Photometric redshifts and model spectral energy distributions of galaxies from the SDSS-III BOSS DR10 data

N. Greisel,[1,2,3]★ S. Seitz,[1,2] N. Drory,[2,4] R. Bender,[1,2] R. P. Saglia[1,2] and J. Snigula[1,2]

[1]*University Observatory Munich, Scheinerstrasse 1, D-81679 Munich, Germany*

[2]*Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, D-85748 Garching, Germany*

[3]*Centro de Estudios de Física del Cosmos de Aragón (CEFCA), Plaza San Juan 1, E-44001 Teruel, Spain*

[4]*Department of Astronomy, The University of Texas at Austin, 2515 Speedway, Stop C1400, Austin, TX 78712, USA*

ABSTRACT

We construct a set of model spectra specifically designed to match the colours of the Baryon Oscillation Spectroscopic Survey CMASS galaxies and to be used with photometric redshift template fitting techniques. As a basis, we use a set of spectral energy distributions(SEDs) of single and composite stellar population models. These models cannot describe well the whole colour range populated by the CMASS galaxies at all redshifts, wherefore we modify them by multiplying the SEDs with $\lambda^{-\beta}$ for $\lambda > \lambda_i$ for different values of λ_i and β. When fitting these SEDs to the colours of the CMASS sample, with a burst and dust components in superposition, we can recreate the location in colour spaces inhabited by the CMASS galaxies. From the best-fitting models, we select a small subset in a two-dimensional plane, whereto the galaxies were mapped by a self-organizing map. These models are used for the estimation of photometric redshifts with a Bayesian template fitting code. The photometric redshifts with the novel templates have a very small outlier rate of 0.22 per cent, a low bias $\langle \Delta z/(1+z) \rangle = 2.0 \times 10^{-3}$, and scatter of $\sigma_{68} = 0.026$ in the rest frame. Using our models, the galaxy colours are reproduced to a better extent with the photometric redshifts of this work than with photometric redshifts of Sloan Digital Sky Survey.

Key words: galaxies: distances and redshifts – galaxies: evolution.

1 INTRODUCTION

Spectroscopic surveys provide very precise measurements of the cosmological redshift, but they are time consuming and cannot be applied to fainter galaxies. Significantly larger volumes of the cosmos can be probed on shorter time-scales with photometric surveys by observation through selected filter bands. Since spectral features cannot be resolved by medium- or broad-band photometry, one has to apply statistical methods to derive the photometric redshift(photo-z). Redshift measurements are necessary in many cosmological contexts whenever information about the redshift tomography, or the distribution of galaxies as a function of redshift is required. The results of cosmological applications strongly depend on the photometric redshift accuracies. Large biases and scatter in the photo-zs can deteriorate any such study, e.g. the dark energy constraints from shear tomography (Ma, Hu & Huterer 2006), or the baryonic acoustic oscillation scale (Benítez et al. 2009; Sánchez et al. 2011). However, statistical errors can be accounted for if they are well known (e.g. Huterer et al. 2006). Furthermore, in opposition to a single photoz estimate, including the probability density

function (PDF) $P(z)$ in the analysis of the photo-z uncertainties enhances the accuracy of cosmological measurements (e.g. Mandelbaum et al. 2008; Hildebrandt et al. 2012).

The techniques for photoz estimation are commonly divided into two categories: template fitting and empirical methods. Empirical methods learn a relation between the photometric observables and the spectroscopic redshift (spec-z) of a training set of galaxies in order to employ that relation to objects without spectroscopic information (e.g. Collister & Lahav 2004; Gerdes 2009; Carrasco Kind & Brunner 2013, 2014). These techniques have the great advantage that they take calibration errors explicitly into account, and that they can include photometric observables other than magnitudes and colours. Nevertheless, they strongly rely on the training sample which has to be a good representation of the query galaxies to yield accurate photo-zs. Furthermore, empirical methods have to be trained anew when they are applied to different surveys, filter systems, or extracted magnitudes or fluxes. Also, many empirical methods do not take photometric measurement uncertainties into account.

