

The stellar evolution of Luminous Red Galaxies, and its dependence on colour, redshift, luminosity and modelling

Rita Tojeiro*¹, Will J. Percival¹, Alan F. Heavens², Raul Jimenez³

¹*Institute of Cosmology and Gravitation, Dennis Sciama Building, University of Portsmouth, Burnaby Road, Portsmouth, PO1 3FX*

²*SUPA, Institute for Astronomy, Royal Observatory Edinburgh, University of Edinburgh, Blackford Hill, EH9 3HJ*

³*ICREA & Institute of Sciences of the Cosmos (ICCUB), University of Barcelona, Barcelona 08028, Spain*

11 November 2010

ABSTRACT

We present a series of colour evolution models for Luminous Red Galaxies (LRGs) in the 7th spectroscopic data release of the Sloan Digital Sky Survey (SDSS), computed using the full-spectrum fitting code VESPA on high signal-to-noise stacked spectra. The colour-evolution models are computed as a function of colour, luminosity and redshift, and we do not a-priori assume that LRGs constitute a uniform population of galaxies in terms of stellar evolution. By computing star-formation histories from the fossil record, the measured stellar evolution of the galaxies is decoupled from the survey's selection function, which also evolves with redshift. We present these evolutionary models computed using three different sets of Stellar Population Synthesis (SPS) codes. We show that the traditional fiducial model of purely passive stellar evolution of LRGs is broadly correct, but it is not sufficient to explain the full spectral signature. We also find that higher-order corrections to this model are dependent on the SPS used, particularly when calculating the amount of recent star formation. The amount of young stars can be non-negligible in some cases, and has important implications for the interpretation of the number density of LRGs within the selection box as a function of redshift. Dust extinction, however, is more robust to the SPS modelling: extinction increases with decreasing luminosity, increasing redshift, and increasing $r - i$ colour. We are making the colour evolution tracks publicly available at http://www.icg.port.ac.uk/~tojeiror/lrg_evolution/.

Key words: galaxies: evolution - cosmology: observations - surveys

1 INTRODUCTION

In the currently favoured model of galaxy formation, structure builds up hierarchically and galaxies build up their stellar mass within host dark matter haloes via a complex mixing of two processes - star formation from cold gas, and merging. This model is supported by observed large-scale structure and its evolution, which can be well understood and remarkably well modelled within the Λ Cold Dark-Matter scenario (LCDM). Because of the complicated astrophysics involved, the link between mass build-up and star-formation is not direct. There are two main observables with which to understand this link: stellar light, which holds information about the formation history of the stars present in any given galaxy at the time of observation, and direct observation of size, density and clustering of individual galaxies and their evolution with redshift. The combination of these

two observables has the potential to resolve a long-standing quest of galaxy evolution - to match samples of galaxies at different redshifts, without making assumptions about their stellar or dynamical evolution.

Luminous Red Galaxies (LRGs), for which typically the label of Early-Type Galaxy (ETG) also applies, occupy an interesting position within the galaxy evolution puzzle. They occupy and dominate the high-mass end of the galaxy mass function and are predicted to occupy the most massive dark matter halos. According to LCDM, these halos must have started their assembly earlier, and such a signature is indeed seen in the galaxy stellar populations - LRGs are clearly dominated by old stars, with 70-90 per cent of the present stellar mass having been formed at $z > 1$, with the detailed measurement dependent on the study and exact sample (Wake et al. 2006; Brown et al. 2007; Cool et al. 2008). van Dokkum et al. (2010) have extended this to higher redshift, and measured growth of massive galaxies - albeit not exclusively red galaxies - as factor of 2 since $z = 2$. They

* E-mail: rita.tojeiro@port.ac.uk

concluded this growth was mostly due to mergers, rather than star formation. There is also evidence that more massive LRGs have, on average, older stellar populations, giving strength to the hierarchical formation scenario (e.g. Thomas et al. 2005; Jimenez et al. 2007). What is not clear, however, is why star-formation seems to have shut down in these massive galaxies since roughly a redshift of two (see Peng et al. 2010 for a recent proposal, and references within). The low-level of recent star formation suggests that these galaxies have either been dynamically passively evolving, or that any mass build up through merging in the recent Universe must have occurred through either dry merging (resulting in no new star-formation), or minor mergers (perhaps inducing only very small levels of star formation). It is also possible that the recent star formation observed in LRGs is due completely or in part to the gas returned by stars in the giant branch (AGB) to the interstellar medium. Given the low levels of star formation activity in these galaxies, such a source of gas would provide a natural explanation for the low level of star formation activity observed in LRGs at $z < 1$.

Measuring the mass build-up of massive galaxies is therefore a crucial and stringent test of models of galaxy formation. In spite of being a puzzle in itself, the process that shut down star formation in massive early-type galaxies has made this particular population extremely attractive for observational cosmologists. The fact that they are luminous and are thought to have a simple stellar evolution make them easy to target and observe at a range of redshifts. The ability to predict the colour evolution as a function of redshift ensures an efficient allocation of fibres or slits for spectroscopic measurements, and the fact they are very massive makes them observable to larger distances. Finally, LRGs are often approximated as being a uniform population of galaxies that experiences no significant amount of merging. This is attractive as the computation of their predicted density and velocity bias evolution then becomes trivial (Fry 1996), enabling more information to be recovered using the clustering of this population to form a standard ruler to measure cosmological geometry or using them to measure Redshift-Space Distortions.

There is a wealth of literature on LRGs and ETGs, their properties and evolution. Studies have been performed based on (see also references within): the mass or luminosity function (Wake et al. 2006; Brown et al. 2007; Faber et al. 2007; Cool et al. 2008); colour-magnitude diagram (Cool et al. 2006; Bernardi et al. 2010); photometry SED fitting (Kaviraj et al. 2009; Maraston et al. 2009); absorption line fitting to individual galaxies' spectra (Trager et al. 2000; Thomas et al. 2005, 2010; Carson & Nichol 2010) or to stacked spectra (Eisenstein et al. 2003; Graves et al. 2009; Zhu et al. 2010); full spectral fitting (Jimenez et al. 2007); close-pair counts (Bundy et al. 2009) and clustering (Sheth et al. 2006; Masjedi et al. 2006; Conroy et al. 2007; White et al. 2007; Masjedi et al. 2008; Wake et al. 2008; Tojeiro & Percival 2010; De Propris et al. 2010).

These papers can be broadly split into two types: those that aim to gain knowledge of the stellar content and evolution of these galaxies, and those that are interested in their dynamical evolution, or merging history. The picture seems set in its broad terms, but there is disagreement in the details. It is well established, from the work cited above

and many more, that ETGs constitute a very uniform population of galaxies; are dominated by old and metal rich stellar populations; their mean ages (either mass- or light-weighted) decrease with luminosity; and the most luminous occupy more dense environments. There is, however, an increasing amount of evidence pointing towards some amount of recent star formation in intermediate-mass ETGs coming, e.g. from UV excess measurements (Kaviraj et al. 2007; Salim & Rich 2010). This amount of star formation is not in conflict with the hierarchical model of structure formation, and Kaviraj et al. (2010), through evidence coming from small morphological disruptions in early-type galaxies, argue that it can be explained from the contributions from minor-mergers. The overall dynamical evolution is best constrained via a clustering analysis, although this traditionally involves assuming a model for the stellar evolution of the galaxies in the sample. Measurements vary (see Table 4 in Tojeiro & Percival 2010 for a summary), but luminosity growth seems to be confined to less than 20 per cent since a redshift of 1.

1.1 This work

The motivation for this work is two fold. In Tojeiro & Percival (2010) we constrained the dynamical evolution of Sloan Digital Sky Survey (SDSS) LRGs by assuming the stellar evolution of these galaxies was known, and followed the passive evolution model (with a small metal-poor component) of Maraston et al. (2009) (M09). Given the evidence that ETGs have different histories according to their luminosity, as well as evidence for recent star formation in at least intermediate-mass ETGs, the assumption of passive evolution after a single star burst for *all* LRGs is clearly an oversimplification. The first of our goals is therefore to provide the community with empirically derived colour evolution models for the SDSS LRG sample that depend on luminosity, colour and redshift.

Secondly, our approach of using the fossil record to infer their star-formation and metallicity histories, as well as dust content, and from there compute their colour evolution as a function of redshift makes minimal prior assumptions about the evolution of the galaxies. The method adopted is to compute the star-formation and metallicity histories using VESPA (Tojeiro et al. 2007) - a full-spectrum fitting code that assumes no prior information about the star-formation or metallicity history of a galaxy other than it is non-negative. Non-parametric full-spectral fitting has advantages and disadvantages when compared to other methods. The main advantage comes from a well time-resolved and completely unconstrained star formation history, which does not suffer from any problems and biases traditionally associated with SSP-equivalent ages and analyses (e.g. Trager & Somerville 2009). On the other hand, full-spectral fitting can be more sensitive to inaccuracies in the modelling (particularly dust), and spectrophotometric calibrations. Parameter degeneracies are not a problem, in the sense that they can be estimated and incorporated into any analysis.

There are also clear advantages in using the fossil record to compute the colour evolution of a sample of galaxies. Primarily, one does away with the circular methodology of pre-selecting a sample across a range of redshifts that one

believes is itself a uniform population of galaxies. This is traditionally done according to an evolutionary model and after applying the necessary K+e corrections and survey selection functions, and culling the observed sample to one that is self-consistent (i.e., any galaxy in this culled sample could have been observed at any redshift probed by the survey, according to a model). The obvious danger of such an approach is that the deductions about the evolution of the sample simply reflect the assumptions used to define it. In our approach, however, evolutionary colour paths that stray from the colour selection box *are* acceptable. We therefore de-couple the sample of galaxies observed at any given redshift from galaxies observed at earlier epochs, and we are insensitive to changing survey selections. Of course, if one wants to interpret this evolution in terms of the overall evolution of a population of galaxies then sample selection becomes crucially important. We leave that for a series of follow-up papers. In this work we introduce a new set of empirical models for the evolution of galaxies labelled as Luminous Red Galaxies, as a function of their observed colour, luminosity and redshift.

This paper is organised as follows. We start by describing the dataset used in this work in Section 2; in Section 3 we describe in detail the several steps of our methodology; in Section 4 we describe the three sets of SPS models used throughout the paper; we follow by presenting our results in Section 5, which we interpret in Section 6; and we finally summarise and conclude in Section 7.

2 DATA

The SDSS is a photometric and spectroscopic survey, carried from a dedicated 2.5m telescope in Apache Point, New Mexico. Photometry was taken in five bands: u, g, r, i and z , corresponding to central effective wavelengths of 3590Å, 4819Å, 6230Å, 7640Å, and 9060Å respectively. For details on the hardware, software and data-reduction see York et al. (2000) and Stoughton et al. (2002). In summary, the survey is carried out on a mosaic CCD camera (Gunn et al. 1998), two 3-arcsec fibre-fed spectrographs, and an auxiliary 0.5m telescope for photometric calibration.

Objects were selected for spectroscopic follow-up according to two main targeting algorithms. The main galaxy sample (Strauss et al. 2002) is a magnitude-limited, high-completeness (>99%) sample, selected in the r -band. The targeting is done down to $r = 17.77$, resulting in a median redshift of around $\bar{z} = 0.11$. The luminous red galaxies sample extends the redshift range to $0.15 < z < 0.5$ by targeting luminous and very red objects according to the target algorithm described in Eisenstein et al. (2001).

In this paper we analyse the latest SDSS LRG sample (data release 7, Abazajian et al. 2009), with a spectroscopic footprint of around 8000 sq. degrees and which includes around 180,000 objects. Fig. 1 shows the r -band absolute magnitudes as a function of redshift for a random subsample of our galaxies. The k-corrections were calculated using the k-correct code of Blanton & Roweis (2007). Strictly, to be self-consistent and completely model independent, we should derive k-corrections from the observed spectra. However, the difference between this approach and using the

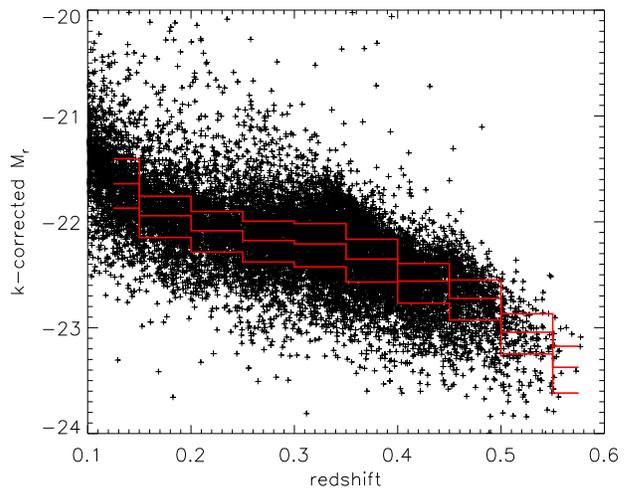


Figure 1. The redshift - absolute magnitude diagram of the DR7 LGRs used in this paper (random subsample of 20,000 objects). The magnitudes are rest-frame r -band absolute magnitudes, and are k-corrected only. The red lines show the limits that split our sample into 4 magnitudes bins at each redshift (see Section 2 for more details). k-corrections were calculated using the k-correct code of Blanton & Roweis (2007).

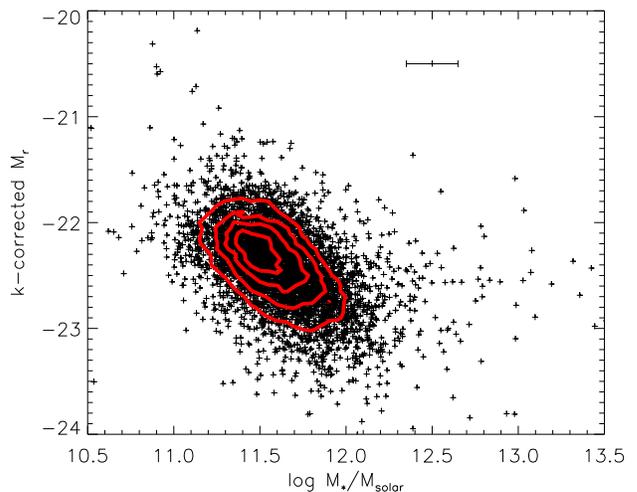


Figure 2. Absolute magnitude vs. stellar mass for the LRG sample. The stellar masses are from the public VESPA database (Tojeiro et al. 2009). The error bar at the top-right corner shows a typical error on the stellar mass.

model fits of the k-correct package will be small, and this will not significantly affect our analysis.

The goal of our work is to provide independent evolutionary histories for samples of galaxies. Luminosity evolution corrections are therefore a result from our work, and cannot be assumed a-priori. We therefore choose samples based only on k-corrected luminosities, with limits designed to optimise our sampling at each redshift, such that the number of galaxies in each luminosity range is kept approximately constant. We show these boundaries in Fig.1.

In Fig. 2 we show the stellar masses of the LRGs in our sample, and how they relate to the absolute magnitudes. The scatter in this relationship is dominated by uncertainties in the stellar mass estimation, but they are not sufficient to explain all of the observed scatter. The remaining stochasticity indicates intrinsic differences in star formation histories and mass-to-light ratios between different LRGs.

3 METHOD

Our methodology can be summarised in the following series of steps:

(i) we stack the observed spectra in regions of colour-luminosity-redshift space, in order to obtain a consistent signal-to-noise ratio (SNR) across this plane;

(ii) we analyse the stacked spectra with VESPA, in order to obtain a star-formation history (SFH), a metallicity history and dust content representative of the galaxies in any given cell of colour-luminosity-redshift space;

(iii) we estimate systematic errors due the limitation of the models, so that we can later study their impact on the recovered parameters;

(iv) we smooth the recovered SFH in lookback time, to obtain a continuous function with time; and

(v) we evolve the observed spectrum with lookback time, to obtain a rest-frame spectrum as a function of time from which we can compute any suite of observed-frame colours.

In the rest of this section we explain in detail each of these steps.

3.1 Stacking the data

We stack galaxies of similar colour, luminosity and at the same redshift. The resulting spectrum gives a weighted evolution of all the galaxies within the stack.

We work on a fixed grid in redshift, from 0.15 to 0.5 in steps of 0.05. At each redshift, we split the sample into four luminosity slices, such that each slice has approximately the same number of galaxies. Within each redshift bin and luminosity slice, we split the galaxies into stacks based on $r-i$ colour, with the number of stacked objects chosen such that each stack contains a fixed number of objects, set to be around 1000¹. This number was chosen so that the SNR of the stacked spectra would permit differentiation between ages of interest in this paper (see Section 3.3 for details). The typical photometric errors increase steeply into the blue, so the spread in $g-r$ colour of LRGs would be expected to be larger than that in $r-i$ even if LRGs did not have any intrinsic scatter in their colours. There is also a linear relationship between redshift and $g-r$ up to $z \lesssim 0.35$, which we capture by binning in redshift. We therefore stack in $r-i$ colour rather than $g-r$. The result is a number of non-uniform grids (one per luminosity slice) in $r-i$ colour, depending on the density of objects in the colour-redshift plane. We choose

¹ For the sake of clarity, we will continue to refer to each region bound by colour, luminosity redshift as a cell, to each column of cells for the same redshift as a redshift bin, and to each grid of colour-redshift for a given luminosity range as a luminosity slice.

to do this in favour of a uniform grid in order to keep a roughly constant SNR of the resulting stacked spectra. The disadvantage is that sparse regions in these diagrams need to be larger in order to reach the target number of galaxies, and variations in the intrinsic nature of the galaxies within each cell may be larger. This does not invalidate the resulting evolution model for this cell, which will simply be a roughly luminosity-weighted average evolution for all the galaxies within the cell, although it may mask larger departures from the mean within the cell. The above procedure occasionally results in cells that are smaller, in $r-i$, than the typical error on $r-i$ for *individual* galaxies. This means that there is likely scatter of galaxies across cells, and that the VESPA solutions from adjacent cells should only change smoothly (or as smoothly as the observational uncertainty in whatever quantity we bin). In other words the true resolution in $r-i$ is given by the photometric errors, and not by the size of the cells.

The goal of this paper is to compute a set of empirical evolutionary models for the *colour* of LRGs, and once again we assume that within a cell this is independent of luminosity. This assumption is strengthened by us taking more than one luminosity range, but there may be residual dependencies within a given chosen range. The purpose of the stacking procedure, therefore, is to compute an optimal estimate of the underlying spectrum, of which we assume each galaxy represents a random realisation. We discard luminosity information within each cell by normalising all spectra to a common wavelength, and we compute an inverse-variance weighted average, per wavelength point:

$$F \equiv \langle f_\lambda \rangle = \frac{\sum_i f_{\lambda,i} / \sigma_{\lambda,i}^2}{\sum_i 1 / \sigma_{\lambda,i}^2}, \quad (1)$$

where F is the stacked spectrum, $f_{\lambda,i}$ is the flux point at wavelength λ of galaxy i and $\sigma_{\lambda,i}$ the corresponding error. The sum is done over galaxies in any given cell.

