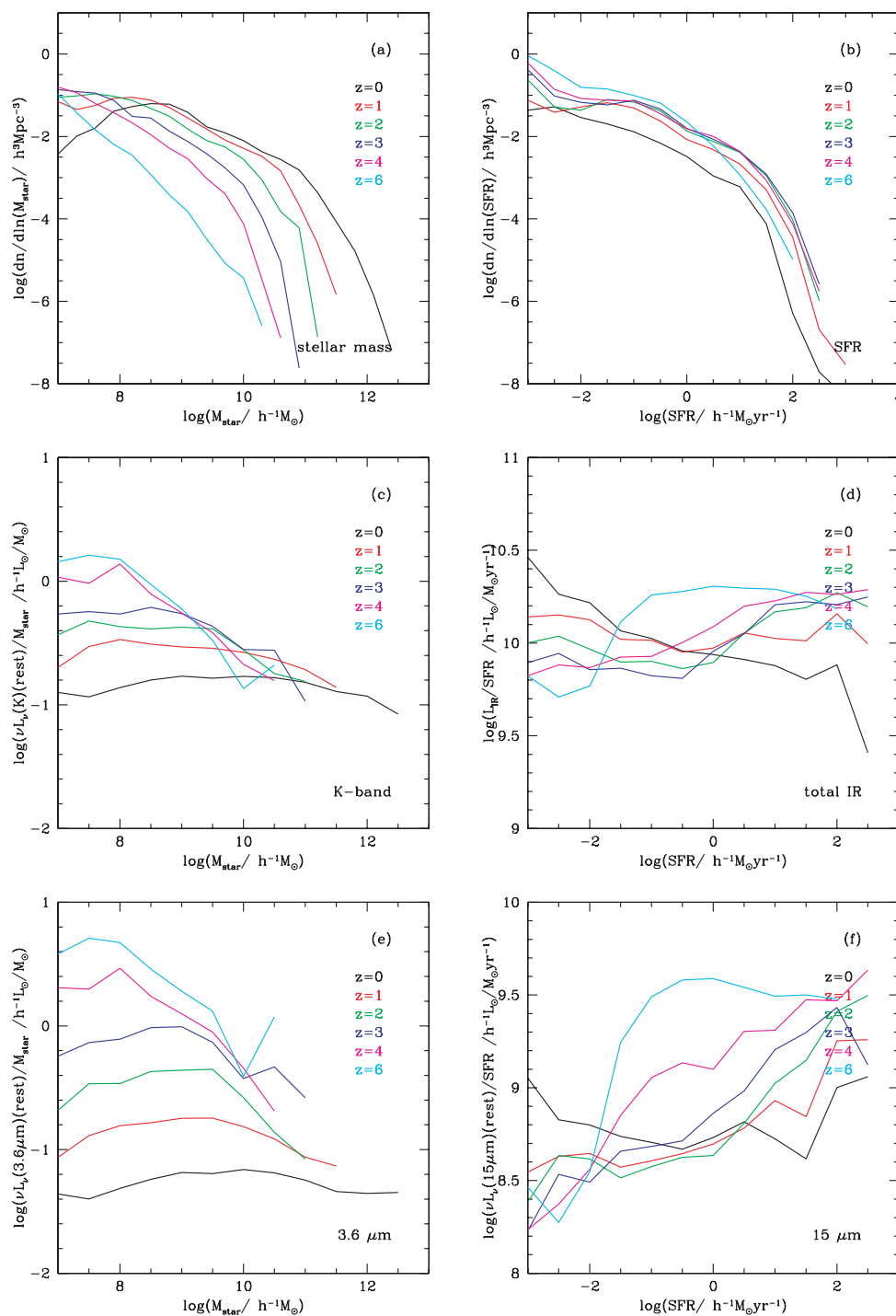


Figure 13. Model predictions for properties related to stellar masses (left-hand column) and SFRs (right-hand column), for redshifts $z = 0, 1, 2, 3, 4$ and 6 : (a) GSMF; (b) GSFrd; (c) mean ratio of rest-frame K -band luminosity to stellar mass, as a function of stellar mass; (d) mean ratio of total mid- + far-IR luminosity to SFR, as a function of SFR; (e) mean ratio of rest-frame $3.6\text{-}\mu\text{m}$ luminosity to stellar mass, as a function of stellar mass; (f) mean ratio of rest-frame $15\text{-}\mu\text{m}$ luminosity to SFR, as a function of SFR. (The $15\text{-}\mu\text{m}$ luminosity is here calculated through top-hat filter with a fractional wavelength width of 10 per cent.)

and GSFrd evolution predicted by our model with observational estimates is therefore not meaningful. Instead, the comparison between models and observations must be made via directly observable (rather than inferred) quantities, such as the K -band luminosities to constrain stellar masses, and the total IR luminosities to constrain SFRs.

6 CONCLUSIONS

We have computed predictions for the evolution of the galaxy population at IR wavelengths using a detailed model of hierarchical galaxy formation and of the reprocessing of starlight by dust, and compared these predictions with observational data from the *Spitzer*

Space Telescope. We calculated galaxy formation in the framework of the Λ CDM model using the GALFORM semi-analytical model, which includes physical treatments of the hierarchical assembly of dark matter haloes, shock-heating and cooling of gas, star formation, feedback from supernova explosions and photoionization of the IGM, galaxy mergers and chemical enrichment. We computed the IR luminosities and SEDs of galaxies using the GRASIL multiwavelength spectrophotometric model, which computes the luminosities of the stellar populations in galaxies, and then the reprocessing of this radiation by dust, including radiative transfer through a two-phase dust medium, and a self-consistent calculation of the distribution of grain temperatures in each galaxy based on a local balance between heating and cooling. The GRASIL model includes a treatment of the emission from PAH molecules, which is essential for understanding the mid-IR emission from galaxies.