In contrast to empirical methods, template fitting techniques essentially perform a maximum likelihood fit to the data (e.g. Arnouts et al. 1999; Benítez 2000; Bolzonella, Miralles & Pelló 2000; Ilbert et al. 2006; Feldmann et al. 2006; Brammer, van Dokkum & Coppi

*E-mail: ngreisel@cefca.es

2008). A set of template SEDs are shifted to several redshift steps, where they are multiplied with the filter functions and integrated over. The such predicted fluxes of the templates are then fitted to the photometry of the data. Therefore, template fitting techniques can be applied to data from any photometric system, and, concurrently to the photoz estimation, also provide restframe properties. In order to predict the photo-zs to a high precision, the data have to be well calibrated (or a catalogue with precise redshift measurements has to be available to re-calibrate the zero points). Furthermore, it is of paramount importance that the underlying model SEDs represent the data in question. In order to achieve that, some codes use combinations of model SEDs, or 'repair' the templates (e.g. Csabai et al. 2003; Feldmann et al. 2006; Brammer et al. 2008). However, the underlying SEDs have to be already well selected if a modification of them should succeed as a representation of the data. Moreover, if one introduces templates that do not match the data, the quality of the photometric redshifts is deteriorated. Therefore, we generate in this work a model set designed to match the data in question.

In Greisel et al. (2013, hereafter G13), we created model SEDs for spectroscopically observed luminous red galaxies (LRGs) from the Sloan Digital Sky Survey data release 7 (SDSS-II; York et al. 2000; Eisenstein et al. 2001; Abazajian et al. 2009). These models were then used for the estimation of photo-zs and yielded accurate results. The luminous red galaxy(LRG) sample of SDSS-II included LRGs only up to redshift $z \approx 0.5$. In this paper, we extend the work done in G13 and generate a set of model SEDs on the basis of the CMASS sample of the Baryon Oscillation Spectroscopic Survey (BOSS; Eisenstein et al. 2011; Dawson et al. 2013; Ahn et al. 2014). These templates are created on the basis of the colours of the CMASS galaxies. They are specifically designed and selected to produce accurate photometric redshifts with template fitting techniques.

This paper is organized as follows. We present the template fitting photometric redshift code used in this work in Section 2. In Section 3, the data used in this work are described. We briefly explain the SED fitting routine in Section 4. After that, we will go into detail about the generation of model SEDs and how we modify them in order to fit the data colours to a better extent. At the end of Section 4. we select a model set which is to be used with template fitting photometric redshift codes. In Section 5, we present the photometric redshift results of our new models and compare them with the two different photo-zs available on the SDSS data base. We also predict the colours of our models when fitted to the more accurate one of the SDSS photo-zs in comparison to their prediction with the redshifts of this work. Finally, we give a summary and conclusions in Section 6.

Throughout this paper, we assume a Λ cold dark matter cosmology with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ and $H_0 = 70\,\mathrm{km\,s^{-1}Mpc^{-1}}$. Magnitudes are $\log_{10}$ magnitudes given in the AB system (Oke & Gunn 1983).

2 PHOTOMETRIC REDSHIFTS

For photometric redshift estimation in this paper, we use the Bayesian template fitting code PHOTOZ (Bender et al. 2001). Template fitting codes essentially determine the photometric redshift by performing a maximum likelihood fit of the predicted colours of a set of template SEDs, calculated by the multiplication of the SEDs with the survey filter functions and integrated, to the observed galaxy colours at varying redshifts. In order to lift degeneracies in the colours of model SEDs, Bayesian codes include the possibility to assign prior probabilities in redshift and luminosity to individual models. The resulting probability of a model-redshift combination then reads

$$P(\boldsymbol{\mu}|C, m) \propto \mathcal{L}(C, m|\boldsymbol{\mu}) \cdot P(\boldsymbol{\mu}),$$

where C and m are the colours and magnitudes of the photometric data, while $\boldsymbol{\mu}$ denotes the absolute magnitude M and the redshift z. The second factor is the prior probability $P(\boldsymbol{\mu}|T) = P_M(M|T) \cdot P_z(z|T)$, with the PDFs of M and z for a template T. In the case of the PHOTOZ code, the prior functions follow

$$P_x(x|T) \propto \exp\left(-\ln(2) \cdot \left(\frac{x - \mu_x}{\sigma_x}\right)^{p_x}\right),$$

with $x = z, M$. μ_x, σ_x, and p_x can be chosen for each template separately, while p_x is even. The individual photometric redshift value of a galaxy is determined by the mode of the $P(z)$ distribution z_{mode} of the best-fitting model. We estimate the error of the individual photoz prediction to

$$\delta z = \left(\sum_z (z - z_{best})^2 \cdot P(z)\right)^{1/2}. \tag{1}$$

Furthermore, we determine the stacked probability densities for all models in the template set which we will use in Section 5 in the calculation of photoz quality metrics.