The resulting error in each wavelength point of the stacked spectrum is simply

$$\sigma_\lambda^2 = \frac{1}{\sum_i 1 / \sigma_{\lambda,i}^2}. \quad (2)$$

Note that by weighting the spectra within each stack based on signal-to-noise we are not modelling the total luminosity of the galaxies within the bin when we fit to the stacked spectrum. Instead our results should be interpreted as giving the history of a *weighted function* of the galaxies within that bin, although this weighting is actually close to a luminosity weighting, as spectral S/N is closely related to luminosity.

The typical number of galaxies per stack is 1000, and we have a total of 124 stacks.

3.2 VESPA and the fossil record

We use VESPA to interpret the stacked spectrum of each cell in terms of a star formation and metallicity history. The algorithm is described in detail in Tojeiro et al. (2007), and it has been applied to SDSS's LRG and Main Galaxy samples resulting in a queryable database, described in Tojeiro et al. (2009). We refer the reader to these two papers for details

but, for completeness, we present now a very brief summary of the method.

In short, VESPA inverts the equation $F_\lambda = \hat{F}_\lambda$, where F_λ is the observed rest-frame flux, and the model flux is

$$\hat{F}_\lambda(t_0) = \int_0^{t_0} f_{dust}(\tau_\lambda, t) \psi(t) S_\lambda(t, Z) dt, \quad (3)$$

to solve for the star formation rate $\psi(t)$ (solar masses formed per unit of time) and the age t and metallicity Z of the components of the stellar populations that each give luminosity per unit wavelength $S_\lambda(t, Z)$, per unit mass. The dependency of the metallicity on age is unconstrained, turning this into a non-linear problem. To get around this, we interpolate between the tabulated values of Z given by the SPS models (see Section 4) giving a piecewise linear behaviour:

$$S'_\lambda(t, Z) = g(t) S_\lambda(t, Z_a) + [1 - g(t)] S_\lambda(t, Z_b), \quad (4)$$

where $S(t, Z_a)$ and $S(t, Z_b)$ are equivalent to $S_\lambda(t, Z)$ in Eq. (3), but at fixed metallicities Z_a and Z_b , which bound the true Z .

Solving the problem then requires finding the correct metallicity range. One should not underestimate the complexity this implies - trying all possible combination of consecutive values of Z_a and Z_b in a grid of 16 age bins would lead to a total number of calculations of the order of 10^9 , which is unfeasible even in today's fast personal workstations. We work around this problem using an iterative approach, which we describe in Tojeiro et al. (2007).

As for the treatment of dust, we follow the two-parameter dust model of Charlot & Fall (2000) in which young stars are embedded in their birth cloud up to a time t_{BC} , when they break free into the inter-stellar medium (ISM):

$$f_{dust}(\tau_\lambda, t) = \begin{cases} f_{dust}(\tau_\lambda^{ISM}) f_{dust}(\tau_\lambda^{BC}), & t \leq t_{BC} \\ f_{dust}(\tau_\lambda^{ISM}), & t > t_{BC} \end{cases} \quad (5)$$

where τ_λ^{ISM} is the optical depth of the ISM and τ_λ^{BC} is the optical depth of the birth cloud. In previous runs of VESPA, we took $t_{BC} = 0.03$ Gyrs. However, in this run we have reduced resolution at the young end (see Section 3.3 for a discussion on the revised age grid) and we do not resolve star-formation under 74 Myr. We therefore take this boundary as t_{BC} , and note that removing this dust component altogether has minimal effect in our results and conclusions.

There is a variety of choices for the form of $f_{dust}(\tau_\lambda)$. To model the dust in the ISM, we use the mixed slab model of Charlot & Fall (2000) for low optical depths ($\tau_V \leq 1$), for which

$$f_{dust}(\tau_\lambda) = \frac{1}{2\tau_\lambda} [1 + (\tau_\lambda - 1) \exp(-\tau_\lambda) - \tau_\lambda^2 E_1(\tau_\lambda)] \quad (6)$$

where E_1 is the exponential integral and τ_λ is the optical depth of the slab. This model is known to be less accurate for high dust values, and for optical depths greater than one we take a uniform screening model with

$$f_{dust}(\tau_\lambda) = \exp(-\tau_\lambda). \quad (7)$$

We only use the uniform screening model to model the dust in the birth cloud and we use $\tau_\lambda = \tau_V (\lambda/5500\text{\AA})^{-0.7}$ as our extinction curve for both environments.

Recovering the star formation history, and its dust content, is solved by making assumptions about the form of f_{dust} and $S_\lambda(t, Z)$ and by minimizing

$$\chi^2 = \frac{\sum_\lambda (F_\lambda - \hat{F}_\lambda)^2}{\sigma_\lambda^2}. \quad (8)$$

Even though the problem has an analytical solution, a dataset perturbed by noise or that is otherwise deteriorated leads to instabilities in the matrix inversion and the recovered solutions can be entirely dominated by noise (Ocvirk et al. 2006). VESPA has a self-regularization mechanism which estimates how many independent parameters one should recover from a given a dataset that has been perturbed. The result is a parametrization which varies from galaxy to galaxy, depending on its SNR and wavelength coverage.

It is the loss of resolution in lookback time, for low SNR spectra, that prompts the need for stacking in data-space. In other words, given the self-regularization mechanism and the inherent loss of - noisy! - information in cases of poor-quality data, it is *not* equivalent to stack in data or solution space. So, in general, the average of the star formation histories of the individual galaxies is not the same as the star formation history of the average spectrum. This becomes less and less true for increasingly better data of the individual objects, but in the case of the LRGs the data quality is sufficiently poor - especially at higher redshift - that stacking becomes necessary.

The physical quantities output by VESPA for each stacked spectrum are:

- star formation fraction in time interval Δt_j , x_j ;
- mass-weighted metallicity for mass formed in time interval Δt_j , Z_j ;
- dust attenuation for stars younger than 0.074 Gyrs, τ_{BC} ;
- dust attenuation for all stars, τ_{ISM} .

Given our choice for stacking method and the lack of overall normalisation, we cannot convert the mass fractions, x_j , into an absolute mass recovered in each time interval. This, however, does not pose a problem as we aim to compute a set of *colour* evolution tracks, for which an overall normalisation is not needed. For any given galaxy within a cell, the overall normalisation is given by its redshift and apparent magnitudes.

3.3 The age grid of VESPA

In its original configuration, the VESPA age grid runs between 0.002 and 14 Gyr, in logarithmically spaced time bins. This, however, means that the width of the final bin is around 5 Gyr. This in turn is comparable to the stretch of lookback time that the LRG sample probes (roughly between $z=0.5$ and $z=0.1$) and we therefore have very little sensitivity to the exact formation epoch of LRGs. The original configuration was so designed because typical SDSS data for individual objects does not normally allow us to distinguish between populations that are separated in age by less than the bin widths.

The large SNR obtained from stacking, however, opens up the possibility to use a revised grid that has more sensitivity to older populations, at the expense of losing time

resolution at the young end (the number of bins and overall structure must remain the same, for reasons constrained by the algorithm). We therefore construct an alternative age grid that we use to analyse the LRGs with VESPA. The boundaries of this new grid are: 0.002, 0.074, 0.177, 0.275, 0.425, 0.658, 1.02, 1.57, 2.44, 3.78, 5.84, 7.44, 8.24, 9.04, 10.28, 11.52 and 13.8 Gyr. Whereas these seem ambitious, statistically the data should be able to differentiate between models in adjacent bins: the differences between models of adjacent ages (normalised to a common value at a given wavelength) are several times that of the statistical noise in the stacked data. In practice, however, systematic errors can dominate - see the next Section for details.

Finally note that given the high SNR of the stacked spectra, VESPA always runs to its highest time-resolution.

3.4 Error analysis

VESPA was designed to deal with limitations coming from photon-noise. However, we are purposely putting ourselves in the limit where the photon-noise is negligible, and the quality of the fits is dominated by limitations in the modelling. We do not have a set of models that give statistically good fits in the limit of high-SNR and this has two potential effects: i) an error budget dominated by systematics in the modelling inaccuracies; and ii) the recovered parameters are biased.

To deal with the model inaccuracies, we make the assumption that if the models (SSPs, dust and parametrization) could represent galaxies perfectly, then the recovered solution would match the stacked spectrum exactly. Under this assumption, the residuals from a fit in each cell can be interpreted as an overall error including statistical photon-noise errors and systematic model errors.

(i) we run VESPA on the stacked spectrum and errors given by equations (2), and calculate the residuals $r_\lambda = |\hat{F}_\lambda - F_\lambda|$;

(ii) we smooth r_λ with a box-car filter to keep the large-scale variation but attenuate pixel-to-pixel fluctuations;

(iii) we take the smoothed r_λ as the new error, σ_λ^{total} and re-fit the stacked spectrum using this new error.

Fig. 3 shows a typical example of the residuals and the new estimated error. Therefore we allow extra freedom in the regions where the models cannot fit the data, by assuming that the residual between data and best-fit model gives an indication of the 1-sigma confidence region for systematic problems with the models, combined with the statistical error. As covariances will vary between different models, we assume this systematic is uncorrelated for flux values at different wavelengths, to avoid placing further biases on acceptable solutions. This estimate of the combined error replaces the instrumental statistical error in all of our fits. Our estimated systematic error will almost certainly be too small, because the best-fit parameters were chosen to match the data, and minimise this difference. Fundamentally, this is a problem that cannot be solved without better models. However the methodology outlined in this section should at least provide a approximation for the error and will be better than not allowing for this problem.

We translate this error into an error bar on the recovered physical parameters by constructing N random reali-

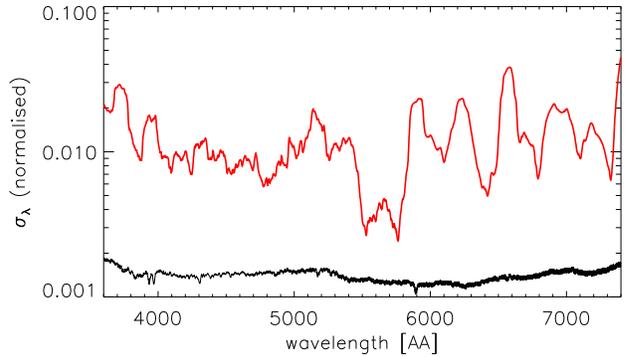


Figure 3. Estimating a new error vector. Red: residuals smoothed with a box-car filter; black: photon-noise errors from the stack given by Eq. (2). The new estimated error larger than the stack error by around a factor of 10.

sations of the best-fit spectrum, using the error calculated as per this section, and analysing these random realisations using VESPA. From this set of N solutions we can estimate the variance and co-variance of all the parameters of interest. More details are given in Tojeiro et al. (2007, 2009).

While these residuals provide a relative comparison of our current ability to model galaxies by comparing solutions obtained with different sets of models (SSPs, dust, etc), we must be careful not to interpret any variation in the recovered parameters as an absolute error bar - different models can be wrong in the same way. We analyse these differences in Section 6.

Finally, we should also note that the difference between models in adjacent bins (see Section 3.3) can be of the order of - or smaller than - this estimate of the systematic error. Ignoring the wavelength dependence, the average systematic error would correspond to roughly a 1 Gyr error on an 8 Gyr population.

3.5 Smoothing the fossil record

VESPA assumes that the star-formation rate (SFR) is constant within a time bin, and this gives results with discontinuities in SFR across bin boundaries. This is an obvious over-simplification of the problem, as any star formation history is likely to be continuous over an arbitrary set of boundaries, if sampled at high enough resolution. For the purposes of this paper, this over-simplification is problematic because a discontinuous SFR will result in a physically discontinuous colour evolution.

To solve this problem, once VESPA has produce model parameters, we replace the top-hat bins of constant SFR with Gaussian distributions. The widths and heights are chosen to be such that the mass formed within each pair of the old sharp boundaries does not change by more than 0.3%. The mass integrated over all ages is kept constant. These new Gaussian representations for each population (previously given by a single top-hat bin) are sampled at a much higher rate in lookback time, such that we can calculate a smooth colour evolution. Each new Gaussian bin has the same metallicity as the old corresponding histogram.

We do not expect this change in the SFR as a func-

tion of time to make a large difference in the interpretation of the spectra. This is because, by construction, the original boundaries are such that the spectral evolution between each boundary is not too significant. The offset between the spectra calculated top-hat and Gaussian bins is either smaller or of the order of the statistical error shown in Fig. 3. The flux obtained with the Gaussian bins is only used to compute the observed colours and their evolution with redshift, and the effect of this difference in the computed colours is negligible.

3.6 Evolving the observed spectrum

Our goal is to compute the rest-frame spectrum of each cell (representing an ensemble of galaxies) at any point $t_e > t_g$, where t_g is the lookback time of the cell in question. We compute

$$F_\lambda(t_e) = \sum_{t_k < t_e} m_k f_\lambda(t_k - t_e, Z_k), \quad (9)$$

where $f(t, Z)$ is the flux of a population of age t and metallicity Z , and $m_k = g(t_k)\Delta t_k$ is the mass formed at t_k , assuming a SFR given by $g(t)$. For the case where $t_e = t_g$ then we simply recover the observed spectrum, as shown in the section above.

From $F_\lambda(t_e)$ we can easily compute any observed frame colours, by applying the relevant K+e corrections (fully known). As noted before, we cannot compute apparent magnitudes due to a lack of normalisation. However, for any individual galaxy the normalisation is set by the observed apparent magnitude and redshift, and $F_\lambda(t_e)$ gives then not only the colour, but also the luminosity evolution of that galaxy, assuming that the SFR in the galaxies follows the weighted average within the cell.

3.7 Modelling dust

Dust evolution does not leave a footprint in a galaxy's spectrum today, so it does not affect our modelling. However, in order to compare the colour tracks presented in Section 5.3.2 to observed populations of galaxies at different redshifts, we do need to treat the evolution of dust. Note that for the two-parameter dust model of Charlot & Fall (2000), in addition to the inter-stellar dust component, young stars are allowed extra dust extinction to account for the effects of their birth cloud, which takes some time to dissipate, of the order t_{BC} . Both of these components are expected to evolve with redshift. We can envisage three possible solutions to this problem:

- (i) we can assume that the values we recover for τ_{BC} and τ_{ISM} remain valid for the entire history of the galaxy. I.e., when t_e is such that a population becomes younger than t_{BC} then we apply the value of τ_{BC} obtained at t_g , and we always apply the same value of τ_{ISM} to all population; and
- (ii) we can estimate the evolution of τ_{BC} and τ_{ISM} from values observed for particular galaxies as a function of redshift, and apply these accordingly.
- (iii) we can do the comparison in a dust-extinction corrected colour-colour or colour-redshift plane, in which case no dust evolution should be applied to the colour tracks.

There are advantages and disadvantages to all methods. In the case of (i), we are over-simplifying the problem by ignoring a potential evolution with redshift, but we are keeping the information that relates to each individual cell. In the case of (ii) we are using a model for the evolution of dust, but we are assuming that the galaxies that we observe at low- and high-redshift are part of the same population. This latter assumption is exactly what we are looking to avoid by tapping into the fossil record. Option (iii) is probably the most robust, but requires modelling of galaxies at high redshifts. In this paper we plot the colour-tracks over *observed* colours, and for simplicity we apply the correction as given by (i) above. For other applications, colour evolution can be supplied with or without dust evolution.

3.8 Summary of outputs

It is worth summarising what the outputs of the analysis described in this section are. They can be broadly grouped into two types. On one hand we have *measured* quantities that relate to the position of the cell at a given redshift and colour, e.g. dust information (τ_{ISM}, τ_{BC}), mass-weighted metallicities or mass-weighted ages. On another hand we have the colour-tracks as a function of redshift, which are inferred *evolved* quantities. All of these results have a dependence on the SSP synthesis codes. We describe the SSP models we use in the next section.

4 MODELS

VESPA models the spectrum of a galaxy as a superposition of simple stellar populations (SSPs) of different ages and metallicities. At the heart of this process lies the assumption that we know the spectral signature of different SSPs, and that we sample the full parameter space needed in order to model observed galaxies.

There are three key steps that go into constructing an SSP. Firstly we need a description of a star's evolution given its mass and metallicity, in terms of observable parameters such as effective temperature or bolometric luminosity. This is traditionally given by the isochrones. Several groups have published sets of isochrones, which aim to model all stages of stellar evolution. Secondly, we must assume an initial mass function (IMF), which gives the correct weight as a function of mass, for stars formed in a single cloud of gas. Stars of different mass evolve with different time-scales, so by combining the isochrones and the IMF one can correctly populate a colour-magnitude diagram across the different stages of stellar evolution. The final step is to assign a spectrum to stars with different parameters, and this is done using the spectral libraries. Spectral libraries can be empirical or synthetic, with the former suffering from poor sampling of the parameter space of stellar evolution (dictated by the chemical enrichment of our own solar neighbourhood), and the latter suffering from deficiencies in our current ability to model stellar evolution.

Crucially, the choices, treatments or calibrations involved in all of these three steps differ across different SPS codes, leading to unavoidable differences in the spectrum of SSPs of fixed age and metallicity. These differences are naturally centred around the least understood

stages of stellar evolution, such as the thermally-pulsating asymptotic giant branch (TP-AGB) phase, the supergiant phase, or horizontal branch stars, and the datasets used for internal calibrations. The first efforts to accurately estimate the effects of the Horizontal Branch, Red Giant and TP-AGB branch on SSPs came from Jimenez et al. (1998); Maraston (1998); Jimenez et al. (2004); Maraston (2005). Jimenez et al. (1998, 2004) built an analytic model to describe this stages of stellar evolution paying careful attention to use observations of individual and resolved stars to calibrate the stellar evolution models (see Jimenez et al. 1995). They pointed out how overestimations of the TP-AGB and lack of morphological description of the HB compromised seriously conclusions drawn from SSPs. They were, also, the first ones to emphasize the importance of stellar interior tracks being computed self-consistently with the stellar atmospheres models (Jimenez et al. 1998, 2004). Several authors have since followed this approach to estimate the broad-band SSPs (see e.g. Conroy et al. 2009 among others).

Conroy et al. (2009), for example, look at the impact of these choices and assumptions on the recovered physical properties of galaxies, using broadband colours, with the aim of estimating the systematic error that arises from the limitations of the models. Here we do not take such a systematic approach, but we do conduct our analysis using three sets of popular SPS codes and investigate the resulting differences. The choices of IMF, stellar libraries or method for computing the stellar energetics cannot always be matched across different SPS codes. In each case we take the combination of these that best suits our application. We briefly describe the three sets of models next.

4.1 BC03

With the BC03 models (Bruzual & Charlot 2003) we adopt a Chabrier initial mass function (Chabrier 2003) and Padova 1994 evolutionary tracks (Alongi et al. 1993; Bressan et al. 1993; Fagotto et al. 1994a,b; Girardi et al. 1996). We use the BC03 models with the empirical STELIB library (Le Borgne et al. 2003) in the optical (3200Å to 9500Å) and the theoretical BaSeL 3.1 stellar library (Lejeune et al. 1997, 1998; Westera et al. 2002) to either side of that range.