Our galaxy formation model incorporates two different IMFs: quiescent star formation in galaxy discs occurs with a normal solar neighbourhood IMF, but star formation in bursts triggered by galaxy mergers happens with a top-heavy $x = 0$ IMF. In a previous paper (Baugh et al. 2005), we found that the top-heavy IMF in bursts was required in order that the model reproduces the observed number counts of the faint SMGs detected at $850\ \mu\text{m}$, which are typically ultraluminous starbursts at $z \sim 2$, with total IR luminosities $L_{\text{IR}} \sim 10^{12} - 10^{13} L_{\odot}$. This conclusion was arrived at following a search of a large grid of model parameters, with the imposition of a variety of detailed observational constraints. The parameters in the Baugh et al. (2005) model were chosen before the publication of any results from *Spitzer*, without reference to any IR data apart from the local $60\text{-}\mu\text{m}$ LF and the $850\text{-}\mu\text{m}$ galaxy counts. We have kept the same parameter values in the present paper, in order to test what the same model predicts at other wavelengths and other redshifts. By doing this, we hope to address the criticism made of many semi-analytical models that they have no predictive power, because their parameters are always adjusted to match the observational data being analysed at that instant.

We first compared the predictions from our model with the galaxy number counts measured in all seven *Spitzer* bands, from 3.6 to $160\ \mu\text{m}$. We found broad agreement between the model and the observations. In the four IRAC bands ($3.6\text{--}8.0\ \mu\text{m}$), where the counts are mostly dominated by emission from older stellar populations, we found that the predicted counts were insensitive to whether we had a top-heavy or normal IMF in bursts. On the other hand, in the MIPS bands ($24\text{--}160\ \mu\text{m}$), where the counts are dominated by emission from dust in star-forming galaxies, the predicted counts are more sensitive to the choice of IMF, and the counts are fit better by the model with a top-heavy IMF. We next investigated the evolution of the galaxy LF at IR wavelengths, where several groups have now used *Spitzer* data to try to measure the evolution of the galaxy LF over the redshift range $z \sim 0\text{--}2$, at rest-frame wavelengths from 3.6 to $24\ \mu\text{m}$.

Our model predicts that at mid- and far-IR rest-frame wavelengths, the LF evolution is very sensitive to the choice of IMF in bursts. We found that our standard model with a top-heavy IMF in bursts fits the measured evolution of the mid-IR LF remarkably well (when allowance is made for complexity of predicting dust emission in the mid-IR), without any adjustment of the parameters. On the other hand, a model with a normal IMF in bursts predicts far too little evolution in the mid-IR LF compared to what is observed. We made a similar comparison with the evolution of the total IR LF, where in the case of the observations, the total IR luminosities at high redshifts have been inferred from the $24\text{-}\mu\text{m}$ fluxes by fitting SEDs, and reached the same conclusion. The evolution of the

galaxy LF in the mid-IR found by *Spitzer* thus supports our original conclusion about the need for a top-heavy IMF in bursts, which was based only on the submillimetre counts. This conclusion will be further tested by ongoing *Spitzer* surveys at longer wavelengths. To assist this, we have also presented predictions for the evolution of the LF in the *Spitzer* 70- and $160\text{-}\mu\text{m}$ bands.

We have also presented predictions for the evolution of the stellar mass function and SFR distribution of galaxies. We investigated how the L/M_* and L/SFR ratios varied with galaxy mass, SFR and redshift in different IR wavelength ranges, and considered the implications for observational estimates of stellar masses and SFRs from IR observations. Even in the near-IR, the predicted variations in L/M_* with mass and redshift can be surprisingly large. The variations in L/M_* are much larger at a rest-frame wavelength of $3.6\ \mu\text{m}$ than at $2.2\ \mu\text{m}$, implying that the $2.2\text{-}\mu\text{m}$ luminosity is a more robust tracer of stellar mass.

Finally, we have presented in Appendix A the predictions of our model for the redshift distributions of galaxies selected at different IR fluxes in the *Spitzer* bands.

One significant limitation of our model is that it does not include the effects of AGN. Two effects are relevant here. The first is feedback from AGN on galaxy formation. In several recent galaxy formation models, AGN feedback is invoked to prevent the formation of too many massive galaxies at the present day. In the model presented here, we instead posit feedback from supernova-driven galactic superwinds, which perform a similar role to AGN feedback in suppressing the formation of very massive galaxies. Both the superwind and AGN feedback models include free parameters which are tuned to give a match to the present-day optical galaxy LF. However, the redshift dependence of the feedback will be different between our superwind model and the various AGN feedback models, so in general they will all predict different evolution of the galaxy population with redshift. We will investigate galaxy evolution in the IR in a model with AGN feedback in a future paper. The second effect of AGN which we have not included is the emission from AGN and their associated dust tori. In order to compensate for this, we have wherever possible compared our model predictions with observations from which the AGN contribution has been subtracted out. This was possible for most of our comparisons of LF evolution. This was not possible for the number counts comparisons, but in this case the contribution from AGN is thought (based on observations) to be a small fraction of the total over the flux range explored by *Spitzer*, even in the mid-IR where the dust tori are the most prominent. We therefore believe that emission from AGN does not seriously affect our conclusions about the IR evolution of star-forming galaxies. We hope to include AGN emission directly into our models in the future.

We have thus shown that *Spitzer* data provide a stringent test of galaxy formation theory, by probing galaxy evolution, constraining SFRs and the role of dust to $z \sim 2$. We find that an *ab initio* Λ CDM model gives an acceptable fit to the *Spitzer* data provided that ~ 10 per cent of the stars in galaxies today formed in bursts of star formation with a top-heavy IMF. Future facilities like *Herschel*, *SPICA*, *JWST* and *ALMA* will continue to exploit the valuable information on galaxy formation contained in the IR part of the electromagnetic spectrum.

ACKNOWLEDGMENTS

We thank T. Babbedge, K. Caputi, A. Franceschini, E. Le Floch and P. Perez-Gonzalez, for providing us with their observational data

in a convenient form. CMB acknowledges the receipt of a Royal Society University Research Fellowship. CSF is the recipient of a Royal Society Wolfson Research Merit Award. This work was also supported by the PPARC rolling grant for extragalactic astronomy and cosmology at Durham.