The PHOTOZ code has been applied in the past to a variety of photometric catalogues (Drory et al. 2001; Gabasch et al. 2004; Drory et al. 2005; Feulner et al. 2005; Brimioulle et al. 2008; Gabasch et al. 2008; G13; Brimioulle et al. 2013; Gruen et al. 2013, 2014; Sánchez et al. 2014). It is also part of the PS1 Photometric Classification Server (Saglia et al. 2012).

3 THE BOSS CMASS SAMPLE

In this paper, we use the spectroscopic data from the BOSS (Eisenstein et al. 2011; Dawson et al. 2013). The latest public data release of BOSS (DR10; Ahn et al. 2014) targets 14 555 deg^2 of the sky obtaining spectra for 1 848 851 galaxies. The two spectrographs used in BOSS are rebuilt from the original SDSS spectrographs (Smee et al. 2013) and cover a wavelength range of 3600–10 400 Å at a resolution of 1560–2650 Å. BOSS was designed to measure the spatial distribution of LRGs and quasars to investigate the baryon acoustic oscillations that are imprinted on the large-scale structure of today's Universe. Galaxies surveyed by BOSS have redshifts up to $z \sim 1.0$. The target selection is discussed in Eisenstein et al. (2011). BOSS target galaxies are selected from SDSS-II imaging data in such a way that they have high luminosities and masses. Furthermore, the BOSS target selection requires approximate uniform stellar masses throughout the redshift range of $0.2 \lesssim z \lesssim 0.6$. Above $z \sim 0.6$, the BOSS sample is magnitude limited. Similar to the cuts of the SDSS-II LRG sample (Eisenstein et al. 2001; Padmanabhan et al. 2005), a number of magnitude and colour cuts are applied to ensure the above requirements are fulfilled. They are selected in such a way to track the colours of a passively evolving galaxy from Maraston et al. (2009, hereafter M09). The BOSS sample is divided into two subsamples, a lower redshift sample tagged *LOWZ* at $z \leq 0.4$, and a higher redshift sample, dubbed *CMASS* for the constant mass requirement (Eisenstein et al. 2011).

The SDSS-III data can be acquired from the SDSS CasJobs website.[1] For the selection of our sample, we employ the SDSS

---

[1] http://skyserver.sdss3.org/CasJobs/.
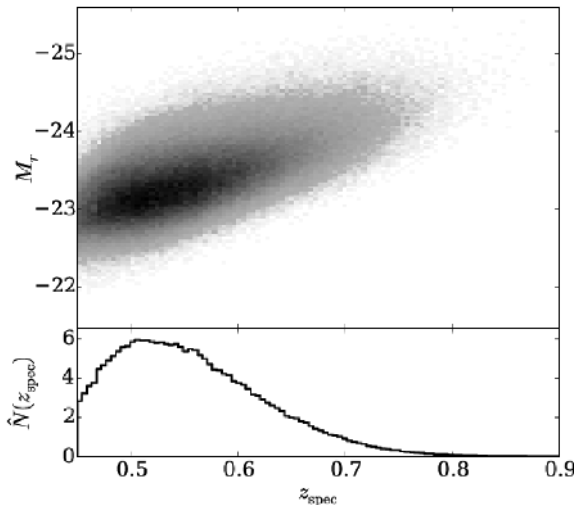
Figure 1. Upper panel: spectroscopic redshift versus absolute magnitude in the *r* band derived from a fit of G13 LRG models to the CMASS catalogue of this work. Lower panel: normalized frequency in spectroscopic redshift of the catalogue.

CLEAN photometry flag. This flag ensures that we do not have duplicates in our sample by removing multiple detections on different frames. Also, objects with deblending problems are removed, as well as the ones where more than 20 per cent of the point spread function flux is interpolated over (that is, only 80 per cent of the flux is actually detected). Additionally to that, we demand that objects were detected in the first pass (unbinned image), are not saturated, and that a radial profile could be constructed.

We require furthermore that galaxies in our sample have spectroscopic redshifts determined to a high accuracy. Therefore, we chose only objects for our catalogue whose spec-*z* warning flag is equal to zero. We want to create models specifically designed for galaxies at higher redshifts than in G13. Also, CMASS galaxies are very sparse at redshifts $z \gtrsim 0.8$, which is why we reduce the sample to galaxies at $0.45 \le z \le 0.9$. The resulting catalogue then contains 486 934 CMASS galaxies. The `CasJobs` SQL query used to acquire this catalogue is given in Appendix B.