4.2 M10

The M10 models (Maraston & Stromback, submitted) are the new, high-resolution version of the M05 models (Maraston 2005). The energetics and stellar evolution calculations are unchanged from the models in Maraston (1998) and M05, so the post main-sequence continues to be treated with the fuel-consumption theorem (Renzini & Buzzoni 1986; Buzzoni 1989; Maraston 1998), leading to a more robust treatment of advanced stages of stellar evolution, such as the TP-AGB phase. The M10 models and publication focus not only on delivering a high-resolution version of the M05 models, but also on exploiting differences arising from stellar libraries and their calibration. Therefore, models are provided in a range of stellar libraries, each with different coverage in age, metallicity and wavelength. Here we choose the MILES library (Sánchez-Blázquez et al. 2006),

as it provides the best coverage in age and metallicity. The wavelength coverage was extended to the UV following the method in Maraston et al. (2009). We adopt a Kroupa IMF (Kroupa 2001).

4.3 FSPS

The Flexible Stellar Population Synthesis (FSPS) code (Conroy et al. 2009; Conroy & Gunn 2009) takes a novel approach to the computation of SSP spectra. By parametrizing uncertain stages in stellar evolution and allowing these parameters to change freely, the authors attempt to both quantify our ignorance about stellar evolution in terms of derived galaxy properties, but also to calibrate their models by finding the combination of parameters that best describe certain observations. Furthermore, models are provided with a choice of isochrones and stellar library. Here we chose the latest Padova evolutionary tracks calculations (Marigo & Girardi 2007; Marigo et al. 2008), and the MILES stellar library. An UV extension was obtained by using the theoretical BaSeL spectral library. In spite of the freedom provided by these models, here we use them simply at their default settings and with the combination of parameters that describe the TP-AGB phase that were found to best fit star cluster data by Conroy & Gunn (2009). The opportunity remains to explore further the flexibility of these models - we leave that for a follow up paper. We use a Chabrier (2003) IMF.

4.4 α - enhancement

The metallicities implicit in the SPS models and quoted throughout this paper refer to [Fe/H] abundances. However, it is now observationally well established that ETGs and bulges have a larger abundance of the so-called α -elements (O, Ne, Mg, Si, S, Ca, Ti) than that predicted by SPS models, given the measured Fe abundances (e.g. Worthey et al. 1992; Davies et al. 1993; Fisher et al. 1995; Longhetti et al. 2000; Maraston et al. 2003; Thomas et al. 2003, 2005 and references therein).

α -elements are mostly produced through the collision of α -particle nuclei in type-II supernovae of very massive stars, whereas Fe-peak elements are associated with type-Ia supernovae, typically associated with older and lower mass stars (e.g. Woosley & Weaver 1995). Given the different delay times for the two types of stellar explosions, different [α /Fe] abundances have therefore been interpreted as a signature for the length of time over which the star formation happened (e.g. Pagel & Tautvaisiene 1995; Thomas et al. 2005). The mismatch with the Fe abundances predicted by SPS models therefore occurs because empirical stellar libraries are primarily constructed by stars in the solar neighbourhood, which has a different chemical enrichment and star formation history than that which is typical of bulges and massive elliptical galaxies.

Whereas some effort has gone into calibrating certain spectral indices for different α -element abundances (e.g. Thomas et al. 2003), this has not yet been done for the full spectrum. Different chemical abundances are therefore a known limitation of current SPS models based on empirical libraries, if one wants to use a full-spectral fitting

technique. Theoretical libraries however, can provide some insight on how different chemical abundances ratios affect galactic spectra (e.g. Coelho et al. 2007). For the moment, such discrepancies (surely to be systematic when looking at a population such as the LRGs) are automatically bundled with other unknown systematic errors in Section 3.4.

5 RESULTS

The main product of this work is a set of colour evolutionary tracks as a function of redshift and colour. In conjunction with high redshift data to match to, these can be used to solve some galaxy evolution cases, and we will consider such an analysis in a companion paper. In this paper we now explore some of the most immediate results, such as dependence of dust, age and metallicity on colour and redshift.

The dependence on the underlying SSP modelling is investigated in Section 6. In this section, for clarity, we will often show results by choosing one of the three sets of SPS sets of models, and we append the matching plots using the other two sets at the end of this paper. Our choice of SSP model in the main text in each case should not be interpreted as supporting one over the others. When practical, we will show results from all three sets of models alongside each other.

We start by showing some a typical fit and residuals.

5.1 Typical fit and residuals

In Fig. 4 we show a fit for one of the stacks, and in Fig. 5 the corresponding residuals in units of the stack noise, as given by Eq. (2). This fit was obtained using the FSPS models (results for the same stack, with Mastro and BC03 models can be found in Appendix A). In this case we can see a visually good match on the blue end, and an increasing tension between data and models towards the red. The detail of the residuals plot reveals that there are significant departures from the data at all wavelengths, but more notoriously blueward of 3800Å and redwards of 5800Å .

Fig. 5 also shows regions of the spectrum that were excluded from fits that give the results presented in this section. Inclusion or exclusion of these regions does not affect the results in this section in any significant way.

5.2 Measured quantities

In this section we show how ages, dust and metallicity vary with colour, luminosity and redshift in our LRG samples. The luminosity ranges refer to those shown in Fig. 1.

When we show or quote errors in this Section, they refer to variations in colour, for a given redshift bin and luminosity slice. These variations are typically larger than the systematic error obtained as per Section 3.4, which is likely to be an underestimation of the true systematic error. The effect of the statistical error on the recovered solutions is negligible.

A summary of the results in this section can be found in Fig. 6.

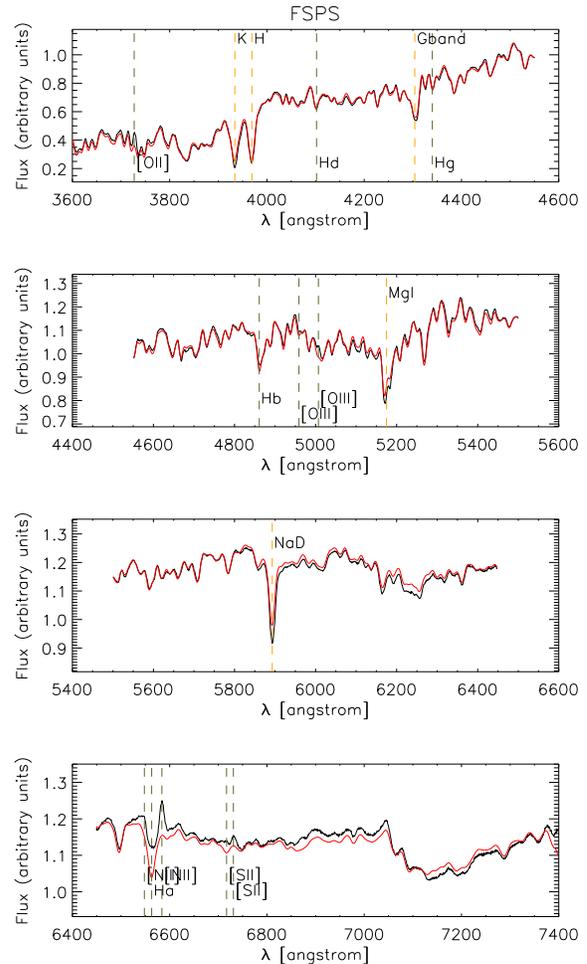


Figure 4. A typical fit, using the FSPS models (equivalent plots, for Mastro and BC03 models can be found in Appendix A1). The black line is the data, and the red line the best-fitting model. The vertical yellow and green dashed lines guide the eye by showing some common emission and absorption features. Although plotted, some regions of the spectrum were excluded from the fit - see Fig. 5

5.2.1 The ages of LRGs

The fiducial model is that LRGs formed mostly in one short epoch of star-formation at high-redshift. Therefore, the age of the oldest stars in LRGs of different redshifts should be consistent with a single epoch of formation. This is in principle testable and our revised age grid is sufficiently fine to test this hypothesis (see Section 3.3).

In contrast, we find best-fit solutions given by the models that are systematically older than the age of the Universe (Komatsu et al. 2009). Moreover, we found this behaviour with all three SPS models. We find that the best-fit solution is still too old in many cases, even after adjusting the error in the flux according to the method in Section 3.4 This suggests that the behaviour is systematic rather than statistically random. We therefore decide to impose a strong prior on possible solutions, such that star-formation cannot happen in a bin whose youngest boundary has an age older

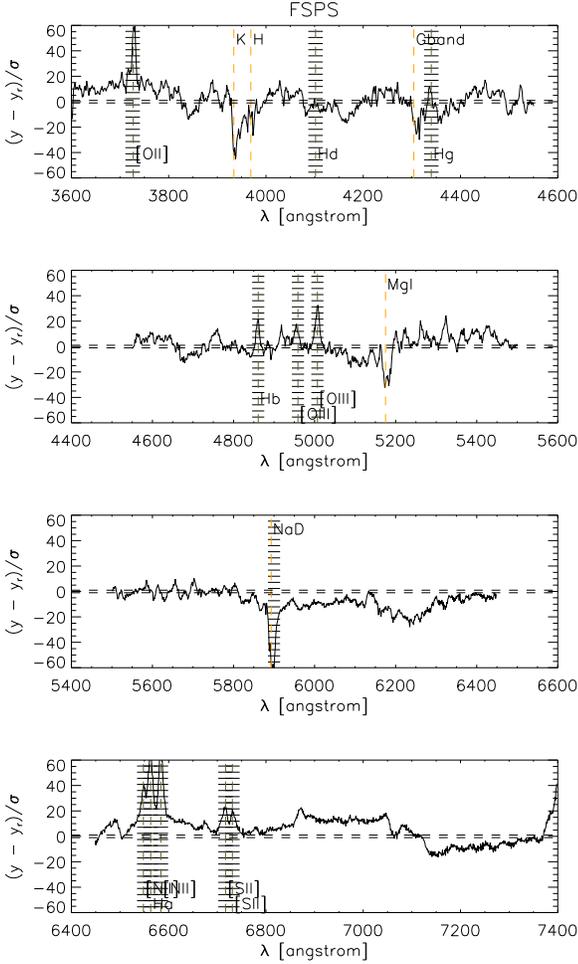


Figure 5. The residuals in units of the stack noise, for the fit showed in Fig. 4. The vertical dashed lines guide the eye by showing some common emission and absorption features. The regions shaded by horizontal lines were excluded from the fit.

than the age of Universe. All results from hereon have this prior with the age of Universe set to be 13.8 Gyrs at $z=0$. The best-fitting solutions obtained with this prior have a formally worse χ^2 than those without the prior, as expected. There is important information in the residuals of the solutions obtained with and without this prior - one is clearly wrong, but provides a better fit. The difference between the two sets of residuals should highlight which spectral features yield this difference. We leave this analysis to a companion paper.

Fig. 7 shows the dependence of the rest-frame mass-weighted age of each stack of galaxies, as a function of colour and redshift for four luminosity slices, and calculated using the M10 models (identical plots obtained with BC03 and FSPS are given in Appendix A2). We calculate a mass-weighted age as

$$\langle t_{age} \rangle = \frac{\sum_i t_i x_i}{\sum_i x_i} \quad (10)$$

where we take t_i as being the mean age of bin i . We see no strong evidence for a dependence of $\langle t_{age} \rangle$ on colour,

and we see the expected ageing of galaxies towards lower redshifts (largely dictated by our prior on the age of the Universe).

In Fig. 6 (3rd row) we collapse the histogram in colour, and show the evolution in redshift obtained with different SPS models. The error bars show the error on the mean over colour, and so are representative of the variation with colour, in a given redshift bin. In the 4th row of Fig. 6 we add the lookback time to the redshift of each stack, and show the formation time in the Earth-frame.

5.2.2 Recent to intermediate star formation

Collapsing the full SFH into a mass-weighted age, however, can mask interesting information such as the amount of recent star formation. In Fig. 8 and, collapsed in colour, in the 5th panel of Fig. 6, we show the fraction of star-formation, by mass, recovered in bins up to 3.8 Gyr in the rest-frame of each stack. A summary is given in Table 1, where we split this range further in intermediate and young ages.

Although a fixed age in the rest frame is equivalent to a different fraction of the galaxies' ages as we go back in redshift, 3.8 Gyr serves as a good split between stars that formed in the oldest, fiducial burst, and anything that may have followed. We discuss these results in Section 6.

5.2.3 Metallicity

Fig. 9 shows the mass-weighted metallicity as a function of $r-i$ colour and redshift, for four luminosity slices, and calculated using the Mastro models (identical plots obtained with BC03 and FSPS are given in Appendix A). For each cell, we calculate the mass-weighted metallicity using the full star-formation history as

$$\langle Z \rangle = \frac{\sum_i Z_i x_i}{\sum_i x_i}. \quad (11)$$

To make any possible trend with redshift more clear, we average over colour and show the resulting redshift dependence in Fig. 6 (first row). The horizontal dashed lines show the metallicities provided by the models - note that different sets of models sample the metallicity space in different ways. We discuss the different behaviour of each set of models in Section 6. As in the previous section, the error bars are shown to show the error on the mean over colour.

5.2.4 Dust

Fig. 10 shows the dust attenuation as a function of $r-i$ colour and redshift, for four luminosity slices. To make any possible trend with redshift more clear, we average over colour and show the resulting redshift dependence in Fig. 6 (second row). The error bars are errors on the mean, and therefore give an indication of the scatter of dust with colour, for each redshift bin. We discuss these results in Section 6.

5.3 Evolved quantities

We can use the full star-formation history of each cell to estimate the colour evolution of a galaxy within that cell. This

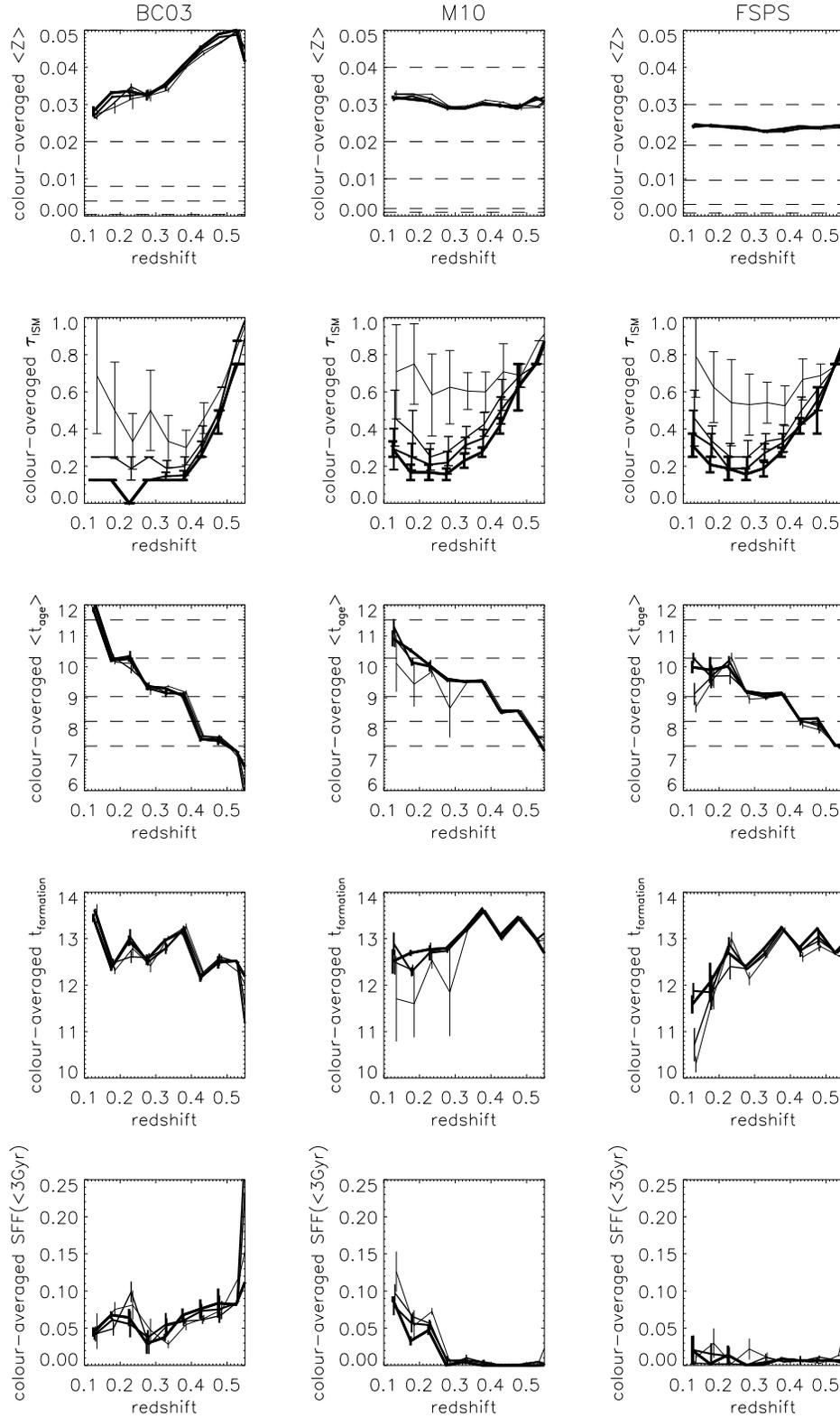


Figure 6. The evolution of mass-weighted metallicity (first row), dust extinction (second row), mass-weighted age (third row), formation epoch (forth row) and recent star-formation fraction (fifth row) with redshift for stacks of different luminosities. The thickness of the line in all plots represents the luminosity of the galaxies in each stack (thinner for the faintest, and thickest for the brightest). At each redshift, the quantities are averaged over $r - i$ colour. The horizontal dashed lines in the first row represent the metallicities provided by each set of models. Extrapolation beyond the extreme values is not permitted. The horizontal dashed lines in the 3rd row show the age boundaries used by VESPA. The error bars are the error on the mean over colour, and are representative of the variation with colour in a given redshift bin. See text for more details.