REFERENCES

- Almeida C., Baugh C. M., Lacey C. G., 2007, *MNRAS*, 376, 1711
- Babbedge T. S. R. et al., 2006, *MNRAS*, 376
- Barnes J., 1998, in Kennicutt R. C. Jr, Schweizer F., Barnes J. E., Friedli D., Martinet L., Pfenniger D., eds, *Saas-Fee Advanced Course 26, Galaxies: Interactions and Induced Star Formation*. Springer-Verlag, Berlin, Heidelberg, p. 275
- Baugh C. M., 2006, *Rep. Prog. Phys.*, 69, 3101
- 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
- Benson A. J., Lacey C. G., Frenk C. S., Baugh C. M., Cole S., 2002, *MNRAS*, 333, 156
- Benson A. J., Bower R. G., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2003, *ApJ*, 599, 38
- Birnboim Y., Dekel A., 2003, *MNRAS*, 345, 349.
- Blain A. W., Jameson A., Smail I., Longair M. S., Kneib J. P., Ivison R. J., 1999, *MNRAS*, 309, 715
- 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
- Bressan A., Granato G. L., Silva L., 1998, *A&A*, 332, 135
- Bressan A., Silva L., Granato G. L., 2002, *A&A*, 392, 377
- Caputi K. I. et al., 2006, *ApJ*, 637, 727
- Caputi K. I. et al., 2007, *ApJ*, 660, 97
- Cattaneo A., Dekel A., Devriendt J., Guiderdoni B., Blaizot J., 2006, *MNRAS*, 370, 165
- Chapman S. C., Blain A. W., Smail I., Ivison R. J., 2005, *ApJ*, 622, 772
- Chary R., Elbaz D., 2001, *ApJ*, 556, 562
- Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, *MNRAS*, 319, 168
- Croton D. J. et al., 2006, *MNRAS*, 365, 11
- Devriendt J. E. G., Guiderdoni B., 2000, *A&A*, 363, 851
- Dole H. et al., 2001, *A&A*, 372, 364
- Dole H., Lagache G., Puget J.-L., 2003, *ApJ*, 585, 617
- Dole H. et al., 2004a, *ApJS*, 154, 87
- Dole H. et al., 2004b, *ApJS*, 154, 93
- Elbaz D. et al., 1999, *A&A*, 351, L37
- Elbaz D., Cesarsky C. J., Chantal P., Aussel H., Franceschini A., Fadda D., Chary R. R., 2002, *A&A*, 384, 848
- Eyles L. P., Bunker A. J., Stanway E. R., Lacy M., Ellis R. S., Doherty M., 2005, *MNRAS*, 364, 443
- Fardal M. A., Katz N., Weinberg D. H., Dave R., 2007, *MNRAS*, 379, 985
- Fazio G. G. et al., 2004a, *ApJS*, 154, 39
- Fazio G. G. et al., 2004b, *ApJS*, 154, 10
- Figier D. F., Kim S. S., Morris M., Serabyn E., Rich R. M., McLean I. S., 1999, *ApJ*, 525, 750
- Franceschini A., Aussel H., Cesarsky C. J., Elbaz D., Fadda D., 2001, *A&A*, 378, 1
- Franceschini A. et al., 2005, *AJ*, 129, 2074
- Franceschini A. et al., 2006, *A&A*, 453, 397
- Frayser D. T. et al., 2006a, *AJ*, 131, 250
- Frayser D. T. et al., 2006b, *ApJ*, 647, L9
- Granato G. L., Lacey C. G., Silva L., Bressan A., Baugh C. M., Cole S., Frenk C. S., 2000, *ApJ*, 542, 710
- Granato G. L., De Zotti G., Silva L., Bressan A., Danese L., 2004, *ApJ*, 600, 580
- Gruppioni C., Lari C., Pozzi F., Zamorani G., Franceschini A., Oliver S., Rowan-Robinson M., Serjeant S., 2002, *MNRAS*, 335, 831
- Gruppioni C., Pozzi F., Lari C., Oliver S., Rodighiero G., 2005, *ApJ*, 618, L9
- Guiderdoni B., Hivon E., Bouchet F. R., Maffei B., 1998, *MNRAS*, 295, 877
- Harayama Y., Eisenhauer F., Martins F., 2007, *ApJ*, in press (arXiv:0710.2882)
- Hauser M. G. et al., 1998, *ApJ*, 508, 25
- Helly J. C., Cole S., Frenk C. S., Baugh C. M., Benson A., Lacey C., 2003, *MNRAS*, 338, 913
- Huang J.-S. et al., 2007, *ApJ*, 664, 840
- Hughes D. H. et al., 1998, *Nat*, 394, 241
- Kauffmann G., 1996, *MNRAS*, 281, 487
- Kauffmann G., White S. D. M., Guiderdoni B., 1993, *MNRAS*, 264, 201
- Kaviani A., Haehnelt M. G., Kauffmann G., 2003, *MNRAS*, 340, 739
- Kennicutt R. C., 1983, *ApJ*, 272, 54
- Kogut A. et al. (the *WMAP* team), 2003, *ApJS*, 148, 161
- Kroupa P., 2001, *MNRAS*, 322, 231
- Lacey C., Cole S., 1993, *MNRAS*, 262, 627
- Lagache G., Dole H., Puget J.-L., 2003, *MNRAS*, 338, 555
- Larson R. B., 1998, *MNRAS*, 301, 569
- Larson R. B., 2005, *MNRAS*, 359, 211
- Lawrence A., Walker D., Rowan-Robinson M., Leech K. J., Penston M. V., 1986, *MNRAS*, 219, 687
- Le Delliou M., Lacey C., Baugh C. M., Guiderdoni B., Bacon R., Courtois H., Sousbie T., Morris S. L., 2005, *MNRAS*, 357, L11
- Le Delliou M., Lacey C., Baugh C. M., Morris S. L., 2006, *MNRAS*, 365, 712
- Le Floch E. et al., 2005, *ApJ*, 632, 169
- Li A., Draine B. T., 2001, *ApJ*, 554, 778
- Maness H. et al., 2007, *ApJ*, 669, 1024
- McCraday N., Gilbert A. M., Graham J. R., 2003, *ApJ*, 596, 240
- Monaco P., Fontanot F., Taffoni G., 2007, *MNRAS*, 375, 1189
- Mortier A. M. J. et al., 2005, *MNRAS*, 363, 563
- Nagashima M., Lacey C. G., Baugh C. M., Frenk C. S., Cole S., 2005a, *MNRAS*, 358, 1247
- Nagashima M., Okamoto T., Lacey C. G., Baugh C. M., Frenk C. S., Cole S., 2005b, *MNRAS*, 363, L31
- Papovich C. et al., 2004, *ApJS*, 154, 70
- Parra R., Conway J. E., Diamond P. J., Thrall H., Lonsdale C. J., Lonsdale C. J., Smith H. E., 2007, *ApJ*, 659, 314
- Patris J., Dennefeld M. II, Lagache G., Dole H., 2003, *A&A*, 412, 349
- Paumard T. et al., 2006, *ApJ*, 643, 1011
- Pearson C., Rowan-Robinson M., 1996, *MNRAS*, 283, 174
- Perez-Gonzalez P. G. et al., 2005, *ApJ*, 630, 82
- Pilbratt G., 2003, in Matter J. C., ed., *Proc. SPIE Vol. 4850, IR Telescopes and Instruments*. SPIE, Bellingham, p. 586
- Puget J.-L., Abergel A., Bernard J.-P., Boulanger F., Burton W. B., Desert F.-X., Hartmann D., 1996, *A&A*, 308, L5
- Rieke G. H., Loken K., Rieke M. J., Tamblyn P., 1993, *ApJ*, 412, 99
- Rowan-Robinson M., 2001, *ApJ*, 549, 745
- Rush B., Malkan M. A., Spinoglio L., 1993, *ApJS*, 89, 1
- Sanders D. B., Mirabel I. F., 1996, *ARA&A*, 34, 749
- Sanders D. B., Mazzarella J. M., Kim D.-C., Surace J. A., Soifer B. T., 2003, *AJ*, 126, 1607
- Saunders W., Rowan-Robinson M., Lawrence A., Efstathiou G., Kaiser N., Ellis R. S., Frenk C. S., 1990, *MNRAS*, 242, 318
- Shupe D. L., Fang F., Hacking P. B., Huchra J. P., 1998, *ApJ*, 501, 597
- Silva L., Granato G. L., Bressan A., Danese L., 1998, *ApJ*, 509, 103
- Silva L., De Zotti G., Granato G. L., Maiolino R., Danese L., 2005, *MNRAS*, 357, 1295
- Smail I., Ivison R. J., Blain A. W., 1997, *ApJ*, 490, L5
- Soifer B. T., Neugebauer G., 1991, *AJ*, 101, 354
- Soifer B. T., Houck J. R., Neugebauer G., 1987a, *ARA&A*, 25, 187
- Soifer B. T., Sanders D. B., Madore B. F., Neugebauer G., Danielson G. E., Elias J. H., Lonsdale C. J., Rice W. L., 1987b, *ApJ*, 320, 238
- Spergel D. N. et al., 2006, *ApJS* 170, 377
- Springel V. et al., 2005, *Nat*, 435, 629
- Stolte A., Brandner W., Grebel E. K., Lenzen R., Lagrange A., 2005, *ApJ*, 628, L113
- Takeuchi T. T., Yoshikawa K., Ishii T. T., 2003, *ApJ*, 587, L89
- Van Dokkum P., 2008, *ApJ*, 674, 29
- Vaisanen P., Tollestrup E. V., Fazio G. G., 2001, *MNRAS*, 325, 1241