Fig. 1 shows the redshift distribution of the such selected CMASS galaxies. In the upper panel, we plot the density of the sample in $z_{\rm spec}$ versus absolute magnitude in the *r* band. M_r is derived by fitting the LRG models of G13 to the data at known spectroscopic redshifts. The lower panel of Fig. 1 presents the frequency in spec-*z*, normalized to an integral of one.

3.1 Colours of SED templates versus colours of BOSS galaxies

In this section, we want to compare the predicted colours of model SEDs to the colours of the BOSS CMASS sample. In Fig. 2, we show the case of the M09 LRG model that is used in the definition of the colour cuts of the BOSS galaxy sample. This model was created by M09 by adding a mass fraction of 3 per cent of metal poor stars ($Z_{\odot}/200$) to a single stellar population(SSP) model with solar metallicity from the Pickles (1998) stellar library in order to match the $g - r$ and $r - i$ colours of SDSS-II LRGs. We sample the M09 model at ages of 1.0 and 8.0 Gyr and predict the colours while redshifting the SED (keeping the age fixed). Additionally, we consider two passively evolving variations of the M09 model with formation redshifts of $z_f = 1.0$ and 3.0. Since the M09 model is available only at distinct ages, we have to interpolate between them

to determine the SEDs as a function of redshift. The passive evolution was computed with the EZGAL software (Mancone & Gonzalez 2012). M09 specifically created this model to match the colours of the SDSS-II LRGs, and it fits their median colours well (see also fig. 5 in G13, or figs 1 and 2 in M09). This is also true for the colours of the CMASS sample, where we find that the model with $z_f = 1.0$ fits the $u - g$, $g - r$, and $i - z$ colours best in the observed cases. In the case of the $g - r$ and $i - z$ colours there are deviations from the data for $z \gtrsim 0.8$, but also the sample size of the CMASS galaxies in this region decreases significantly. However, the predicted colours in $r - i$ are too blue up to a redshift of $z \sim 0.72$ and lie outside the median uncertainties of the data. Furthermore, the colour errors are too small to be responsible for the spread in colour, wherefore the colour values of the data are due to the galaxy population not the photometric uncertainties.

Therefore, the data cannot be matched by the M09 model simultaneously in all colours, also not when taking evolution effects into account. The spread in colour can be explained by differences in ages of the galaxy, but could be also due to different stellar populations, i.e. the metallicity and the distribution of stellar ages.

4 NEW SED TEMPLATES

We aim to create a set of templates that can be used for photometric redshift estimation of galaxies with similar properties as the CMASS sample. To create models specifically designed to match the colours of the BOSS data, we fit a number of model SEDs to the data at their known spectroscopic redshift and select from the best-fitting models a subset that should represent the data in terms of colours, while yielding accurate photometric redshifts.

We expect the galaxy population to vary as a function of redshift. In the process of photoz estimation with template fitting, we can account for that by assigning different redshift prior probabilities to individual model SEDs. One could be tempted to use a huge amount of SEDs with different properties that can match the data in question and yield reliable photoz results (while heavily increasing the computation time). However, this is not the case since one has to deal with degeneracies in colours resulting from different galaxy properties, e.g. age and metallicity. Also, introducing peculiar templates can deteriorate the $P(z)$ of a galaxy. Therefore, one has to carefully select a small set of templates able to match the galaxy catalogue in question. In order to create a set of models that match the data at different redshifts, we fit a variety of model SEDs to the CMASS galaxies within four redshift bins. The *z* bins are centred on $z = 0.5, 0.6, 0.7, 0.8$ (continuing the sequence from G13) with interval widths of $\Delta z = 0.04$.

In the following, we first give a short introduction of the SED fitting procedure and turn afterwards to the description of the models we used as a basis to construct new model SEDs. After that we explain how we select models to represent the data that should then serve as a template set for photometric redshift estimation.