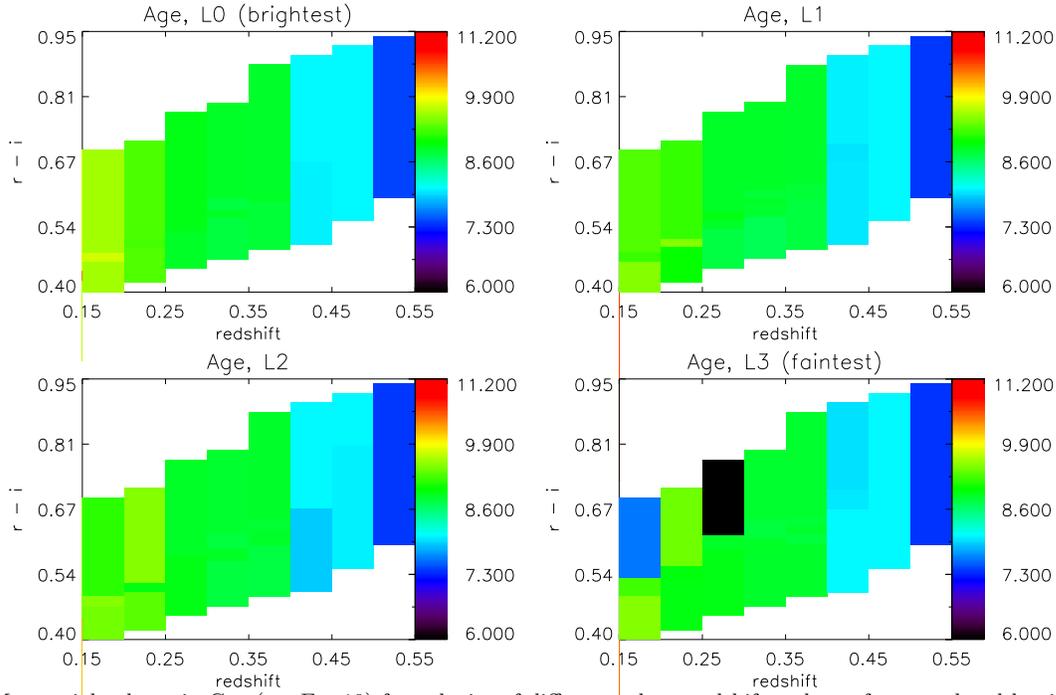


Figure 7. Mass-weighted age in Gyr (see Eq. 10) for galaxies of different colour, redshift and rest-frame r -band luminosity, analysed with the M10 models. The luminosity ranges change with redshift, and are shown in Fig. 1. We only show data for regions of parameter space with a sufficient number density of galaxies. In practice, there are a few galaxies (less than 20 per redshift bin) outside of the coloured areas, but these have an insignificant impact on the recovered solution. In Fig. 6 (3rd row) we average over colour to show the trend with redshift for each luminosity slice.

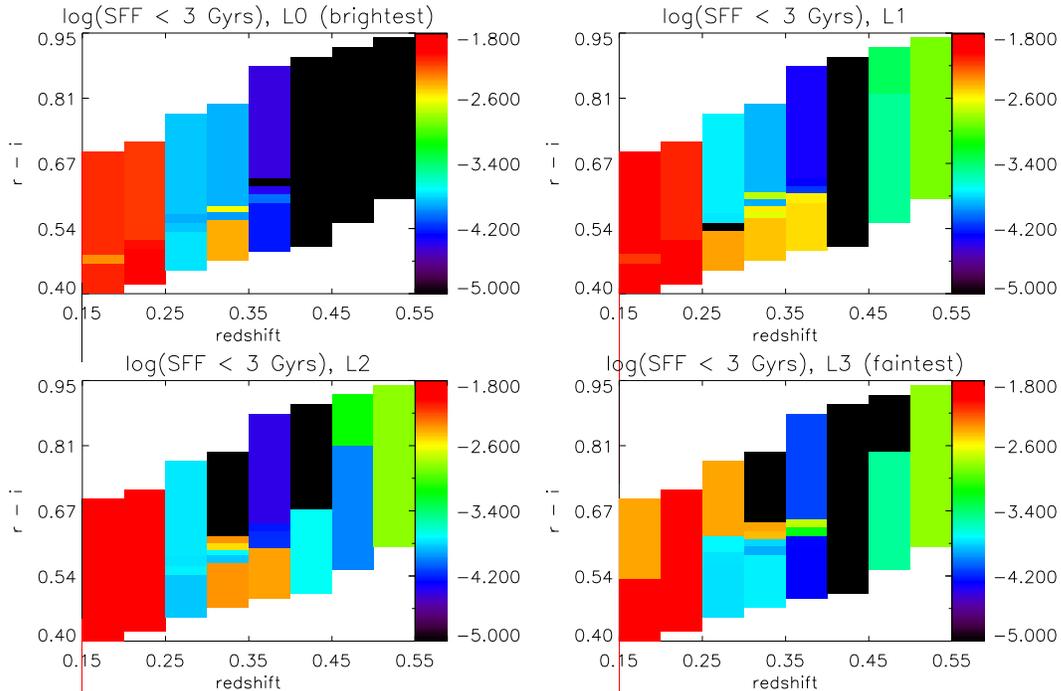


Figure 8. The logarithm of the fraction of star formation (by mass), recovered in bins up to 3.8 Gyr for galaxies of different colour, redshift and rest-frame r -band luminosity, analysed with the M10 models. The luminosity ranges change with redshift, and are shown in Fig. 1. We only show data for regions of parameter space with a sufficient number density of galaxies. In practice, there are a few galaxies (less than 20 per redshift bin) outside of the coloured areas, but these have an insignificant impact on the recovered solution. In Fig. 6 (5th row) we average over colour to show the trend with redshift for each luminosity slice.

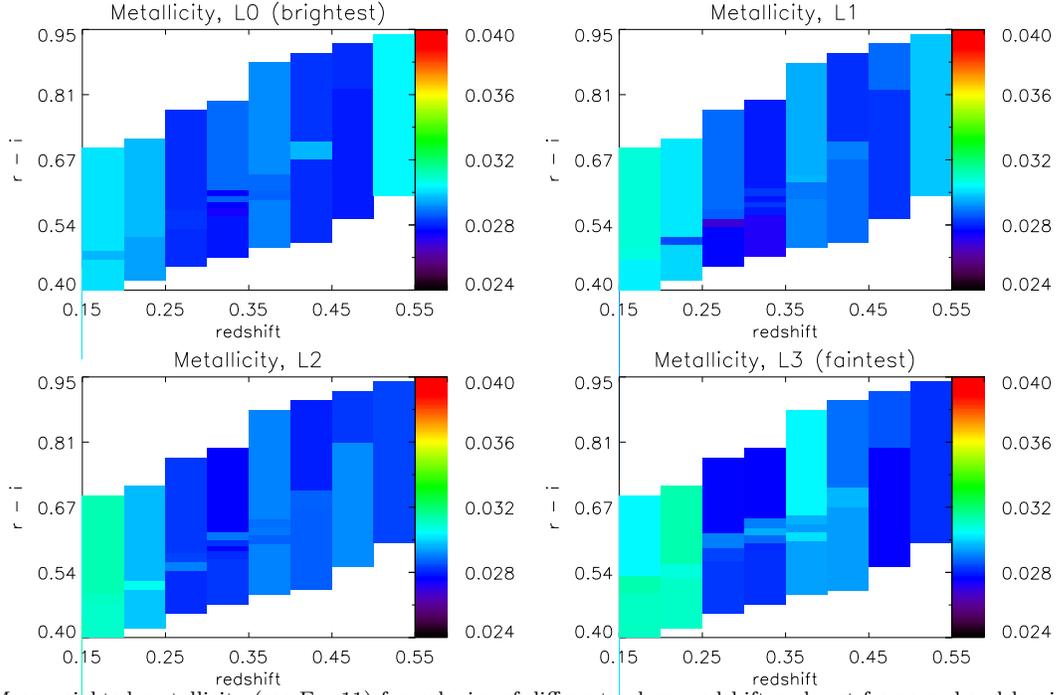


Figure 9. Mass-weighted metallicity (see Eq. 11) for galaxies of different colour, redshift and rest-frame r -band luminosity, analysed with the M10 models. The luminosity ranges change with redshift, and are shown in Fig. 1. We only show data for regions of parameter space with a sufficient number density of galaxies. In practice, there are a few galaxies (less than 20 per redshift bin) outside of the coloured areas, but these have an insignificant impact on the recovered solution. Fig. 6 averages over colour to show the trend with redshift for each luminosity slice.

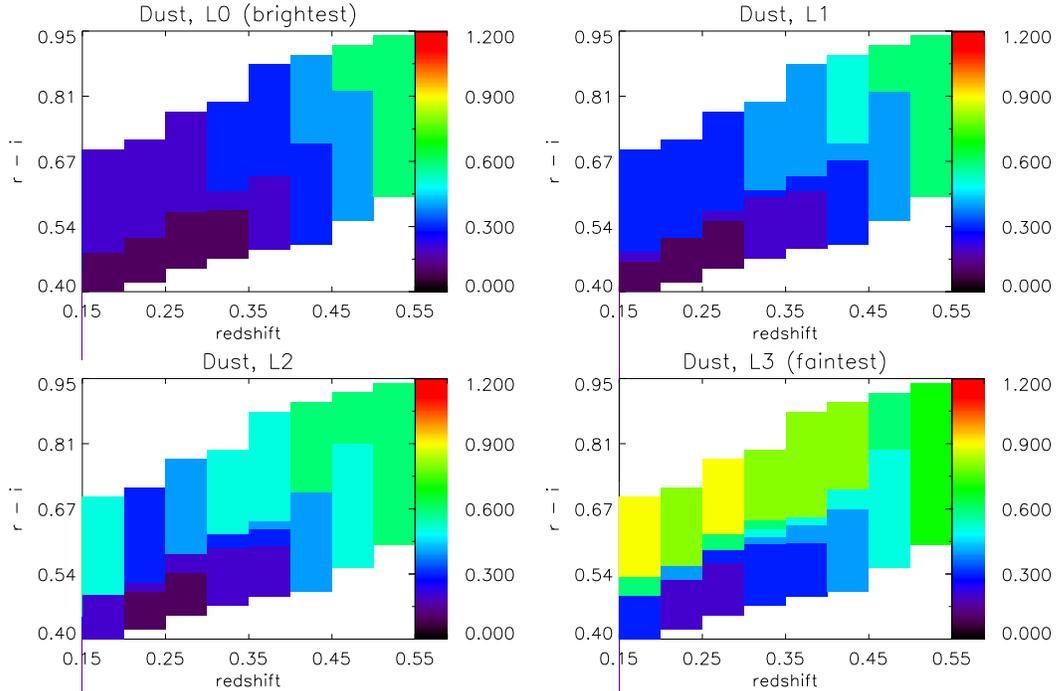


Figure 10. Inter-stellar dust absorption for galaxies of different colour, redshift and rest-frame r -band luminosity, analysed with the M10 models. The luminosity ranges change with redshift, and are shown in Fig. 1. We only show data for regions of parameter space with a sufficient number density of galaxies. In practice, there are a few galaxies (less than 20 per redshift bin) outside of the coloured areas, but these have an insignificant impact on the recovered solution. Fig. 6 averages over colour to show the trend with redshift for each luminosity slice.

SPS model	Luminosity range	0.15 < z < 0.35		0.35 < z < 0.5	
		Young SFF < 0.7 Gyr	Intermediate SFF 0.7 - 3.8 Gyr	Young SFF < 0.7 Gyr	Intermediate SFF 0.7 - 3.8 Gyr
BC03	L0 (brightest)	0.0018±0.00029	0.047±0.0085	0.0068±0.00024	0.068±0.0068
	L1	0.0019±0.00028	0.048±0.011	0.013±0.00028	0.126±0.0057
	L2	0.0019±0.00033	0.054±0.011	0.010±0.00024	0.114±0.0023
	L3 (faintest)	0.0013±0.00023	0.056±0.015	0.013±0.00029	0.0776±0.0040
M10	L0 (brightest)	0.0018±0.00065	0.036±0.0033	0.00014±4e-06	0±0
	L1	0.0019±0.00069	0.043±0.0071	0.00015±3e-06	0.00280±7e-05
	L2	0.0015±0.00057	0.046±0.0051	0.00044±1e-06	0.0038±0.0001
	L3 (faintest)	0.00311±0.0022	0.056±0.0098	0.0009±2e-06	0.018±0.0004
FSPS	L0 (brightest)	0.0068±0.0066	0.0017±0.0007	0±0	0.0046±0.0013
	L1	0.0096±0.0094	0.0020±0.0002	0±0	0.0063±0.0015
	L2	0.0022±0.0020	0.0025±0.0007	0±0	0.0048±0.00076
	L3 (faintest)	0.013±0.0086	0.0076±1e-05	0±0	0.026±0.00078

Table 1. Summary of the recovered star-formation fractions at young and intermediate ages, averaged over two redshift ranges, for all luminosity slices and the three SPS models we consider. As in Fig. 6, the errors come from the variation with colour at a given redshift bin.

information can be coupled with the apparent magnitude and exact redshift of the galaxy to calculate its luminosity evolution. In this section we present a selection of these colour tracks for a sample of galaxies selected in colour, luminosity and redshift. As before, we present the results using only one set of SPS models in this section (FSPS), and provide the corresponding plots using the other two SPS models in Appendix A3.

5.3.1 Spectro and model magnitudes

The fibre aperture in the SDSS has a fixed size of 3 arc-seconds. Apparent sizes of objects are often larger than this, leading to an unavoidable discrepancy between the magnitudes obtained from integrating a spectrum over a filter’s response (the spectro magnitudes), and the best estimate of the photometric magnitudes (the cmodel magnitudes). We do our stacking using the latter, but we use the spectrum to obtain our star formation and metallicity histories, which in turn refer only to the 3 arc-second fibre. Therefore we need to be able to map one quantity to the other, and understand any possible biases in this relationship. The spectrophotometric calibration can also affect this mapping, and there is a known offset between the spectro and fibre magnitudes (obtained from the photometry, within the 3 arc-second aperture) since DR6 due to how this calibration is done (Adelman-McCarthy et al. 2008). This offset, however, has an insignificant dependence on photometric band and it is not a problem for our colour selection.

Fig. 11 shows the distribution of

$$\Delta[r - i] = [r - i]_{\text{spectro}} - [r - i]_{\text{cmodel}}, \quad (12)$$

for a random sample of 10,000 LRGs. This scatter can be better understood by explicitly writing

$$[r - i]_{\text{cmodel}} = [r - i]_{3''} + \delta_{\text{photo}} + [r - i]_{\text{extra}} \quad (13)$$

$$[r - i]_{\text{spectro}} = [r - i]_{3''} + \delta_{\text{spectro}}, \quad (14)$$

where $[r - i]_{\text{extra}}$ is the difference in colour due to a radial colour gradient extending outwith the 3 arc-second fibre, and

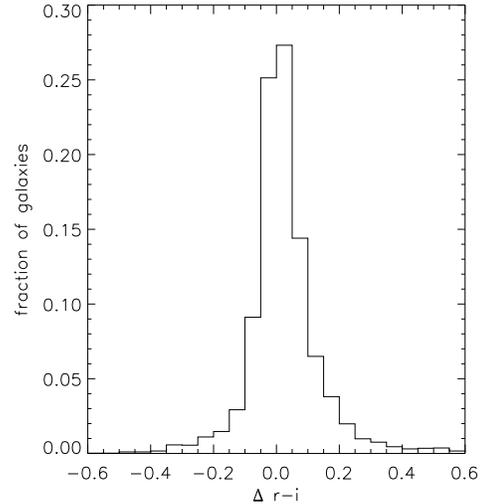


Figure 11. The distribution of $\Delta r - i$ (Eq. 12) for a random sample of 10,000 LRGs.

δ_{spectro} and δ_{photo} are noise in the spectroscopic and photometric measurements, respectively. Assuming that δ_{spectro} and δ_{photo} are stochastic components, an offset from zero in $\Delta[r - i]$ would be indicative of a systematic behaviour in $[r - i]_{\text{extra}}$. Fig. 11 shows no evidence for such an offset. Fig. 12 shows $\Delta[r - i]$ as a function of redshift - there is an indication of a positive slope at high-redshift. A positive $\Delta[r - i]$ means a larger spectro colour, which in turn means a bluer colour inside the fibre, when compared to the full aperture. This is counter-intuitive for two reasons - firstly one would expect the proportion of light falling into the 3 arc-second fibre to go up with redshift, as apparent sizes get smaller; and secondly, one would expect the outer regions to be bluer in colour in comparison to the central regions.

Note, however, that at high redshift we are intrinsically selecting increasingly redder galaxies (in cmodel colours)

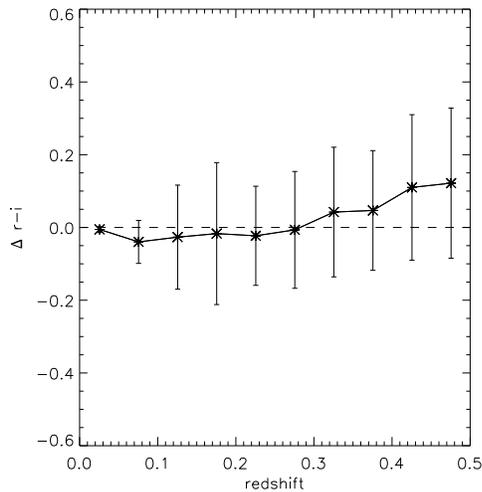


Figure 12. The mean and scatter of $\Delta r - i$ as a function of redshift. See text for an explanation of the upturn at high redshift.

due to the LRG sample selection. These, in turn, are likely to have the largest positive values of δ_{photo} and, in consequence, of $\Delta[r - i]$. There is an intrinsic Malquist-type selection bias in $\Delta[r - i]$ simply due to the fact that we select in $[r - i]_{\text{cmodel}}$. The selection of galaxies for each cell will be biased in a similar way, but this bias is known as we can compute $\Delta[r - i]$ for each cell.

Without independent photometry, it is hard to disentangle the redshift-dependence of this selection bias from a true dependence of $[r - i]_{\text{extra}}$ in colour, or redshift. However, Figs. 11 and 12 show no evidence of a detectable signal in $[r - i]_{\text{extra}}$. Therefore, to translate from spectro colour to cmodel colour we simply apply a correction to the beginning of each track. We do this in all of the plots in the next section.

5.3.2 Colour tracks

Figs. 13 and 14 shows the predicted evolution of $r - i$ and $g - r$ colour with redshift, for four combinations of colour and luminosity and obtained with the FSPS models. Fig. 15 shows this evolution in the $r - i$ vs $g - r$ plane, using the same models. In each case we over-plot with contour lines the number of *observed* LRGs for in the same luminosity range. The contours are wider for the fainter sample, but we cannot immediately tell whether this is due to an increased photometric error for fainter objects, or due to any intrinsic colour-luminosity relation. We explore these results and their model dependence in Section 6.