Vega O., Silva L., Panuzzo P., Bressan A., Granato G. L., Chavez M., 2005, MNRAS, 364, 1286

Werner M. W. et al., 2004, ApJS, 154, 1

Wright G. S., Joseph R. D., Meikle W. P. S., 1984, Nat, 309, 430

Xu C. et al., 1998, ApJ, 508, 576

APPENDIX A: REDSHIFT DISTRIBUTIONS

In this appendix, we present some predictions from our standard model for the redshift distributions of galaxies selected at different fluxes in the *Spitzer* bands. This is principally for completeness, to assist in interpreting data from current surveys, and to assist in planning future surveys based on *Spitzer* data. The set of redshift distributions at all observed fluxes in principle contains equivalent information to that in the LFs at different wavelengths and redshifts. However, comparing models with observations via LFs is more physically transparent than making the comparison via redshift distributions, which is why we have presented our results on LFs in the main part of the paper, and why we make only a limited direct comparison with observed redshift distributions in this appendix. In addition, if one only compares the predicted and observed redshift distributions for galaxies above a single flux limit (e.g. the flux limit of a survey), this has less information than comparing the LFs at different redshifts.

We first show in Fig. A1 how the median redshift, and the 10–90th percentile range, are predicted to change with flux for galaxies

selected in one of the four *Spitzer* bands 3.6, 8.0, 24 or 70 μm . While at most wavelengths the median redshift is predicted to increase smoothly and monotonically with decreasing flux, this is not true at 24 μm , where there is a bump around $S_\nu \sim 100 \mu\text{Jy}$. The structure seen for the 24- μm band as compared to the other wavelengths results from different PAH emission features moving through the band with increasing redshift.

In Fig. A2, we show the predictions from our standard model for the redshift distributions of galaxies in the four IRAC bands. For each band, we show the redshift distribution for galaxies selected to be brighter than $S_\nu > 10 \mu\text{Jy}$ in that band. The flux limit $S_\nu > 10 \mu\text{Jy}$ has been chosen to match that in the observed deep sample selected at 3.6 μm by Franceschini et al. (2006). In each panel, the blue curve shows the predicted dN/dz for all galaxies, normalized to unit area under the curve, and the red and green curves show the separate contributions of bursting and quiescent galaxies to the total. For 3.6 μm , the black line shows the observed redshift distribution from Franceschini et al. (2006), which has also been normalized to unit area under the curve. We see that the observed redshift distribution peaks at a slightly higher redshift than in the model. However, the LF evolution derived from this same sample is in reasonable agreement with the model, as was already shown in Fig. 4. Franceschini et al. (2006) note that the peak seen in their data at $z \sim 0.8$ is partly contributed by large-scale structures in the *Chandra* Deep Field South field.