4.1 Generating model SEDs for objects in the BOSS catalogue by SED fitting

To fit model SEDs to the data, we use the SED fitting software SEDFIT (Drory, Bender & Hopp 2004). The code fits a number of model SEDs to the colours of the data by performing a maximum likelihood fit. Additionally to the models, one can define a burst model which is then fitted in superposition to the main component at several mass fractions. Furthermore, the SEDFIT code applies dust extinction to both, the main and starburst component, following the

Maraston 2009 LRG
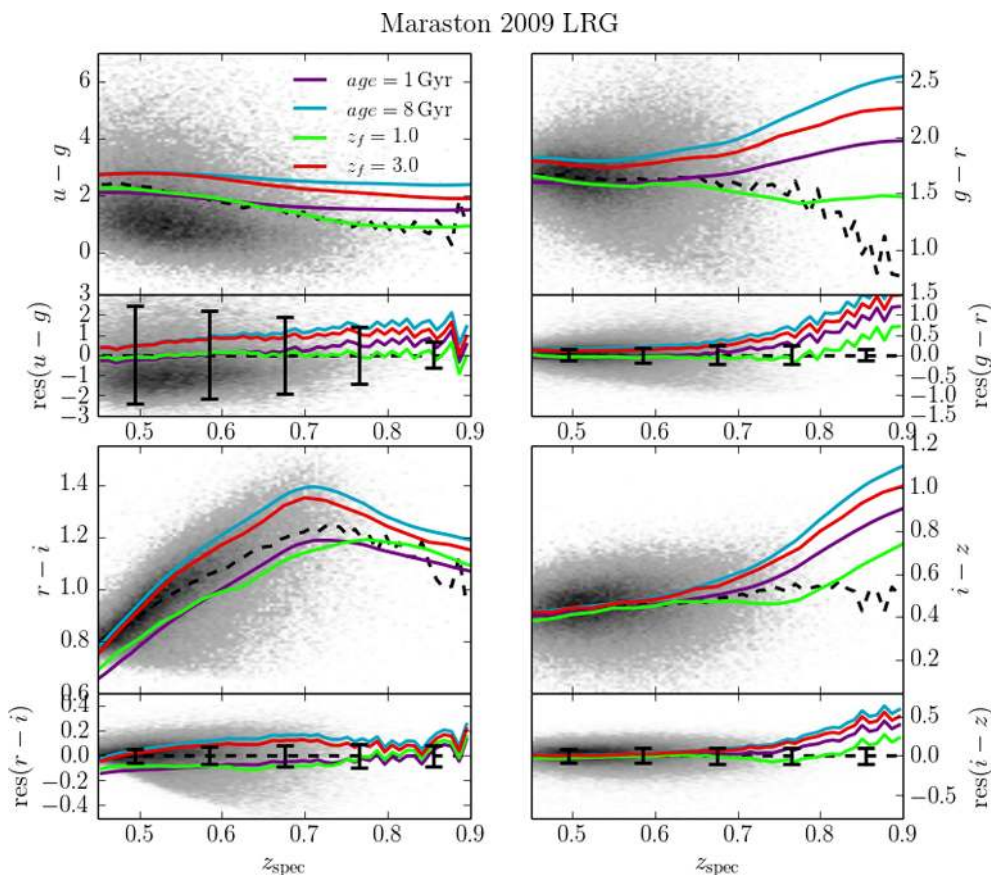
Figure 2. In the upper panels of the four major panels, the CMASS colours are plotted as a function of redshift in grey, where the median is shown by a black dashed line. The lower panels are normalized to the median colour and show the residuals. Error bars show the median data errors in five redshift bins. On top of that the predicted colours of the M09 are plotted for several ages and formation redshifts. The purple and blue lines show the M09 model at constant ages of 1 and 8 Gyr. Furthermore, the green and red lines show a passively evolving version of M09 at formation redshifts of $z_f = 1.0$ and 3.0.

extinction law of Calzetti et al. (2000). We perform the SED fitting procedure in such a way that signal-to-noise ratios smaller than 1 are considered upper limits.

The 4000 Å break is the most significant feature in the spectra of red galaxies and the reason why we can estimate photo-zs from broad-band photometry to a high accuracy. At redshifts populated by the CMASS galaxies the 4000 Å break lies within the r band, wherefore the g band is the bluest band needed to derive the position of the break in wavelength. The u band is very shallow and it may deteriorate the quality of the SED fitting results and the photometric redshifts if the errors are not determined accurately. For these reasons, we decide to omit it in the following SED fitting and photometric redshift estimation. Furthermore, the SED fits are not performed on the whole data of a z bin because of the high computation time, but we randomly select subsamples from within each z bin containing ∼2000 objects.

4.1.1 SED fitting results with BC03 models

We tried a variety of available model SEDs in the SED fitting procedure. The most extensive public libraries originate from Bruzual & Charlot (2003, hereafter dubbed BC03), Maraston (1998, 2005, hereafter M05), and Maraston & Strömbäck (2011, hereafter M11). We fit all of these models in separate trial runs in order to select a basic model set to proceed further, since large sets of models require enormous computation times.