6 MODEL DEPENDENCE AND INTERPRETATION

The results presented in the previous section cannot be interpreted without addressing the fact that each set of models often gives a different result. In this paper we do not directly tackle the question of which set of models is the most cor-

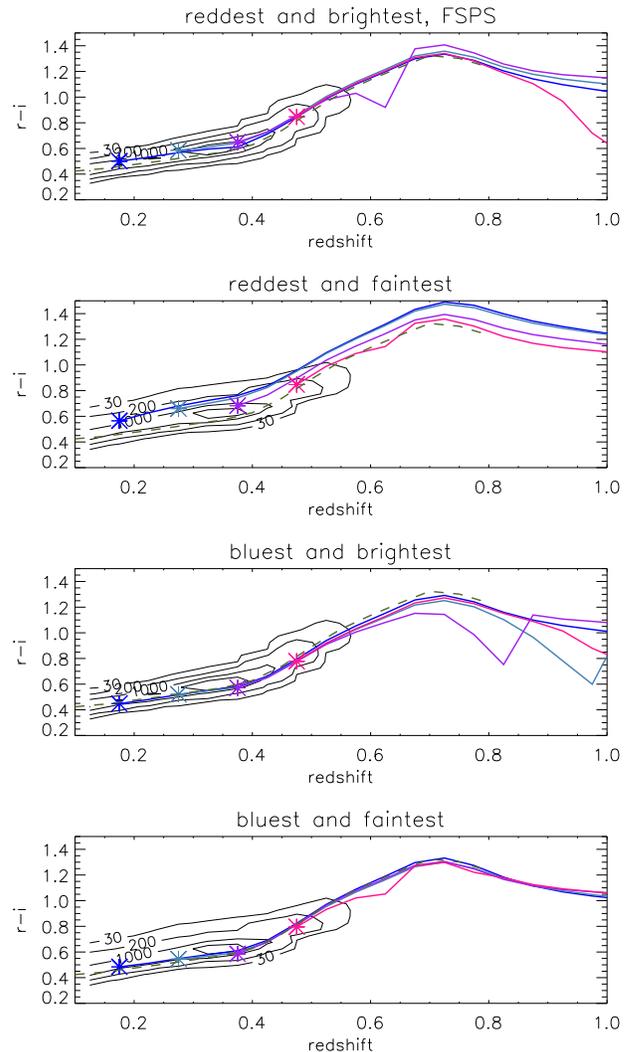


Figure 13. Predicted evolution of $r - i$ colour with redshift, calculated according to Eq. (9), and the solutions presented in Section 5.2. The different panels show examples at different combinations of colour and luminosity. The black contour lines show the number density of LRGs within the respective luminosity slice, and the coloured lines show the tracks from cells at four different redshifts: 0.175, 0.275, 0.375 and 0.475. The star shows the position of the cell the track relates to. The green dashed line shows the LRG model from Maraston et al. (2009), for comparison.

rect, we limit ourselves to a very basic exploration of the residuals in Section 6.3.

Overall, we find agreement with the fiducial model that says that LRGs are dominated by old, metal rich stars. However, we find that some small departure of this model is needed to fit the stacked spectra, and that small order changes to the fiducial model depend on the SPS model used. Ultimately, we are interested in seeing how the stellar evolution of LRGs translates into colour tracks, and the potential implication for galaxy evolution and observational cosmology. We begin by analysing the in-situ results of Section 5.2 and then we discuss the effects on the colour tracks of Section 5.3.

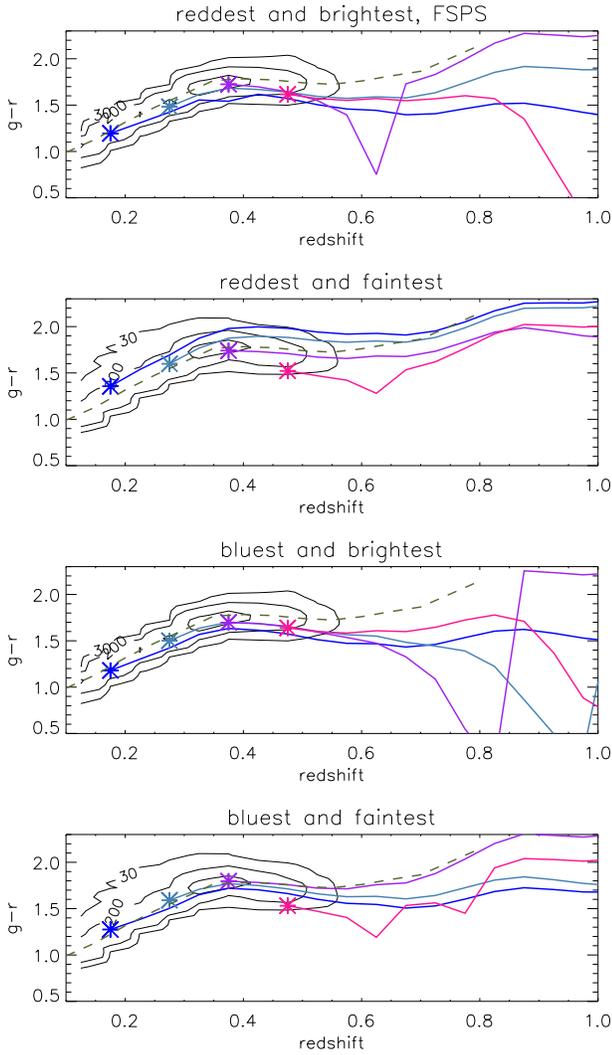


Figure 14. Predicted evolution of $g-r$ colour with redshift, calculated according to Eq. (9), and the solutions presented in Section 5.2. The different panels show examples at different combinations of colour and luminosity. The black contour lines show the number density of LRGs within the respective luminosity slice, and the coloured lines show the tracks from cells at four different redshifts: 0.175, 0.275, 0.375 and 0.475. The star shows the position of the cell the track relates to. The green dashed line shows the LRG model from Maraston et al. (2009), for comparison.

6.1 Measured quantities

Fig. 6 summarises the in-situ measurements presented in Section 5.2.

We see a different behaviour in the metallicity evolution when using different models, but all models consistently show LRGs to be metal-rich, as expected following the fiducial model. This difference is not a surprise - metallicity is known to be more model dependent than age (e.g. Panter et al. 2008; Tojeiro et al. 2009), and even the simple fact that the models sample the metallicities at different intervals will have some effect.

The sharp increase in metallicity seen when using BC03 models is most likely due to the age-metallicity degeneracy.

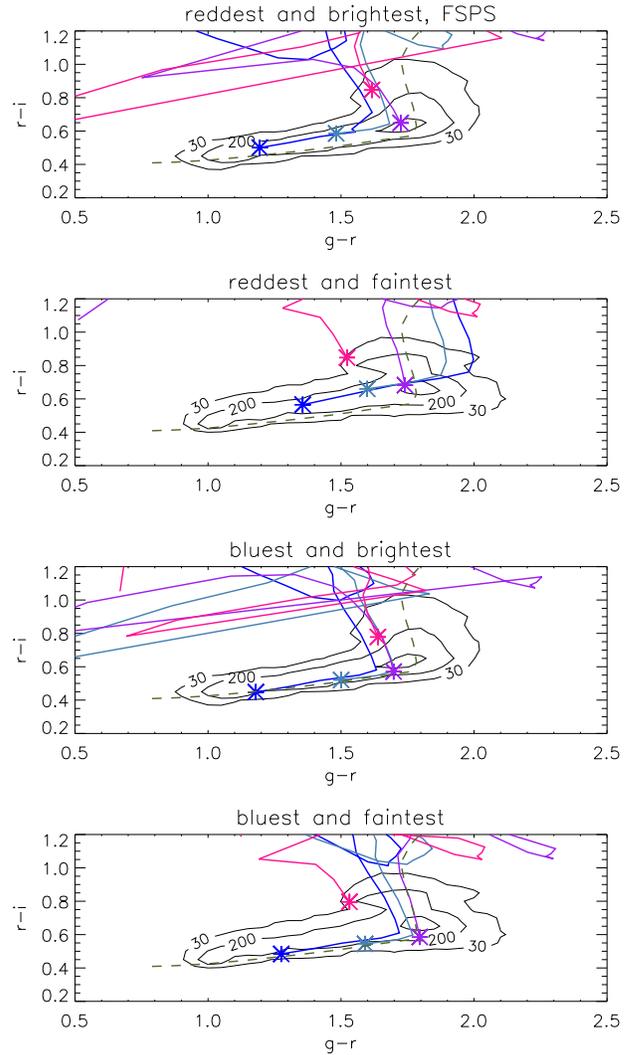


Figure 15. Predicted evolution of $r-i$ colour with $g-r$, calculated according to Eq. (9), and the solutions presented in Section 5.2. The different panels show examples at different combinations of colour and luminosity. The black contour lines show the number density of LRGs within the respective luminosity slice, and the coloured lines show the tracks from cells at four different redshifts: 0.175, 0.275, 0.375 and 0.475. The star shows the position of the cell the track relates to. The green dashed line shows the LRG model from Maraston et al. (2009), for comparison.

BC03 predicts a modest amount of recent star-formation at low-redshift, but this fraction increases to higher redshifts (5th panel). The results is a mass-weighted formation time that gets younger with redshift, and an increasing metallicity. It is worth noting that the fraction of mass in young/intermediate stars is not generally uniformly distributed over the last 3 Gyr of the galaxies. Instead, VESPA generally assigns the mass to one or two age bins, and the particular bins with the most mass change from stack to stack. In the case of BC03 and M10, these are firmly at the intermediate ages regime. This is probably indicative of sporadic star formation episodes rather than a continuous star formation type of history, for any given stack.

The FSPS and M10 models paint a scenario much closer

to the fiducial model of passive stellar evolution, and a constant high-metallicity. However, the M10 models require a small amount of intermediate star formation at low redshift (a little over 10% by mass for the faintest objects, and slightly less for brightest ones), which in turn drives down the mass-weighted ages of the faintest objects at low redshift. Note that this is different from the reason that drives down the mass-weighted ages of the faintest objects in FSPS - in this case it is the dominant, oldest populations that are slightly younger. This type of distinction is possible due to having time-resolved star formation histories, and it becomes crucially important if one aims to use the age-redshift relationship of LRGs to put tight constraints on $H(z)$ (e.g. Jimenez et al. 2003; Simon et al. 2005; Crawford et al. 2010; Stern et al. 2010).

Aside from the dust results, we see little coherent dependence on luminosity, in either age, metallicity or amount of recent star formation. Work on ETGs has consistently shown that fainter objects tend to have younger stars on average (e.g. Thomas et al. 2005; Carson & Nichol 2010; Zhu et al. 2010), but the reason why we do not see clearly in this study is easily understood. The range in luminosity of our LRG sample is much smaller than the traditional ETG samples, which tend to go down to fainter magnitudes, leaving us with a small lever-arm in luminosity with which to constrain this type of relationship. Furthermore, as the LRG sample is neither magnitude nor volume limited, any trend with luminosity over redshift is difficult to interpret. We do see a tendency for fainter objects to be younger, in both the FSPS and M10 cases (although for different reasons, as explained in the previous paragraph), but this is limited to $z \lesssim 0.3$. This is explained both by the lack of faint galaxies in the sample at larger redshifts, and by the larger range in magnitudes in the sample at low redshifts.

Dust extinction results are relatively consistent across different models. This is likely to be due to the fact that the dust modelling is common in all three cases, and that the large-scale changes due to the presence of dust are independent enough of large-scale changes due to age or metallicity. We consistently see a very clear separation in luminosity - fainter objects have more dust extinction than bright objects - and an increase towards high-redshift. In the case of Mastro and FSPS, we also see evidence for an upturn at low redshifts, suggesting that perhaps there is some contamination at his end by dustier galaxies (with some recent star formation in both cases, although up to 10 times more, in fractional mass, in the case of the Mastro models). Additionally, dust extinction is the only of our measurements to show a clear trend with colour - redder cells tend to have larger extinction, across all luminosities.

6.2 Evolved quantities

Our particular interest in this paper is to see how the solutions discussed in the previous section translate into the colour evolution of LRGs of different colour, luminosity and redshift. Inevitably, the model dependence observed in the previous section will manifest itself in the colour tracks.

The most obvious is the effect of recent star formation. An episode of star formation results in both a short and abrupt change in the colour tracks, and a reddening of the colours to higher redshift. The magnitude of these two ef-

fects depends on the size (in mass), age and duration of the episode of star formation. Metallicity and dust induce less abrupt changes in the colour evolution.

The interpretation of the results from the different sets of models is rather different. In the case of FSPS, with negligible amounts of recent star formation and a constant metal-rich population, the colour tracks follow the locus of observed objects very closely. As the selection box was designed around a passively evolving model, this is no less the expected behaviour. Furthermore, we see little dependence on colour or luminosity, suggesting that the differences noted in the previous section (namely dust and age of oldest population) have a small effect on the colour evolution.

It is a different matter in the case of the M10 or BC03 models. With BC03, the substantial and ever-present (in redshift, colour and luminosity) level of recent star formation means that star formation events constantly push the colour tracks outside the contours. Depending on the duration of the burst (and we have limited resolution with our grid) then an object remains outside the selection for box for a given amount of time, until the excessively blue light of the young stars subsides enough to make the object red once again. In principle, there is nothing wrong with this scenario, and it follows that if LRGs go through small and sporadic events of star formation, then their true number density is greater than the number density measured inside the selection box at any given redshift. The fraction of objects not observed due to this is also in principle calculable, with an analysis of this type and good, trustworthy modelling.

In the case of the M10 models, this effect is much less important due to the smaller and more restricted amount of recent star formation. M10 models tend to be redder (especially in $r - i$), but mostly lie within the locus of observed galaxies, even when a small amounts of star formations is detected.

Changes in colour tracks can not be attributed solely to the differences in the physical solutions. These were obtained by fitting the stacked spectra in the optical ($\lambda \gtrsim 3200\text{\AA}$, depending on the models), and we compute the colours to a redshift range that surpasses the optical band, so UV extensions were used to accomplish this. Differences in the spectral libraries, and in the treatment of blue stragglers or blue horizontal branch stars affect the inferred colour evolution in this regime, even when if they have limited influence in the optical.

6.3 Model comparison

A natural question to ask is: which set of models best describes real galaxy spectra? Answering this question, however, is far from trivial. We can use the distribution of the average residual per pixel (as calculated in Section 3.4 and shown for one stack in Fig. 3), as the simplest possible measure of goodness of fit. We show the distribution of these values for our 124 stacks and for all three SPS models in Fig. 16. There is a factor of roughly 1.5 between the mean residual of FSPS and BC03 or M10 models. This suggests FSPS models provide a better match overall for fitting the high SNR data. The physical solutions from this set of models are also the most in line with the LRG fiducial model, although the M10 solutions also broadly fit in with the mostly passive, metal-rich typical LRG picture. However the inter-

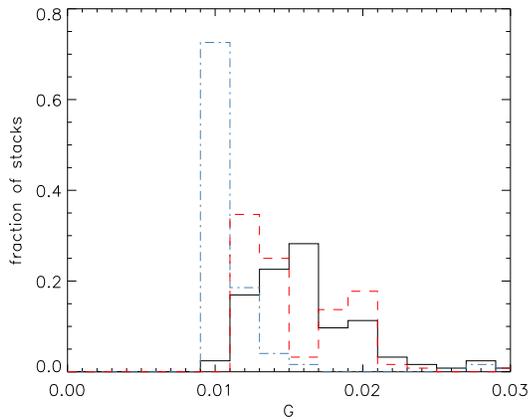


Figure 16. A simple measure of goodness of fit, taken as the average residual per pixel. In the black solid line we show the histogram for the results obtained with the BC03 models, in the red dashed line the results for the Mastro models, and in the blue dotted-dashed line the results for the FSPS models.

pretation of this result is not straightforward and there are a number of caveats.

We are in a regime where neither set provides a statistically good fit to the data, and we do not know what the true answer should be. This means that when interpreting the best-fit solution and its residual when compared against the data, we need to bear in mind that the fitting procedure can compensate for deficiencies in the models by choosing a wrong solution. This behaviour can be explicitly seen in the ages of LRGs, where the best fit solution in data-space is known to be incorrect (see Section 5.2.1 for more details). In that case, we have strong physical reason to impose a prior on the age of LRGs based on the age of the Universe obtained from independent methods. An approach based on strong physical priors, or on requiring different wavelength ranges to give consistent answers, can help to identify specific regions where each model is more likely to fail. These regions can in turn be interpreted in terms of stellar physics and library deficiencies, and be a useful tool for SPS models. Having a wider range of galaxy populations will also provide a broad test of the models. We leave this analysis for another publication.

7 SUMMARY AND CONCLUSIONS

We have computed, for the first time, colour-evolution models for the SDSS LRG sample, that depend on luminosity, colour and redshift, and are completely decoupled from the selection function of the survey. This is possible due to a full-spectral fitting technique (VESPA), which computes highly-resolved non-parametric star formation and metallicity histories, as well as the dust extinction. For 124 high signal-to-noise stacks of spectra for galaxies with different colour, luminosity and redshift, VESPA fits show how mass-weighted ages, metallicities, recent star formation and dust content vary. These solutions were then translated into colour-evolution tracks. For this analysis we have considered

three different SPS codes: BC03, M10 and FSPS, and studied the effect of this choice on the results.

Our results and conclusions can be summarised as follows:

(i) When faced with high-quality data, the models are not able to produce formally good fits to the data. We use an updated error estimate, based on the fit residuals, as a way to deal with these limitations.

(ii) The broad picture that LRGs are dominated by old, metal-rich stellar populations is confirmed by our analysis using all SPS codes. However, whereas FSPS predicts virtually no recent star formation, M10 and BC03 require some amount of young stars to fit the spectra of the LRGs.

(iii) In the case of M10 and FSPS, we see some evidence for dependence of mass-weighted age on luminosity, with the faintest sample in both cases having a lower mass-weighted age. However, whereas in the M10 models this is due to star formation at young/intermediate ages, in the case of the FSPS models this is due to the dominant, old population being younger. Distinguishing between these two scenarios is crucial for studies that aim to use the age-redshift relationship of LRGs for cosmological constraints.

(iv) All three models reveal a similar picture for the dust extinction: extinction increases with decreasing luminosity, increasing redshift, and increasing $r - i$ colour.

(v) Dust extinction is the only quantity we studied which has a significant and coherent gradient with $r - i$ colour.

(vi) The effect of recent star formation - even in small amounts - in the colour tracks can be quite dramatic, and temporarily remove LRGs from the SDSS LRG selection. It follows that, if LRGs really do go through such events, the number density of objects within the selection box is smaller than the true number density of LRGs. The fraction of LRGs going through this phase, as a function of redshift, can in principle be estimated using our models.

(vii) A test of which set of models best fits the data is non-trivial as no set gives a statistically good fit to the data. FSPS solutions produce, on average, lower residuals than the other two sets of models, but the significance of this cannot currently be established.

(viii) The differences in the colour tracks predicted by different SPS codes are the limiting factor in using these models to link populations of galaxies at different redshifts, and suggest that previous analyses that require this link (any based on the evolution of a quantity with redshift, such as number or luminosity densities, clustering, etc) are likely to at best have underestimated their errors, and at worse suffered from a systematic error due to our lack of understanding in stellar evolution.

Lastly, we note that the results in this paper apply only to the LRG population, which is a subset of the ETG population. The relationship is more pronounced in terms of mass than colour - the LRG population occupies the most massive end of the ETG population, which is in itself predominantly red.

The colour evolution tracks obtained with any of the SPS models can be found in http://www.icg.port.ac.uk/~tojeior/lrg_evolution/.

8 ACKNOWLEDGMENTS

We thank the anonymous referee for a thoughtful report that made this manuscript noticeably clearer. We would like to thank Claudia Maraston and Charlie Conroy for helpful discussions and for providing UV extensions of their models and comments on an earlier draft. We also thank David Wake for helpful and encouraging discussions.