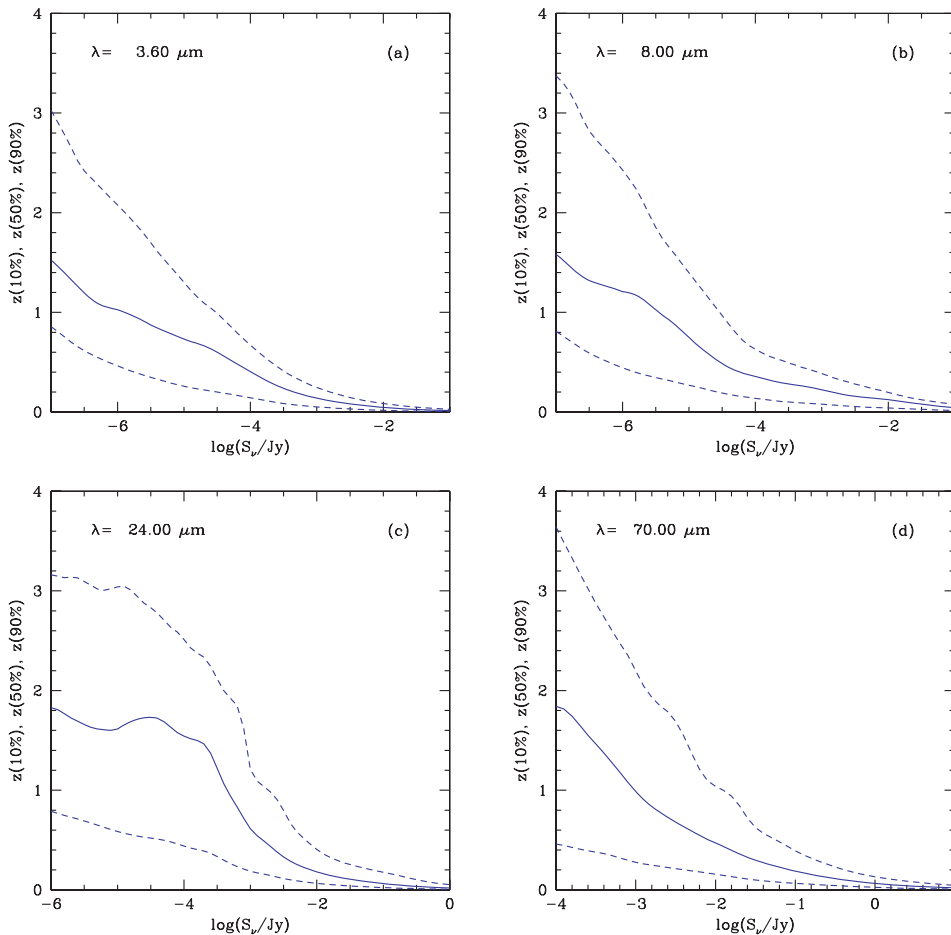


Figure A1. Model predictions for the median redshift as a function of flux in four *Spitzer* bands. (a) 3.6 μm , (b) 8.0 μm , (c) 24 μm , (d) 70 μm . In each panel, the median redshift for galaxies at each flux is shown by a solid line, and the 10th and 90th percentile are shown by dashed lines.