The BC03 models can be generated by the software GALAXEV.[2] We create synthetical models from the BaSeL 3.1 library, using the Padova 1994 evolutionary tracks recommended by BC03 and the initial mass function(IMF) by Chabrier (2003). We generate models with four different metallicities, $Z = 0.2, 0.4, 1,$ and $2.5\,Z_\odot$. From these, we produce SSP models, as well as composite stellar population(CSP) models. SSPs assume that all stars are formed instantaneously in a delta-functional starburst at the birth of the galaxy (age zero) and are passively evolving afterwards. To create CSPs one can assign essentially any function for the star formation history (SFH), such that star formation takes place for longer periods of time. Usually, a star formation rate (SFR) is assumed that behaves proportional to $\exp(-t/\tau)$ (at least for galaxies at lower redshifts $z \lesssim 1$), where t is the age of the galaxy and τ is the e-folding time-scale (e.g. Shapley et al. 2005; Longhetti & Saracco 2009). We produce models with different τ values, $\tau = 0.5, 1.0, 3.0, 5.0, 8.0, 20, 50$ Gyr, where the latter simulate an almost constant SFR. We sample the models at ages ranging from 10 Myr to 13 Gyr. In order to exploit the maximum freedom available, we also create model SEDs with increasing star formation rate(SFR), hence $\tau < 0$. This kind of SFH is mostly important for high(er) redshifts ($z \gtrsim 2$) and yields more physical results (e.g. Maraston et al. 2010; Monna et al. 2014), but was used for SED fitting also at lower redshifts
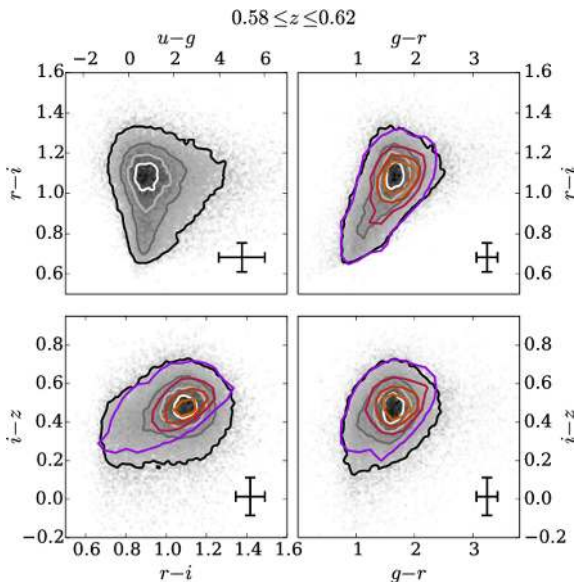
[2] http://www.cida.ve/~bruzual/bc2003   or   http://www.iap.fr/~charlot/bc2003.
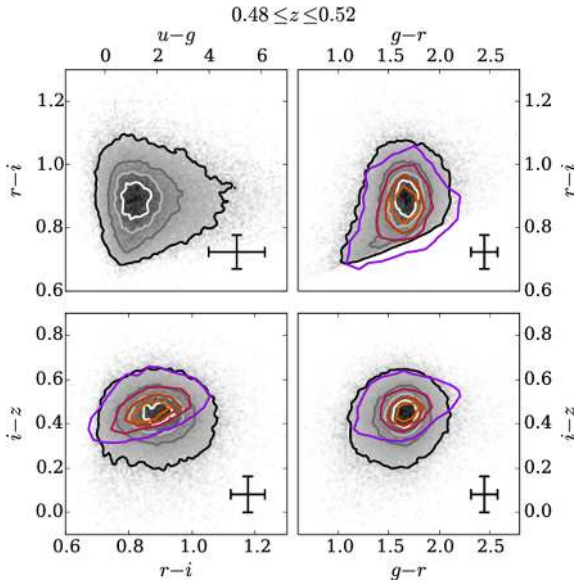
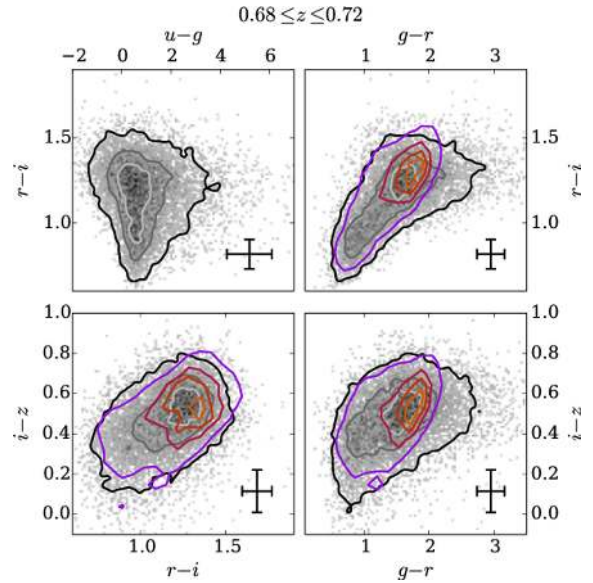**Figure 3.** Colour–colour plots for CMASS galaxies at $0.48 \leq z \leq 0.52$. The BOSS data are shown by grey shades and accompanying grey contours. Error bars denote the median error in colour in this redshift bin. The density of the resulting colours of the SED fitted unmodified BC03 models are shown by coloured contours. Both sets of contours are drawn at the same frequency levels, 5, 25, 45, and 65 per cent.