RT thanks the Leverhulme trust for financial support. WJP is grateful for support from the UK Science and Technology Facilities Council, the Leverhulme trust and the European Research Council.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is <http://www.sdss.org/>. The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

REFERENCES

- Abazajian K. N., et al., 2009, *ApJ Supplement Series*, 182, 543
- Adelman-McCarthy J. K., et al., 2008, *ApJ Supplement Series*, 175, 297
- Alongi M., Bertelli G., Bressan A., Chiosi C., Fagotto F., Greggio L., Nasi E., 1993, *A&AS*, 97, 851
- Bernardi M., Roche N., Shankar F., Sheth R. K., 2010, *ArXiv e-prints*
- Blanton M. R., Roweis S., 2007, *AJ*, 133, 734
- Bressan A., Fagotto F., Bertelli G., Chiosi C., 1993, *A&AS*, 100, 647
- Brown M. J. I., Dey A., Jannuzi B. T., Brand K., Benson A. J., Brodwin M., Croton D. J., Eisenhardt P. R., 2007, *ApJ*, 654, 858
- Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
- Bundy K., Fukugita M., Ellis R. S., Targett T. A., Belli S., Kodama T., 2009, *ApJ*, 697, 1369
- Buzzoni A., 1989, *ApJ Supplement Series*, 71, 817
- Carson D. P., Nichol R. C., 2010, *MNRAS*, pp 1163–
- Chabrier G., 2003, *PASP*, 115, 763
- Charlot S., Fall S. M., 2000, *ApJ*, 539, 718
- Coelho P., Bruzual G., Charlot S., Weiss A., Barbuy B., Ferguson J. W., 2007, *MNRAS*, 382, 498
- Conroy C., Gunn J. E., 2009, *ArXiv e-prints*
- Conroy C., Gunn J. E., White M., 2009, *ApJ*, 699, 486
- Conroy C., Ho S., White M., 2007, *MNRAS*, 379, 1491
- Cool R. J., Eisenstein D. J., Fan X., Fukugita M., Jiang L., Maraston C., Meiksin A., Schneider D. P., Wake D. A., 2008, *ApJ*, 682, 919
- Cool R. J., Eisenstein D. J., Johnston D., Scranton R., Brinkmann J., Schneider D. P., Zehavi I., 2006, *AJ*, 131, 736
- Crawford S. M., Ratsimbazafy A. L., Cress C. M., Olivier E. A., Blyth S., van der Heyden K. J., 2010, *MNRAS*, 406, 2569
- Davies R. L., Sadler E. M., Peletier R. F., 1993, *MNRAS*, 262, 650
- De Propriis R., Driver S. P., Colless M., Drinkwater M. J., Loveday J., Ross N. P., Bland-Hawthorn J., York D. G., Pimblett K., 2010, *AJ*, 139, 794
- Eisenstein D. J., et al., 2001, *AJ*, 122, 2267
- Eisenstein D. J., et al., 2003, *ApJ*, 585, 694
- Faber S. M., et al., 2007, *ApJ*, 665, 265
- Fagotto F., Bressan A., Bertelli G., Chiosi C., 1994a, *A&AS*, 104, 365
- Fagotto F., Bressan A., Bertelli G., Chiosi C., 1994b, *A&AS*, 105, 29
- Fisher D., Franx M., Illingworth G., 1995, *ApJ*, 448, 119
- Fry J. N., 1996, *ApJL*, 461, L65+
- Girardi L., Bressan A., Chiosi C., Bertelli G., Nasi E., 1996, *A&AS*, 117, 113
- Graves G. J., Faber S. M., Schiavon R. P., 2009, *ApJ*, 698, 1590
- Gunn J. E., et al., 1998, *AJ*, 116, 3040
- Jimenez R., Bernardi M., Haiman Z., Panter B., Heavens A. F., 2007, *ApJ*, 669, 947
- Jimenez R., Jorgensen U. G., Thejll P., MacDonald J., 1995, *MNRAS*, 275, 1245
- Jimenez R., MacDonald J., Dunlop J. S., Padoan P., Peacock J. A., 2004, *MNRAS*, 349, 240
- Jimenez R., Padoan P., Matteucci F., Heavens A. F., 1998, *MNRAS*, 299, 123
- Jimenez R., Verde L., Treu T., Stern D., 2003, *ApJ*, 593, 622
- Kaviraj S., et al., 2007, *ApJ Supplement Series*, 173, 619
- Kaviraj S., Peirani S., Khochfar S., Silk J., Kay S., 2009, *MNRAS*, 394, 1713
- Kaviraj S., Tan K., Ellis R. S., Silk J., 2010, *ArXiv e-prints*
- Komatsu E., et al., 2009, *ApJ Supplement Series*, 180, 330
- Kroupa P., 2001, *MNRAS*, 322, 231
- Le Borgne J.-F., et al., 2003, *A&AS*, 402, 433
- Lejeune T., Cuisinier F., Buser R., 1997, *A&AS*, 125, 229
- Lejeune T., Cuisinier F., Buser R., 1998, *A&AS*, 130, 65
- Longhetti M., Bressan A., Chiosi C., Rampazzo R., 2000, *A&AS*, 353, 917
- Maraston C., 1998, *MNRAS*, 300, 872
- Maraston C., 2005, *MNRAS*, 362, 799
- Maraston C., Greggio L., Renzini A., Ortolani S., Saglia R. P., Puzia T. H., Kissler-Patig M., 2003, *A&AS*, 400, 823
- Maraston C., Nieves Colmenáez L., Bender R., Thomas D., 2009, *A&AS*, 493, 425
- Maraston C., Strömbäck G., Thomas D., Wake D. A., Nichol R. C., 2009, *MNRAS*, 394, L107
- Marigo P., Girardi L., 2007, *A&AS*, 469, 239

- Marigo P., Girardi L., Bressan A., Groenewegen M. A. T., Silva L., Granato G. L., 2008, *A&AS*, 482, 883
- Masjedi M., Hogg D. W., Blanton M. R., 2008, *ApJ*, 679, 260
- Masjedi M., Hogg D. W., Cool R. J., Eisenstein D. J., Blanton M. R., Zehavi I., Berlind A. A., Bell E. F., Schneider D. P., Warren M. S., Brinkmann J., 2006, *ApJ*, 644, 54
- Ocvirk P., Pichon C., Lançon A., Thiébaud E., 2006, *MNRAS*, 365, 46
- Pagel B. E. J., Tautvaisiene G., 1995, *MNRAS*, 276, 505
- Panter B., Jimenez R., Heavens A. F., Charlot S., 2008, *MNRAS*, 391, 1117
- Peng Y., et al., 2010, *ApJ*, 721, 193
- Renzini A., Buzzoni A., 1986, in C. Chiosi & A. Renzini ed., *Spectral Evolution of Galaxies Vol. 122 of Astrophysics and Space Science Library*, Global properties of stellar populations and the spectral evolution of galaxies. pp 195–231
- Salim S., Rich R. M., 2010, *ApJL*, 714, L290
- Sánchez-Blázquez P., Peletier R. F., Jiménez-Vicente J., Cardiel N., Cenarro A. J., Falcón-Barroso J., Gorgas J., Selam S., Vazdekis A., 2006, *MNRAS*, 371, 703
- Sheth R. K., Jimenez R., Panter B., Heavens A. F., 2006, *ApJL*, 650, L25
- Simon J., Verde L., Jimenez R., 2005, *Phys. Rev. D*, 71, 123001
- Stern D., Jimenez R., Verde L., Kamionkowski M., Stanford S. A., 2010, *JCAP*, 2, 8
- Stoughton C., et al., 2002, *AJ*, 123, 485
- Strauss M. A., et al., 2002, *AJ*, 124, 1810
- Thomas D., Maraston C., Bender R., 2003, *MNRAS*, 339, 897
- Thomas D., Maraston C., Bender R., Mendes de Oliveira C., 2005, *ApJ*, 621, 673
- Thomas D., Maraston C., Schawinski K., Sarzi M., Silk J., 2010, *MNRAS*, 404, 1775
- Tojeiro R., Heavens A. F., Jimenez R., Panter B., 2007, *MNRAS*, 381, 1252
- Tojeiro R., Percival W. J., 2010, *MNRAS*, 405, 2534
- Tojeiro R., Wilkins S., Heavens A. F., Panter B., Jimenez R., 2009, *ApJ Supplement Series*, 185, 1
- Trager S. C., Faber S. M., Worthey G., González J. J., 2000, *AJ*, 120, 165
- Trager S. C., Somerville R. S., 2009, *MNRAS*, 395, 608
- van Dokkum P. G., et al., 2010, *ApJ*, 709, 1018
- Wake D. A., et al., 2006, *MNRAS*, 372, 537
- Wake D. A., et al., 2008, *MNRAS*, 387, 1045
- Westera P., Lejeune T., Buser R., Cuisinier F., Bruzual G., 2002, *A&AS*, 381, 524
- White M., Zheng Z., Brown M. J. I., Dey A., Jannuzi B. T., 2007, *ApJL*, 655, L69
- Woosley S. E., Weaver T. A., 1995, *ApJ Supplement Series*, 101, 181
- Worthey G., Faber S. M., Gonzalez J. J., 1992, *ApJ*, 398, 69
- York D. G., et al., 2000, *AJ*, 120, 1579
- Zhu G., Blanton M. R., Moustakas J., 2010, *ArXiv e-prints*

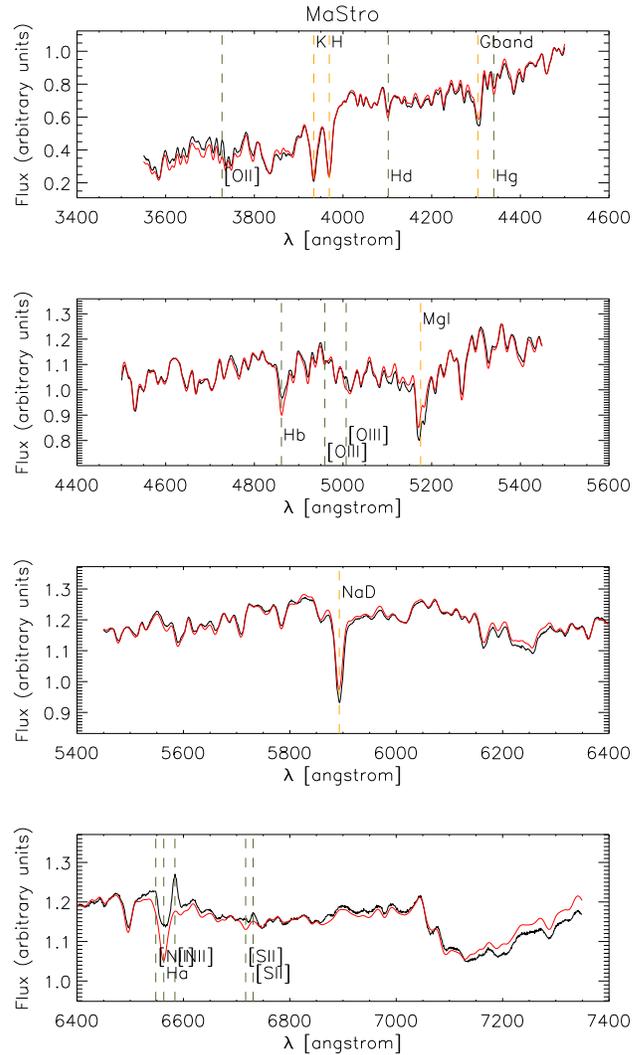


Figure A1. A typical fit, using the M10 models. The black line is the data, and the red line the best-fitting model. The vertical yellow and green dashed lines guide the eye by showing some common emission and absorption features. Although plotted, some regions of the spectrum were excluded from the fit - see Fig. A2.

APPENDIX A: ADDITIONAL PLOTS WITH DIFFERENT SPS MODELS

In this Appendix we show the equivalent of the plots shown in the main text, for all other SPS models considered. All plots are like for like, and we refer the reader to the main text for an explanation.

A1 Fits and residuals

In this section we show the equivalent of Fig. 4 and Fig. 5 for the same stacked spectra, analysed with the M10 models (in Fig. A1 and Fig. A2) and with the BC03 models (in Fig. A3 and Fig. A3).

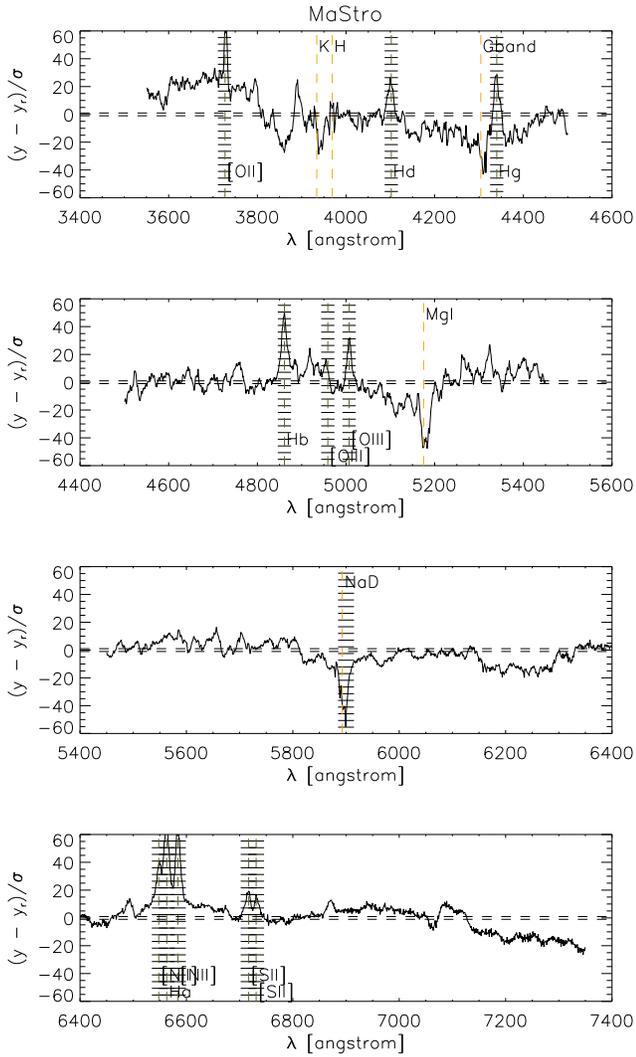


Figure A2. The residuals in units of the stack noise, for the fit showed in Fig. A1. The vertical dashed lines guide the eye by showing some common emission and absorption features. The regions shaded by horizontal lines were excluded from the fit.

A2 2D histograms

In this section we present the 2D histograms of Section 5.2, which show the dependence of age, metallicity, dust extinction and amount of recent star formation (< 3 Gyr) as a function of colour and redshift. Figs. A5 to A8 show the results obtained with using the FSPS code, and Figs. A9 to A12 the results obtained using the BC03 code.

A3 Colour tracks

Here we present the colour tracks, as in shown Section 5.3.2, computed using the M10 models (Figs. A13 to A15) and the BC03 models (Figs. A16 to A18).

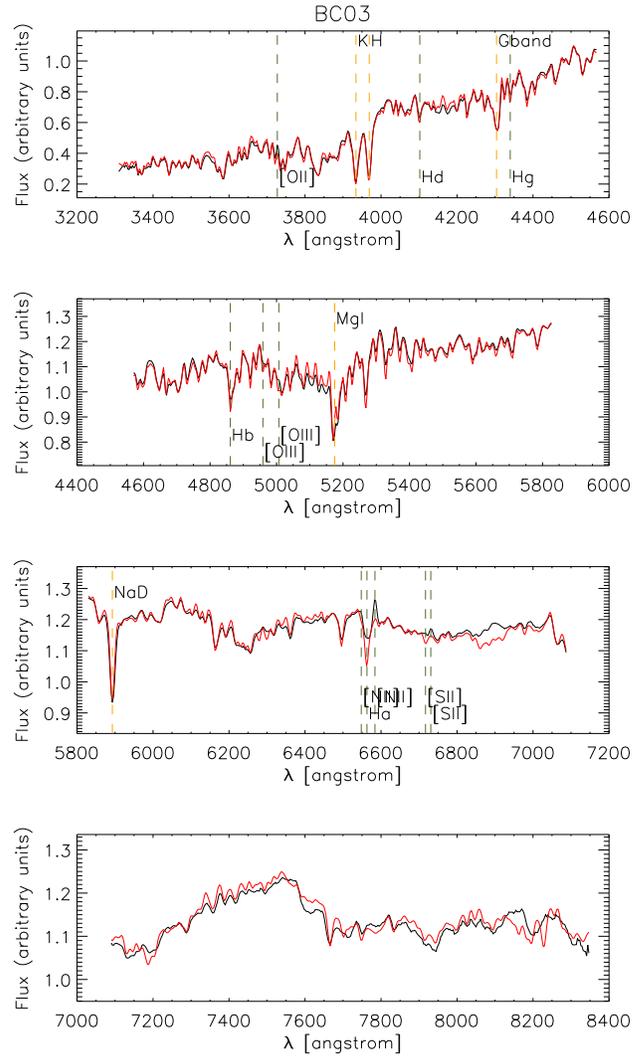


Figure A3. A typical fit, using the BC03 models. The black line is the data, and the red line the best-fitting model. The vertical yellow and green dashed lines guide the eye by showing some common emission and absorption features. Although plotted, some regions of the spectrum were excluded from the fit - see Fig. A4.

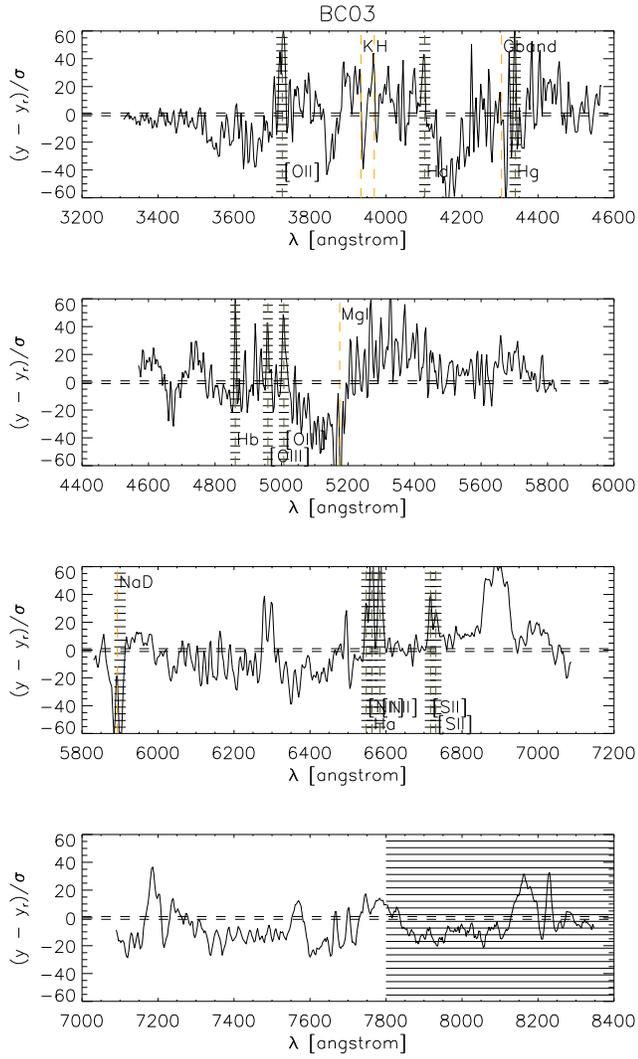


Figure A4. The residuals in units of the stack noise, for the fit showed in Fig. A3. The vertical dashed lines guide the eye by showing some common emission and absorption features. The regions shaded by horizontal lines were excluded from the fit.