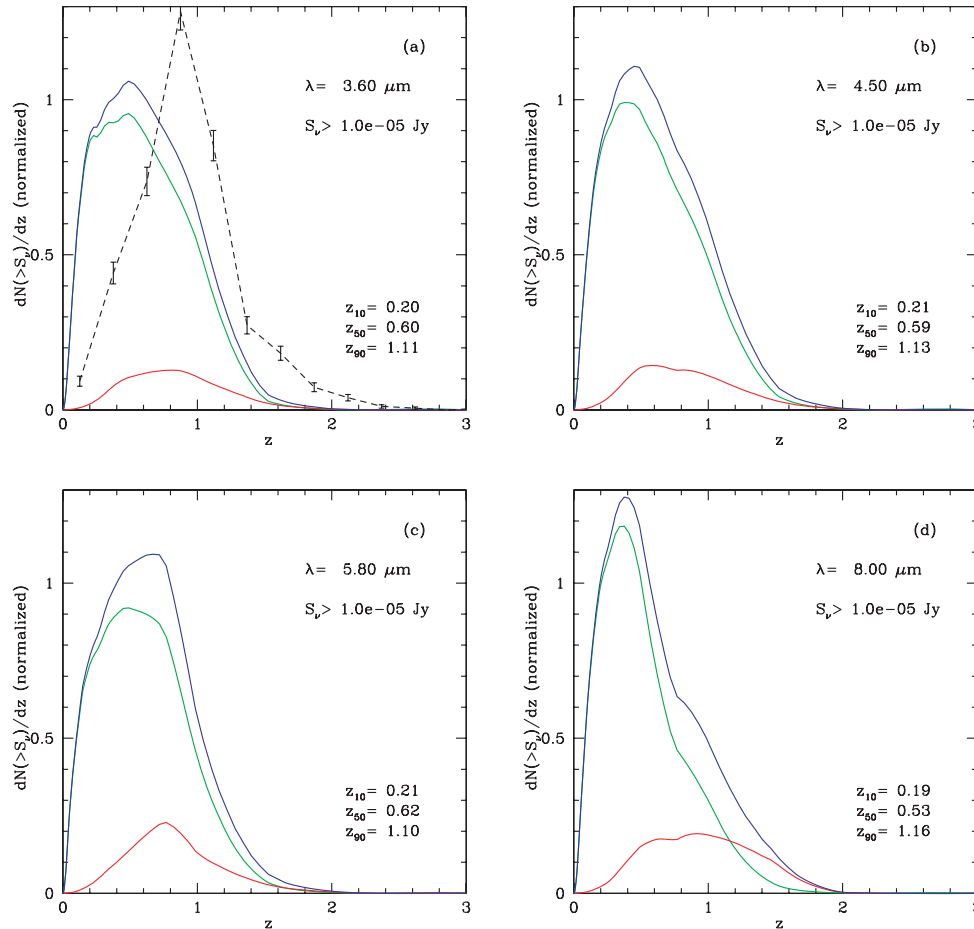


Figure A2. Predicted galaxy redshift distributions in the four IRAC bands, for galaxies brighter than $S_v = 10 \mu\text{Jy}$. (a) $3.6 \mu\text{m}$, (b) $4.5 \mu\text{m}$, (c) $5.8 \mu\text{m}$ and (d) $8.0 \mu\text{m}$. The model curves (which all include the effects of dust) are as follows: blue – total; red – ongoing bursts; green – quiescent galaxies. The curves are normalized to unit area under the curve for the total counts. The median (z_{50}) and 10th and 90th percentile (z_{10} , z_{90}) redshifts for the total counts in each band are also given in each panel. For $3.6 \mu\text{m}$, the model predictions are compared with observational data from Franceschini et al. (2006) (black dashed line), normalized to unit area as for the models. The error bars plotted on the observational data include Poisson errors only.

In Fig. A3, we show predicted redshift distributions for galaxies selected to be at a set of different fluxes in the four IRAC bands. The curves for the different fluxes are all normalized to have unit area as before, but in this figure the galaxies are selected to be at a particular flux, rather than being brighter than a certain flux. As one would expect, the typical redshift increases as the flux decreases.

Figs A4 and A5 show for the three MIPS bands the equivalent of Figs A2 and A3 for the IRAC bands. In Fig. A4, we show the predicted redshift distributions for galaxies brighter than a particular flux, where this flux limit is taken to be $83 \mu\text{Jy}$ at $24 \mu\text{m}$, 10 mJy at $70 \mu\text{m}$ and 100 mJy at $160 \mu\text{m}$. The flux limit at $24 \mu\text{m}$ has been chosen to match that used in the deep observational samples of Le Floc’h et al. (2005), Perez-Gonzalez et al. (2005) and Caputi et al. (2006), while the flux limits at 70 and $160 \mu\text{m}$ have been chosen to be roughly three times brighter than the source confusion limits in these bands. We see in Fig. A5 that the redshift distributions at $24 \mu\text{m}$ show much more structure than at other wavelengths. This results from different PAH emission features moving through the $24\text{-}\mu\text{m}$ band with changing redshift.

In Fig. A4(a), we compare the predicted redshift distribution at $24 \mu\text{m}$ with observational determinations from Perez-Gonzalez et al. (2005) (dashed black line) and Caputi et al. (2006) (solid black line). The observed distributions have been separately normalized to unit

area under the curve, as for the model distribution. Both observed distributions are based primarily on photometric redshifts, but the photometric redshifts of Caputi et al. (2006) are likely to be more accurate than those of Perez-Gonzalez et al. (2005), since the former are based on deeper optical and K -band data than the latter. (Perez-Gonzalez et al. found optical counterparts with $B_{\text{AB}} \lesssim 24.7$ or $R_{\text{AB}} \lesssim 23.7$ for ~ 70 per cent of their $S_v(24 \mu\text{m}) > 83 \mu\text{Jy}$ sources, but relied on IRAC fluxes in deriving photo- z values for the remaining ~ 30 per cent of their sample. On the other hand, Caputi et al. found K -band counterparts with $K(\text{Vega}) < 21.5$ for 95 per cent of their $S_v(24 \mu\text{m}) > 80 \mu\text{Jy}$ sample, and derived photo- z values for essentially all of these sources using optical and K -band data alone). Both observed distributions are similar, but the Caputi et al. distribution shows more structure. This is a combination of the effects of more accurate photometric redshifts but also a nine times smaller survey area, which means that fluctuations due to galaxy clustering are larger. Caputi et al. argue that the separate peaks at $z \sim 0.7$ and 1.1 result from large-scale structure, but that the bump at $z \sim 1.9$ results from PAH emission features entering the observed $24\text{-}\mu\text{m}$ band. We see that the model also predicts peaks in the redshift distribution at $z \sim 0.3$, 1 and 2 , which can be explained by different PAH features moving through the $24\text{-}\mu\text{m}$ band, although the $z \sim 2$ peak is more prominent than is seen in the observational data. Overall, the model