(e.g. Pforr, Maraston & Tonini 2012). In our case, the increasing SFR models do not significantly increase the range in colour space needed to match the CMASS galaxies that is why we omit them in the following to save computation time.

We create the same variety of CSPs for the M05 and M11 models (at least where possible) using the stellar population synthesis code EZGAL (Mancone & Gonzalez 2012), and fit them to the data as well. We show the distribution in likelihoods in Fig. A1 in Appendix A for the SED fits of BC03, M05, and M11 models to the CMASS data in the four redshift bins. The BC03 models outperform the M05 and M11 models in terms of their χ^2 values marginally at
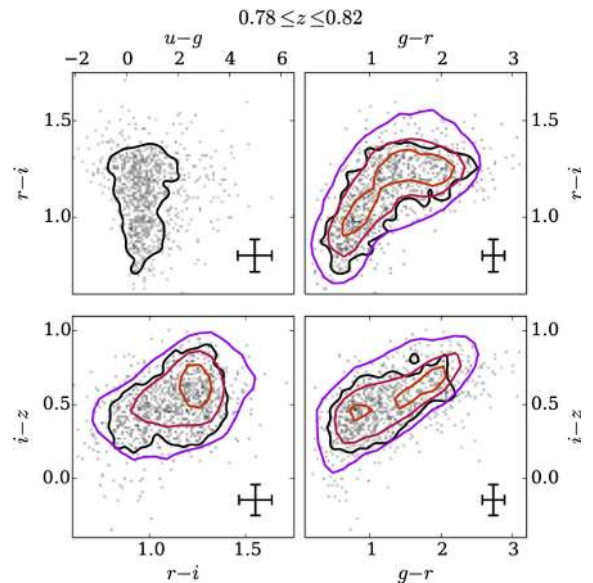
most. Fig. A1 shows that the fits with BC03 models have a slightly higher frequency of lower χ^2 values than M05/11. Furthermore, considering fig. 20 in G13, the BC03 models are a better match to the SDSS-II LRG data in terms of χ^2 values (this is not so evident here, as seen in Fig. A1). Because of the versatility in the model creation with GALAXEV, and the variety of provided ages and metallicities, we choose to use the BC03 models in the remainder of this work.

In Figs 3–6, we present the SED fitting results with BC03 SSP and CSP models with increasing and decreasing SFRs in colour space. The data colour is plotted in grey-scales with according contours, and the median colour errors are shown by error bars in each panel. The density of the predicted colours of the best-fitting models is

Figure 5. Colour–colour plots for CMASS galaxies and best-fitting BC03 models at $0.68 \leq z \leq 0.72$. See Fig. 3 for a detailed description.

Figure 4. Colour–colour plots for CMASS galaxies and best-fitting BC03 models at $0.58 \leq z \leq 0.62$. See Fig. 3 for a detailed description.

Figure 6. Colour–colour plots for CMASS galaxies and best-fitting BC03 models at $0.78 \leq z \leq 0.82$. See Fig. 3 for a detailed description. The contours had to be smoothed to make them visible due to the small sample size in this redshift bin.

plotted over the data distribution in coloured contours. Both sets of contours are plotted at the same frequency levels.

The bluer $i - z$ colour range populated by the CMASS galaxies in the redshift bins at $z \sim 0.5$. $z \sim 0.6$ and $z \sim 0.7$ are not populated by the best-fitting BC03 models. Additionally, the redder $g - r$ colours at $z \sim 0.7$ are also not matched by the BC03 models. All these offsets in colour cannot be accounted for by the photometric errors which are also given in Figs 3–6. We investigate the colour mismatch in the next section and modify the model SEDs to fit the data to a better extent.