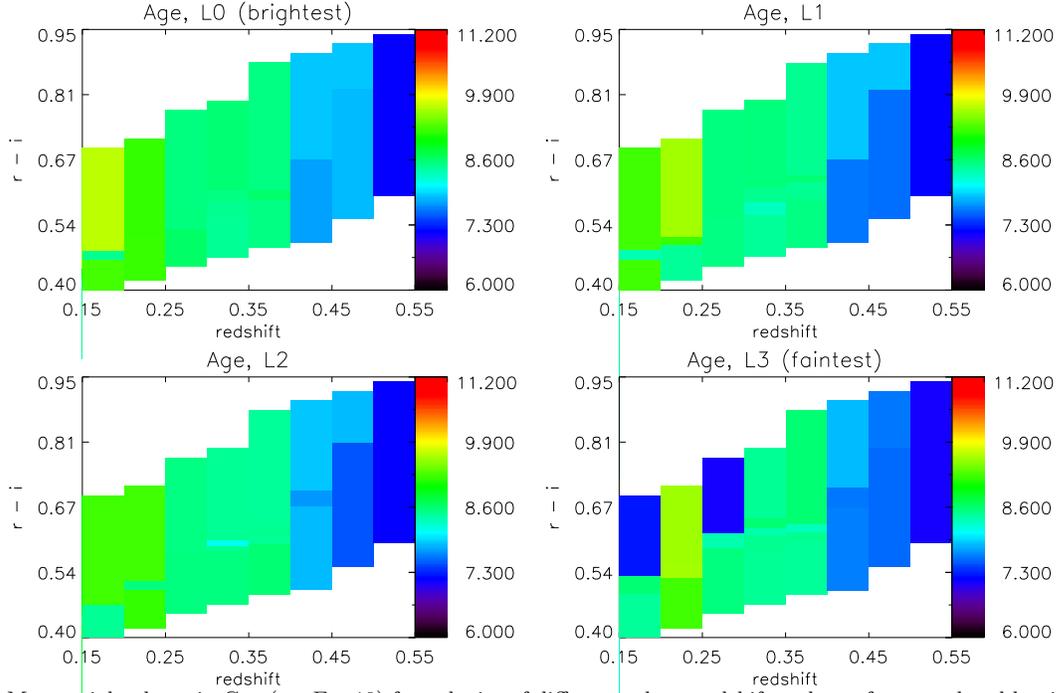


Figure A5. Mass-weighted age in Gyr (see Eq. 10) for galaxies of different colour, redshift and rest-frame r -band luminosity, analysed with the FSPS models. We only show data for regions of parameter space with a sufficient number density of galaxies. In practice, there are a few galaxies (less than 20 per redshift bin) outside of the coloured areas, but these have an insignificant impact on the recovered solution. In Fig. 6 (3rd row) we average over colour to show the trend with redshift for each luminosity slice.

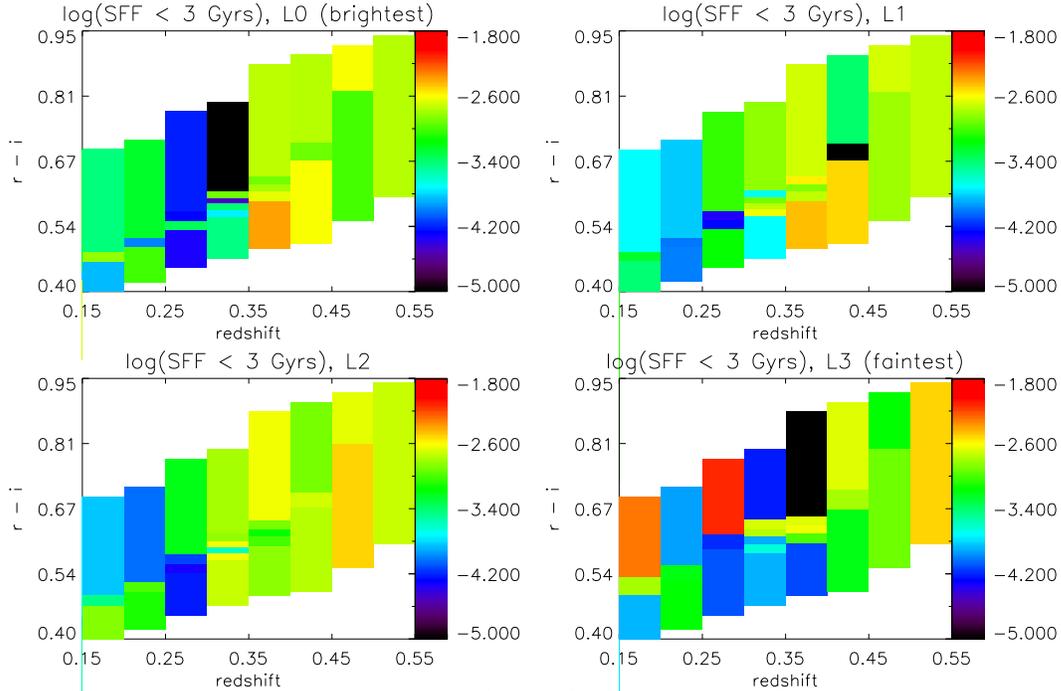


Figure A6. The logarithm of the fraction of star formation (by mass), recovered in bins up to 3.8 Gyr for galaxies of different colour, redshift and rest-frame r -band luminosity, analysed with the FSPS models. We only show data for regions of parameter space with a sufficient number density of galaxies. In practice, there are a few galaxies (less than 20 per redshift bin) outside of the coloured areas, but these have an insignificant impact on the recovered solution. In Fig. 6 (5th row) we average over colour to show the trend with redshift for each luminosity slice.

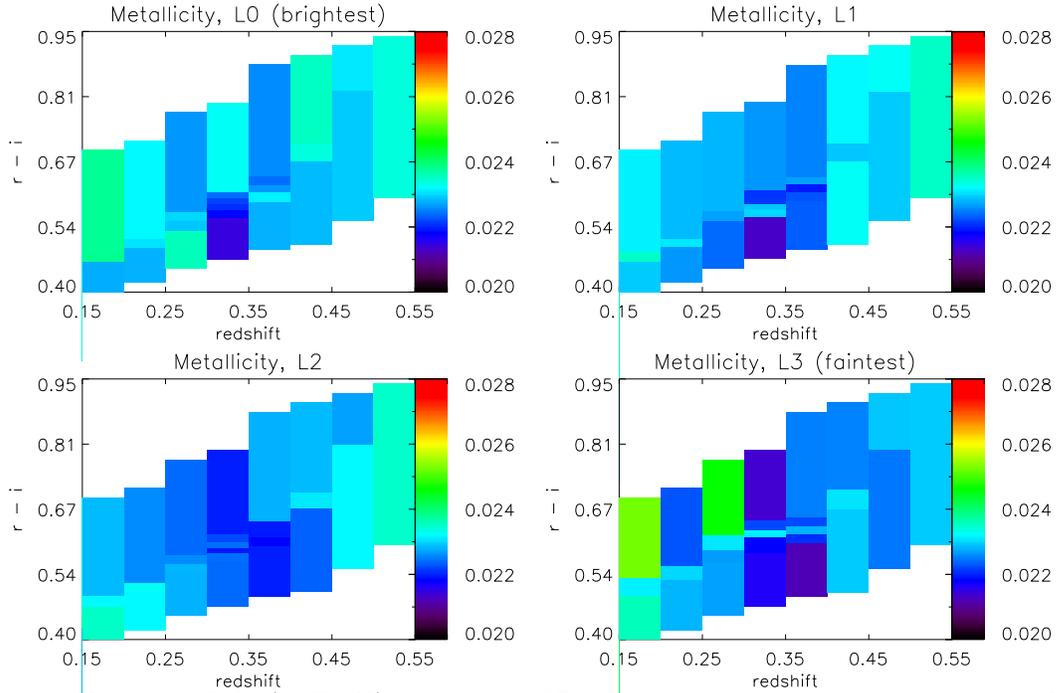


Figure A7. Mass-weighted metallicity (see Eq. 11) for galaxies of different colour, redshift and rest-frame r -band luminosity, analysed with the FSPS models. We only show data for regions of parameter space with a sufficient number density of galaxies. In practice, there are a few galaxies (less than 20 per redshift bin) outside of the coloured areas, but these have an insignificant impact on the recovered solution. Fig. 6 averages over colour to show the trend with redshift for each luminosity slice.

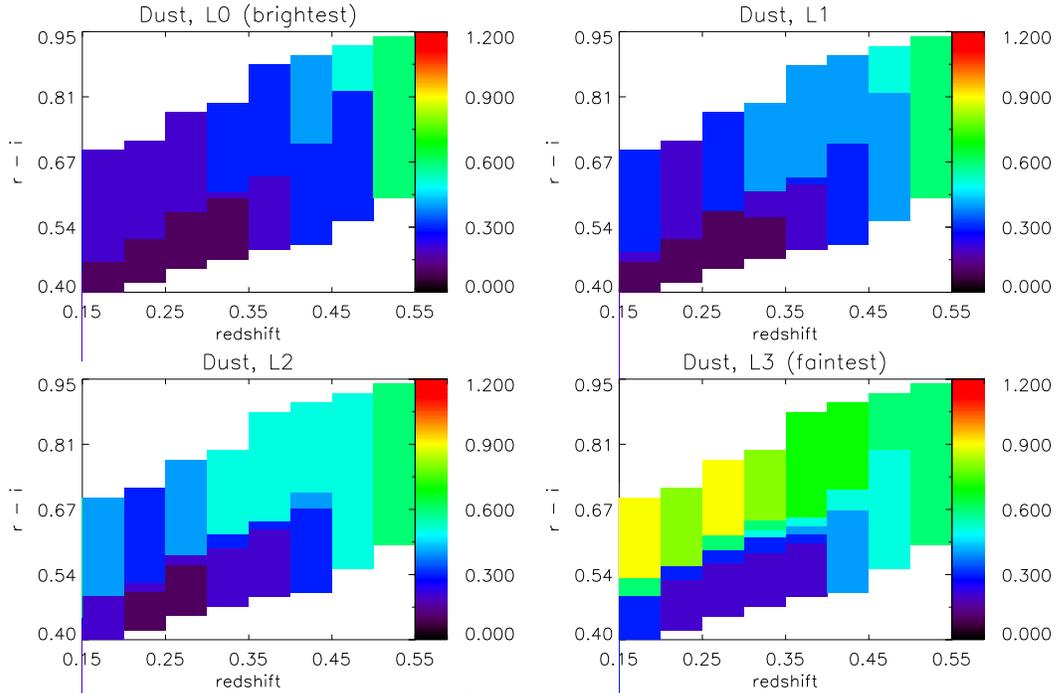


Figure A8. Inter-stellar dust absorption for galaxies of different colour, redshift and rest-frame r -band luminosity, analysed with the FSPS models. We only show data for regions of parameter space with a sufficient number density of galaxies. In practice, there are a few galaxies (less than 20 per redshift bin) outside of the coloured areas, but these have an insignificant impact on the recovered solution. Fig. 6 averages over colour to show the trend with redshift for each luminosity slice.

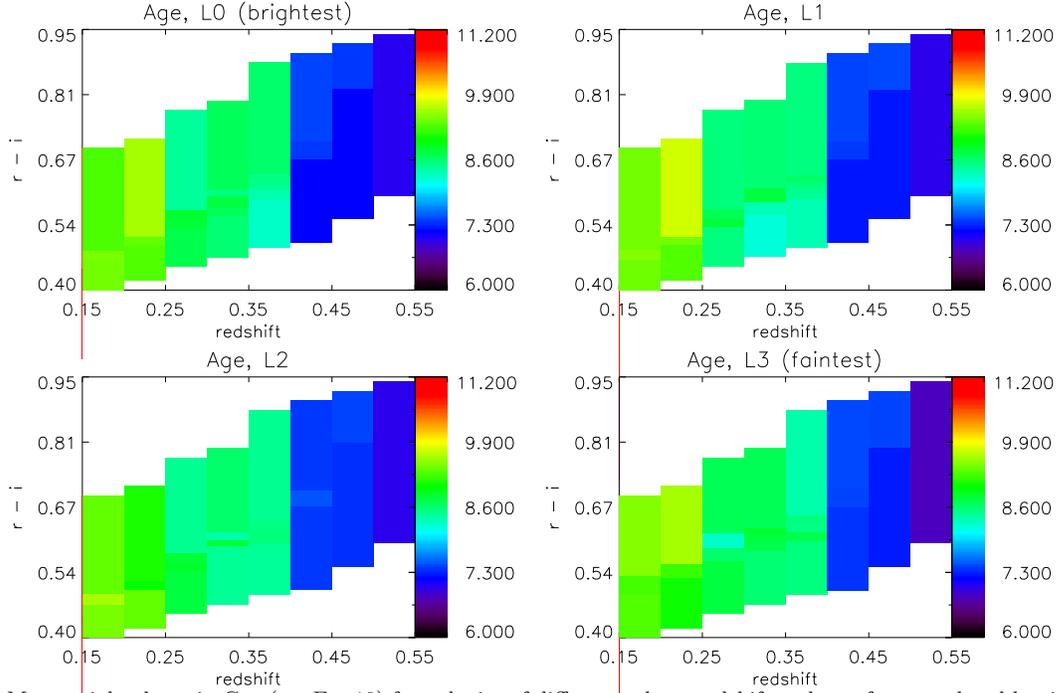


Figure A9. Mass-weighted age in Gyr (see Eq. 10) for galaxies of different colour, redshift and rest-frame r -band luminosity, analysed with the BC03 models. We only show data for regions of parameter space with a sufficient number density of galaxies. In practice, there are a few galaxies (less than 20 per redshift bin) outside of the coloured areas, but these have an insignificant impact on the recovered solution. In Fig. 6 (3rd row) we average over colour to show the trend with redshift for each luminosity slice.

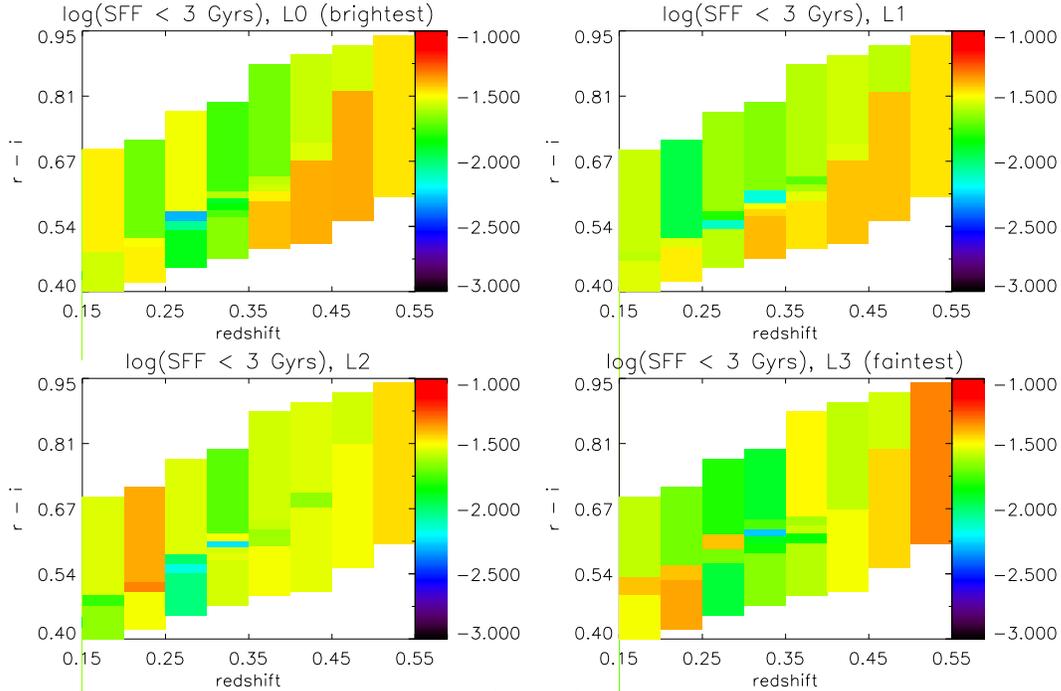


Figure A10. The logarithm of the fraction of star formation (by mass), recovered in bins up to 3.8 Gyr for galaxies of different colour, redshift and rest-frame r -band luminosity, analysed with the BC03 models. We only show data for regions of parameter space with a sufficient number density of galaxies. In practice, there are a few galaxies (less than 20 per redshift bin) outside of the coloured areas, but these have an insignificant impact on the recovered solution. In Fig. 6 (5th row) we average over colour to show the trend with redshift for each luminosity slice.

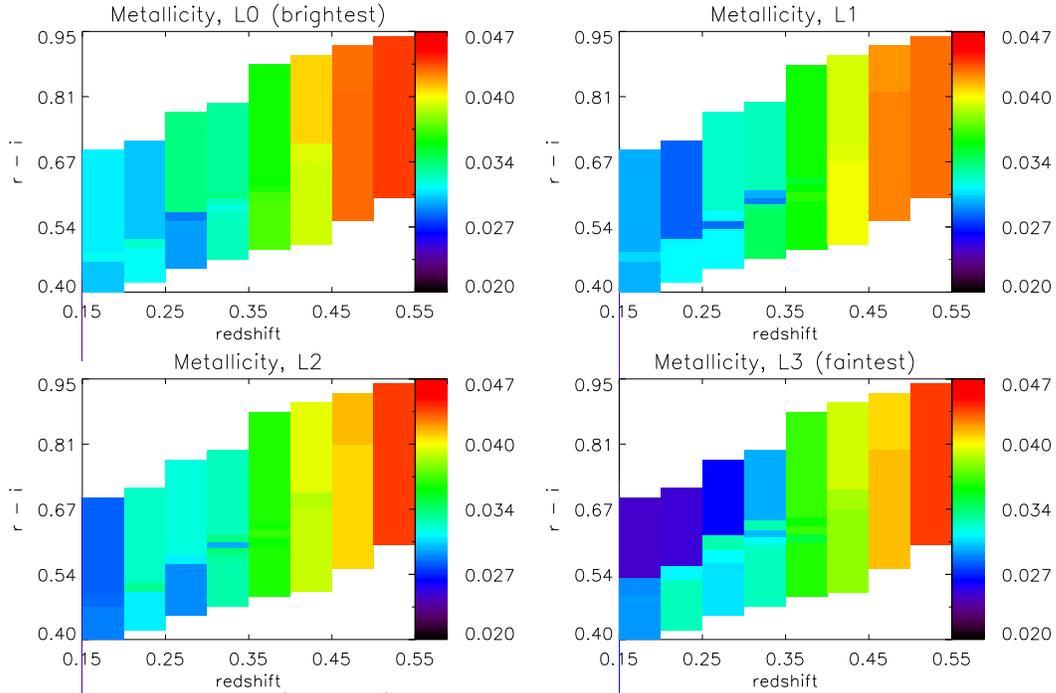


Figure A11. Mass-weighted metallicity (see Eq. 11) for galaxies of different colour, redshift and rest-frame r -band luminosity, analysed with the BC03. We only show data for regions of parameter space with a sufficient number density of galaxies. In practice, there are a few galaxies (less than 20 per redshift bin) outside of the coloured areas, but these have an insignificant impact on the recovered solution. Fig. 6 averages over colour to show the trend with redshift for each luminosity slice.