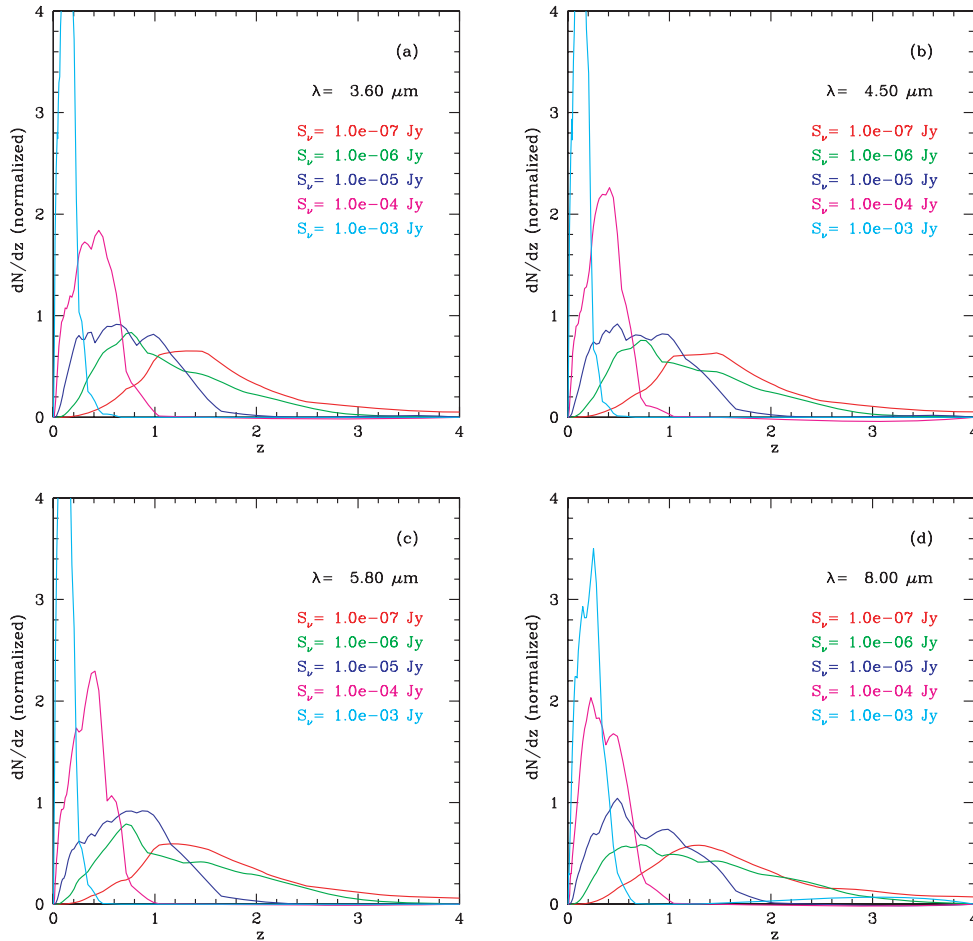


Figure A3. Predicted galaxy redshift distributions in the four IRAC bands, for different fluxes. (a) 3.6 μm , (b) 4.5 μm , (c) 5.8 μm and (d) 8.0 μm . In this figure, the redshift distributions are for galaxies at a particular flux. Predictions are shown for fluxes $S_v = 0.1, 1, 10, 100$ and $1000 \mu\text{Jy}$, as shown in the key. In all cases, the model curves are normalized to unit area, and include the effects of dust.

redshift distribution at this flux limit is too skewed to high redshift compared to the observations, predicting too few galaxies at $z \sim 0.5$ – 1 , and too many in the peak at $z \sim 2$.

We investigate further this apparent discrepancy in the $24 \mu\text{m}$ redshift distribution in Fig. A6, where we show the effects of apparent magnitude limits in the R and K bands on the predicted redshift distributions for $S_v(24 \mu\text{m}) > 83 \mu\text{Jy}$. In this plot, the redshift distributions are plotted as number per solid angle, without normalizing to unit area under the curve. The left- and right-hand panels, respectively, have the redshift distributions of Perez-Gonzalez et al. and Caputi et al. overplotted. We concentrate on the comparison with Caputi et al., since this has the simpler sample selection and more accurate redshifts. The model prediction for $K < 21.5$ (which is the magnitude limit used by Caputi et al.) is shown by the short-dashed blue line, while the prediction with no limit on the K magnitude is shown by the solid blue line. The model dN/dz with no limit on the

K magnitude is most discrepant with the Caputi et al. data at $z \sim 2$, where it predicts around two times too many galaxies. This is directly related to the fact that the predicted LF at $z = 2$ at rest-frame wavelength $8 \mu\text{m}$ (corresponding to observed wavelength $24 \mu\text{m}$) and luminosity $\sim 10^{11} L_\odot$ is also around two times too high compared to what Caputi et al. estimate from their data, as shown in Fig. 5(c). When the effect of the $K < 21.5$ limit is included, the predicted redshift distribution is closer to the observational data, but only 58 per cent of the model galaxies are brighter than this K -band magnitude limit, as against 95 per cent in the observed sample of Caputi et al. We conclude that the main reason for the discrepancy between the predicted and observed redshift distributions at $24 \mu\text{m}$ is that the model predicts a rest-frame $8\text{-}\mu\text{m}$ LF at $z \sim 2$ which is somewhat too high at luminosities $\sim 10^{11} L_\odot$, even though it reproduces quite well the general features of the evolution of the mid-IR LF.