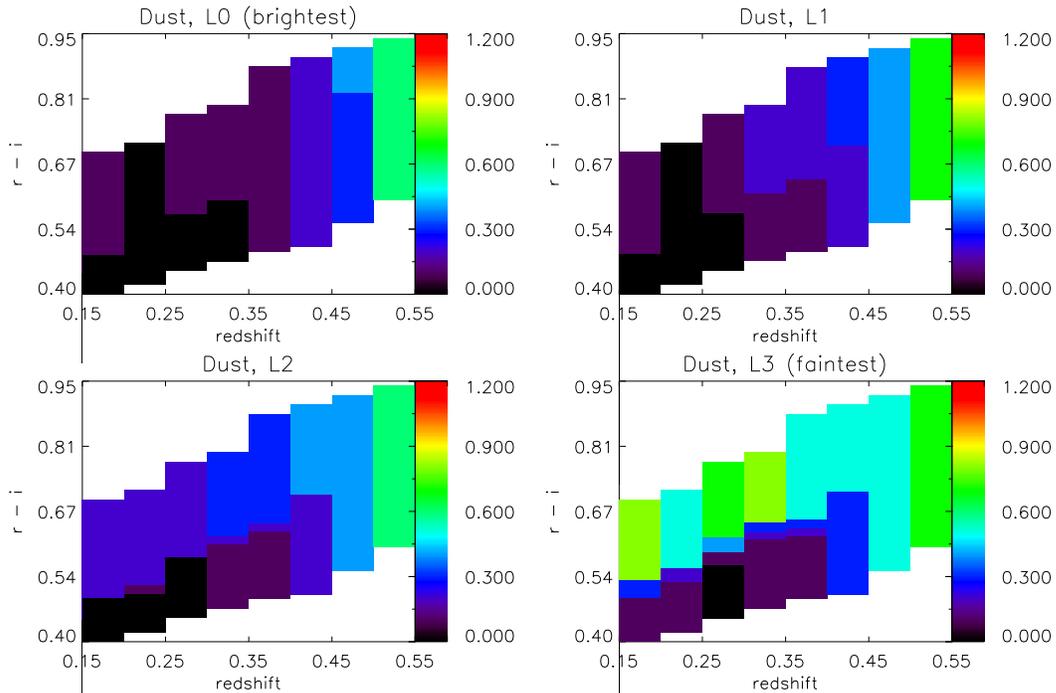


Figure A12. Inter-stellar dust absorption for galaxies of different colour, redshift and rest-frame r -band luminosity, analysed with the BC03 models. We only show data for regions of parameter space with a sufficient number density of galaxies. In practice, there are a few galaxies (less than 20 per redshift bin) outside of the coloured areas, but these have an insignificant impact on the recovered solution. Fig. 6 averages over colour to show the trend with redshift for each luminosity slice.

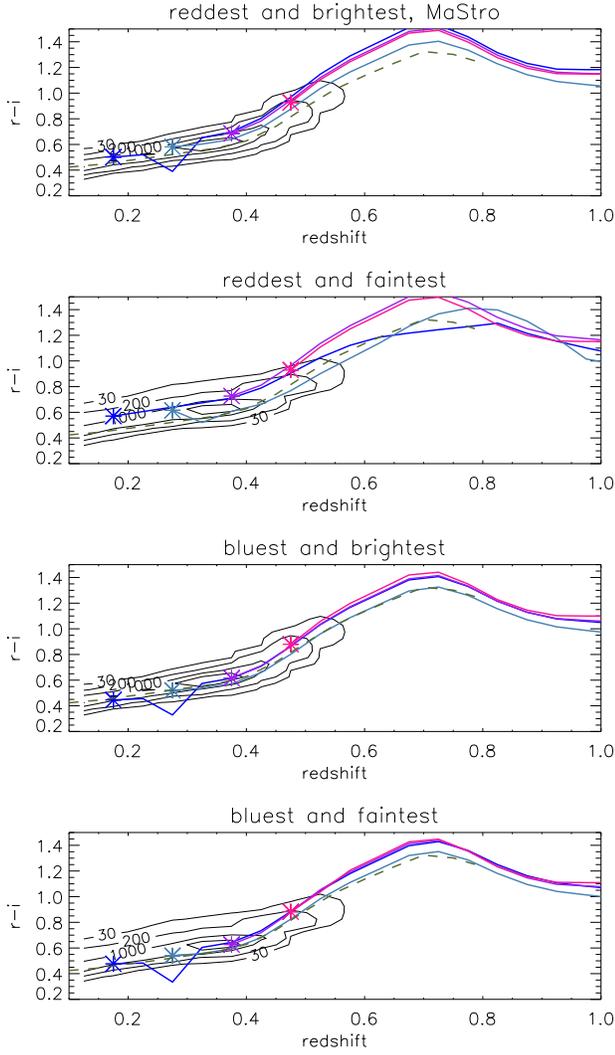


Figure A13. Predicted evolution of $r - i$ colour with redshift, calculated according to Eq. (9), and the solutions presented in Section 5.2, and using the M10. The different panels show examples at different combinations of colour and luminosity. The black contour lines show the number density of LRGs within the respective luminosity slice, and the coloured lines show the tracks from cells at four different redshifts: 0.175, 0.275, 0.375 and 0.475. The star shows the position of the cell the track relates to. The green dashed line shows the LRG model from Maraston et al. (2009), for comparison.

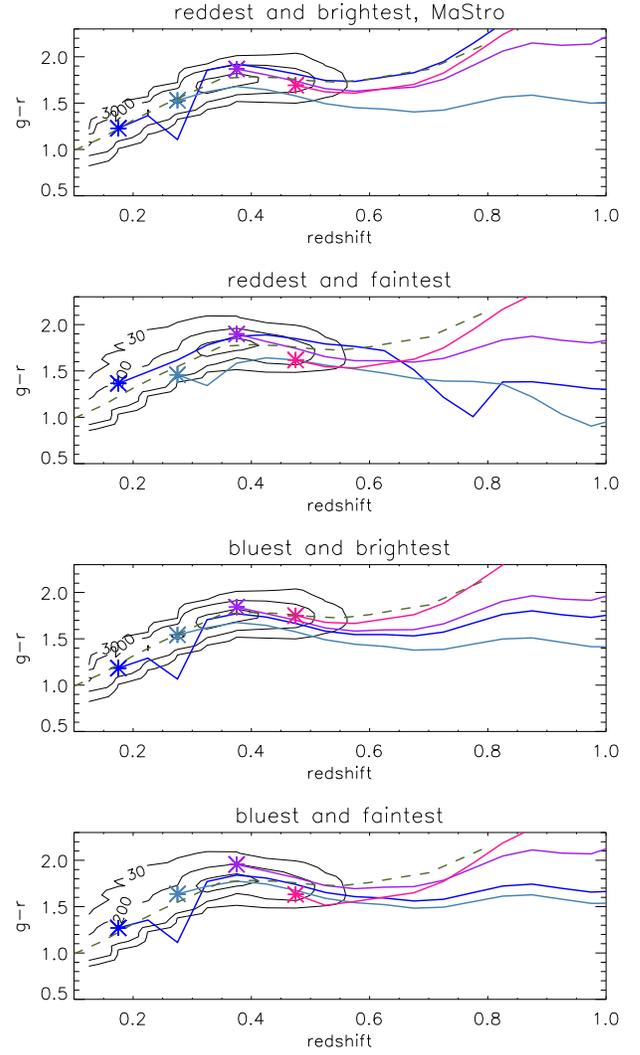


Figure A14. Predicted evolution of $g - r$ colour with redshift, calculated according to Eq. (9), and the solutions presented in Section 5.2, and using the M10. The different panels show examples at different combinations of colour and luminosity. The black contour lines show the number density of LRGs within the respective luminosity slice, and the coloured lines show the tracks from cells at four different redshifts: 0.175, 0.275, 0.375 and 0.475. The star shows the position of the cell the track relates to. The green dashed line shows the LRG model from Maraston et al. (2009), for comparison.

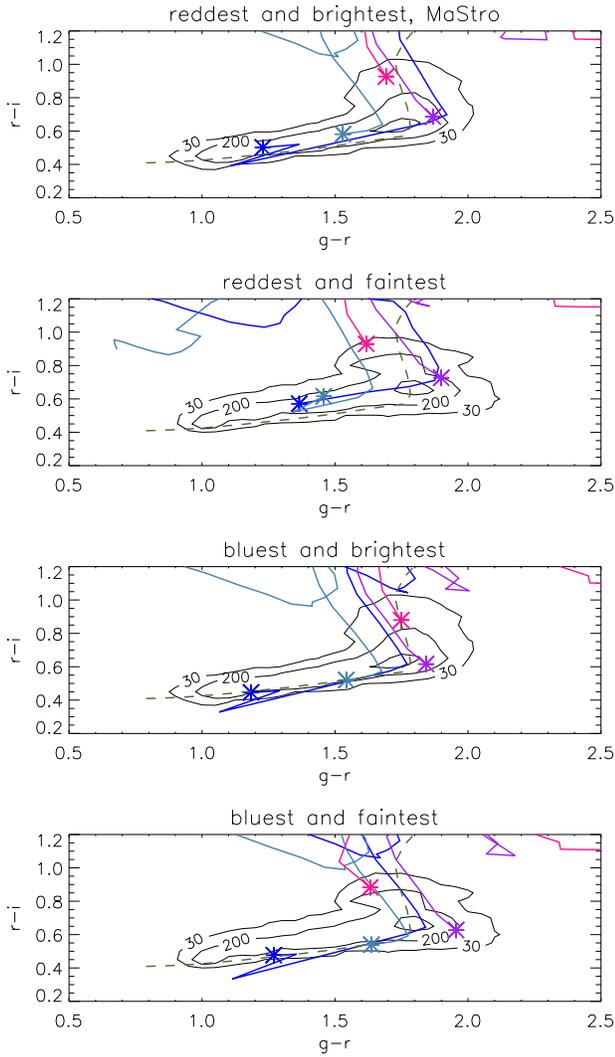


Figure A15. Predicted evolution of $r - i$ colour with $g - r$, calculated according to Eq. (9), and the solutions presented in Section 5.2, and using the M10 models. The different panels show examples at different combinations of colour and luminosity. The black contour lines show the number density of LRGs within the respective luminosity slice, and the coloured lines show the tracks from cells at four different redshifts: 0.175, 0.275, 0.375 and 0.475. The star shows the position of the cell the track relates to. The green dashed line shows the LRG model from Maraston et al. (2009), for comparison.

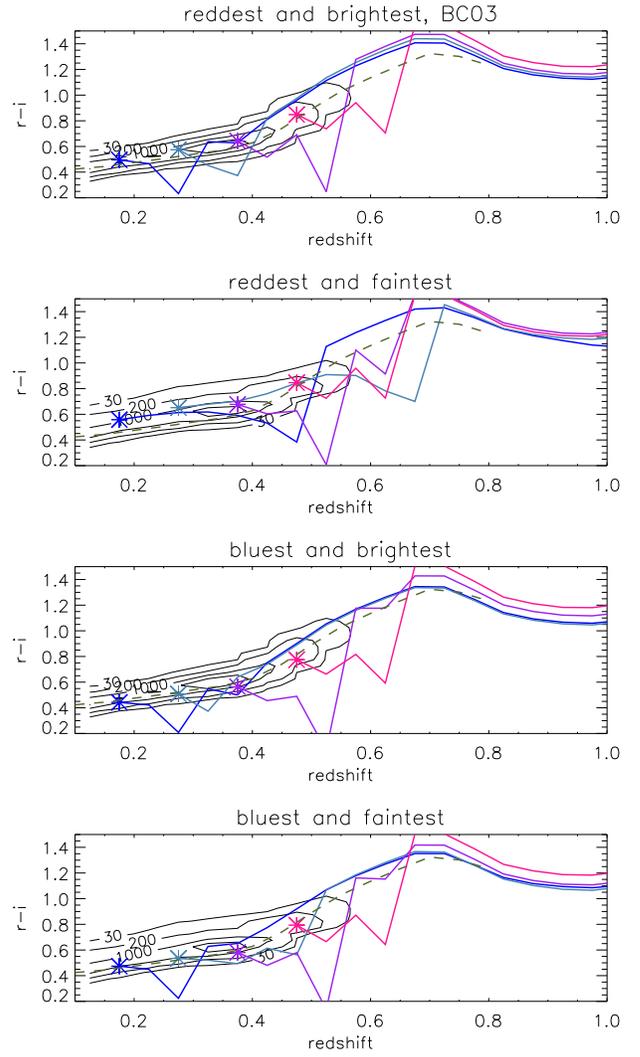


Figure A16. Predicted evolution of $r - i$ colour with redshift, calculated according to Eq. (9), and the solutions presented in Section 5.2, and using the BC03 models. The different panels show examples at different combinations of colour and luminosity. The black contour lines show the number density of LRGs within the respective luminosity slice, and the coloured lines show the tracks from cells at four different redshifts: 0.175, 0.275, 0.375 and 0.475. The star shows the position of the cell the track relates to. The green dashed line shows the LRG model from Maraston et al. (2009), for comparison.

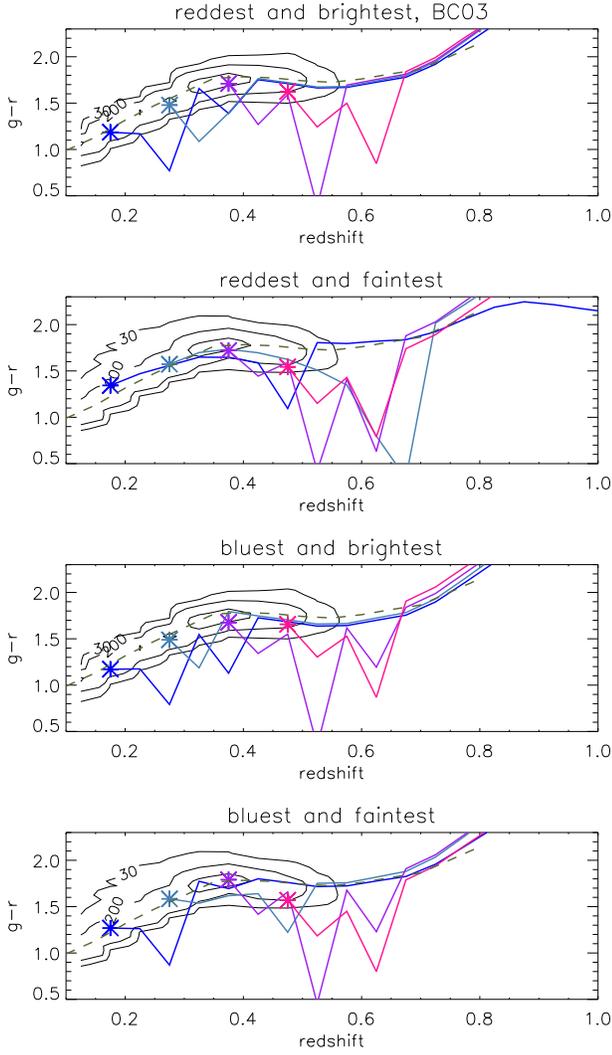


Figure A17. Predicted evolution of $g-r$ colour with redshift, calculated according to Eq. (9), and the solutions presented in Section 5.2, and using the BC03 models. The different panels show examples at different combinations of colour and luminosity. The black contour lines show the number density of LRGs within the respective luminosity slice, and the coloured lines show the tracks from cells at four different redshifts: 0.175, 0.275, 0.375 and 0.475. The star shows the position of the cell the track relates to. The green dashed line shows the LRG model from Maraston et al. (2009), for comparison.

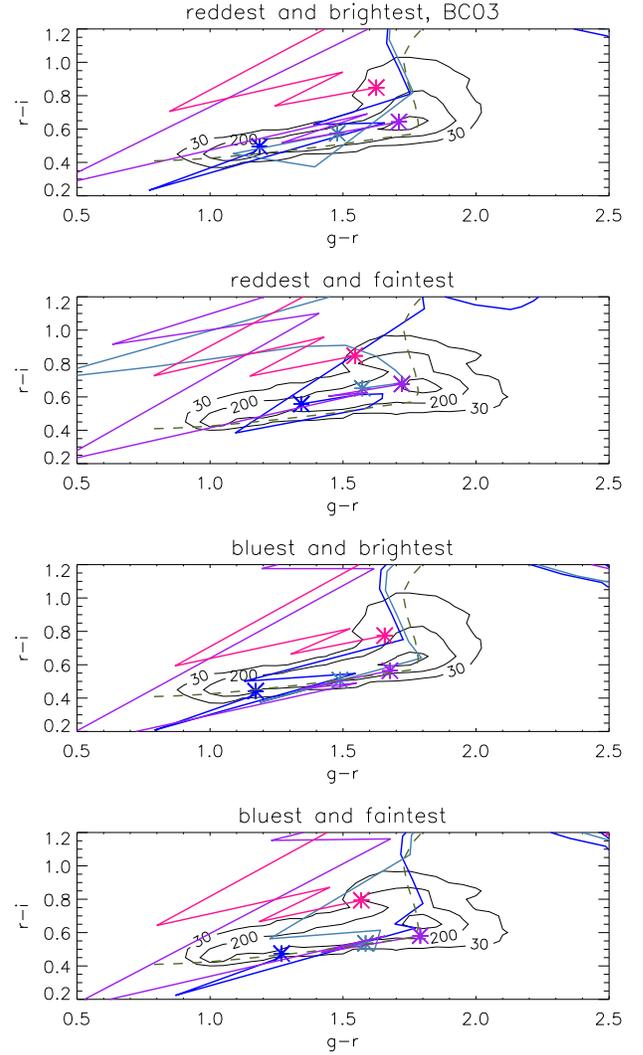


Figure A18. Predicted evolution of $r-i$ colour with $g-r$, calculated according to Eq. (9), and the solutions presented in Section 5.2, and using the BC03 models. The different panels show examples at different combinations of colour and luminosity. The black contour lines show the number density of LRGs within the respective luminosity slice, and the coloured lines show the tracks from cells at four different redshifts: 0.175, 0.275, 0.375 and 0.475. The star shows the position of the cell the track relates to. The green dashed line shows the LRG model from Maraston et al. (2009), for comparison.