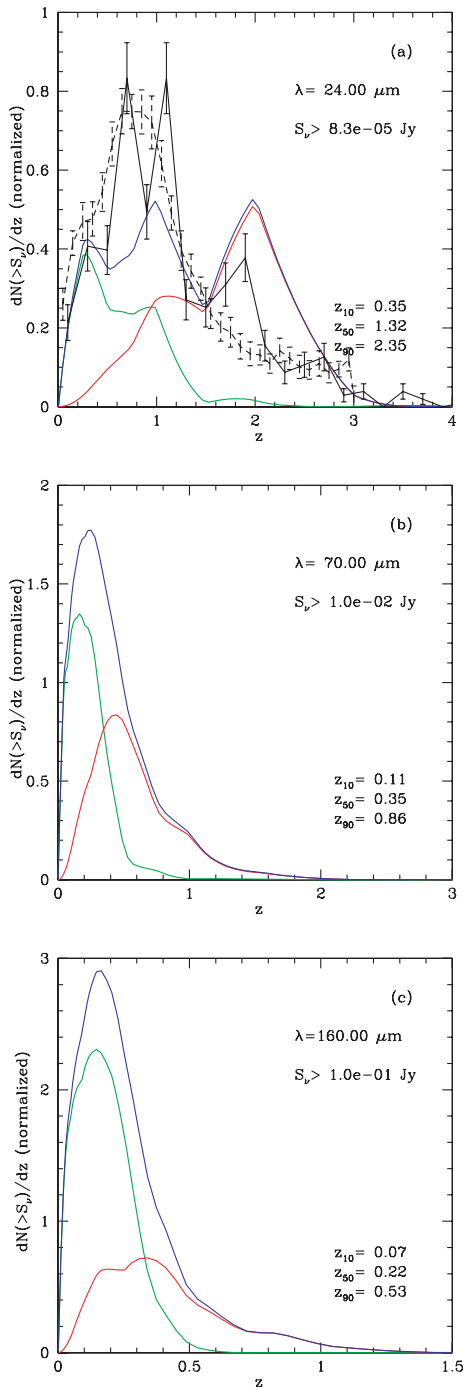


Figure A4. Predicted galaxy redshift distributions in the three MIPS bands, for galaxies brighter than a specified flux. (a) $24 \mu\text{m}$, $S_v > 83 \mu\text{Jy}$, (b) $70 \mu\text{m}$, $S_v > 10 \text{ mJy}$ and (c) $160 \mu\text{m}$, $S_v > 100 \text{ mJy}$. The model curves are as follows: blue – total; red – ongoing bursts; green – quiescent galaxies. The curves are normalized to unit area under the curve for the total counts. The median (z_{50}) and 10th and 90th percentile (z_{10} , z_{90}) redshifts for the total counts in each band are also given in each panel. For $24 \mu\text{m}$, the model predictions are compared with observational data from Caputi et al. (2006) (solid black line) and Perez-Gonzalez et al. (2005) (dashed black line), normalized to unit area as for the models. The error bars plotted on the observational data include Poisson errors only for Caputi et al., but also include errors in photometric redshifts for Perez-Gonzalez et al.

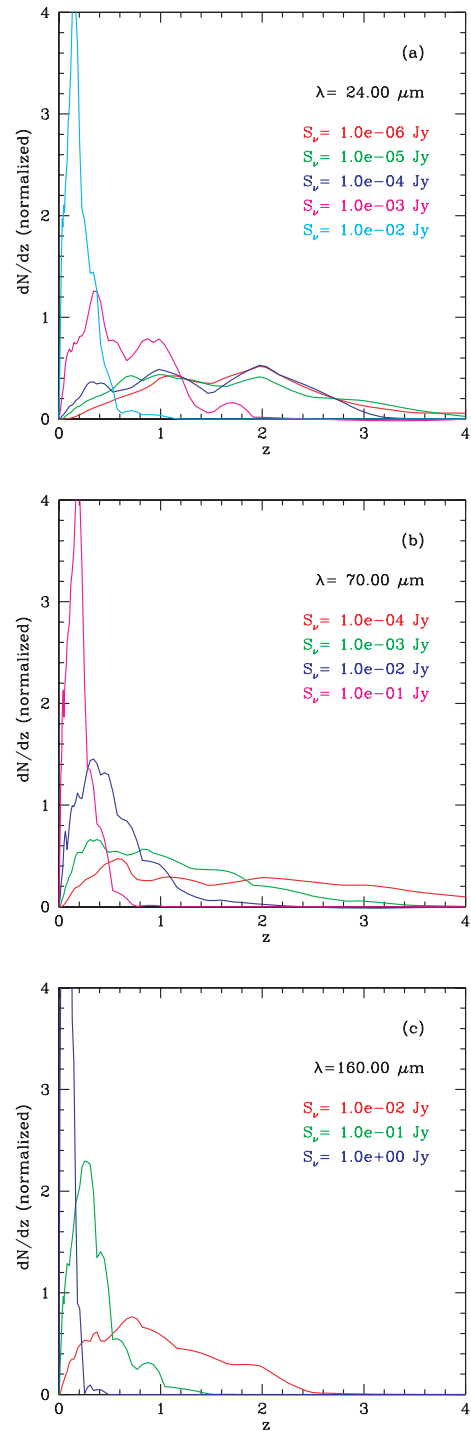


Figure A5. Predicted galaxy redshift distributions in the three MIPS bands, for different fluxes. (a) $24 \mu\text{m}$, (b) $70 \mu\text{m}$ and (c) $160 \mu\text{m}$. In this figure, the redshift distributions are for galaxies at a particular flux, as shown in the key in each panel. In all cases, the model curves are normalized to unit area, and include the effects of dust.

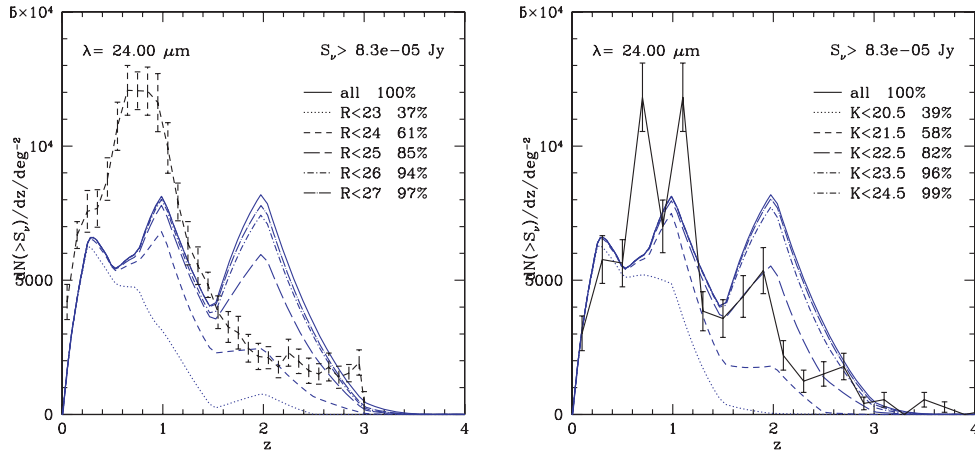


Figure A6. Predicted redshift distributions at 24 μm , showing the effects of optical or near-IR magnitude limits. Model galaxies are selected with $S_p > 83 \mu\text{Jy}$ together with the optical/near-IR magnitude limits as shown in the key. The fraction of 24 μm sources brighter than each magnitude limit is also given. (a) R -band magnitude limit. The observed redshift distribution from Perez-Gonzalez et al. (2005) is overplotted in black. Note Le Floc'h et al. (2005) used $R < 24$ and obtained 54 per cent completeness. (b) K -band magnitude limit. The observed redshift distribution from Caputi et al. (2006) (with $K < 21.5$) is overplotted. Magnitudes are on the Vega system.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.