4.1.2 Modification of the red continuum slope

In G13, we already showed in Fig. 19 that the predicted $i - z$ colour of the SDSS-II LRG sample cannot be matched by the models for the highest redshift bin ~ 0.4, at least not while the other three colours are fitted simultaneously. The model colours were too red in comparison to the data, which means that the decrease in the continuum slope redwards of the 4000 Å break of the model SEDs is not strong enough. The mismatch in $i - z$ is also present for the CMASS sample, which is located at even higher redshifts than the SDSS-II LRGs.

The slope of the continuum is changed by variations in the physical properties of the models which we investigate in the following paragraph. It is unlikely that the choice of the IMF could affect the model SEDs in such a way to produce the colour mismatch we observe. Changing the Chabrier IMF to a Salpeter (1955) or Kroupa (2001) IMF would only result in a change of the abundance of low-mass stars. These should not have a great impact on the continuum slope, since the light in this part of the spectrum is dominated by red giants. Optical colours are not sensitive to the choice of the IMF, which is (when stars are not resolved) often derived from spectroscopy or IR photometry (e.g. Cenarro et al. 2003; Conroy & van Dokkum 2012; Meidt et al. 2012). In turn, the choice of the particular IMF cannot cause the observed mismatch in colours. Different metallicities change the slope in the continuum as well, but we have considered sub-solar to supersolar values. Furthermore, we have exploited the model age and extinction as further parameters up to still reasonable values, i.e. 13 Gyr and $A_V = 3.0$. The burst component significantly affects only the bluer parts of the spectrum and has only marginal impact on the red part, which is dominated by the main stellar population, and not by the small (in total mass) burst fractions. Lastly, the star formation history(SFH) has to be considered. We analysed CSPs with nearly constant SFR and SSPs that create all their stars at one instant in time. We even considered exponentially increasing SFRs. If the too shallow red continuum slope would be a result of a poorly chosen SFH, we would at least expect that the data is bracketed by the considered models in colour space.

Since we cannot isolate physical model parameters that cause the colour mismatch, we can only assume that the issue arises due to inaccuracies in the modelling of the stellar evolution phase most difficult to follow theoretically, i.e. the asymptotic giant branch(AGB). The wavelength ranges covered by the colour mismatch hints at an incorrect modelling of AGB stars, since these dominate the SED in the red optical and NIR parts of the spectrum. E.g. Zibetti et al. (2013) show through NIR spectroscopy (comparing BC03 and M05 models in their fig. 1) that the chemical composition influences the slope of the model SEDs in the NIR, i.e. that these differences are caused by different C/O abundances. Furthermore, the different composition of dust in the circumstellar envelopes of carbon and
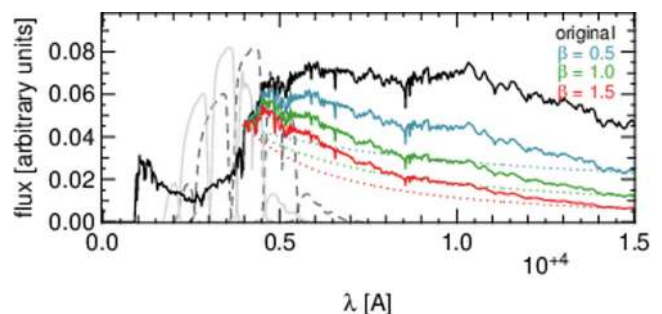
Figure 7. The black line is an LRG model SED from G13. The dotted blue, green, and red lines show the function $\lambda^{-\beta}$ with $\beta = 0.5, 1.0, 1.5$ for $\lambda \geq 4000$ Å. The solid lines are the above SED multiplied by $\lambda^{-\beta}$, also for $\lambda \geq 4000$ Å. In grey (dark grey dashed line and light grey solid line), the SDSS *ugriz* filter curves are shown for a galaxy at $z = 0.5$ and $z = 0.8$, respectively.

oxygen-rich AGB stars affect the efficiency of dust absorption and emission, which can therefore greatly influence the colours (e.g. Marigo et al. 2008; Cassarà et al. 2013; Salaris et al. 2014). Other explanations could be that the simple mixing-length theory models describing the convection in stars in the AGB phase are insufficient to reproduce the colours, or that incorrect assumptions on mass loss leads to the deviation.

A close investigation of the modelling of the AGB phase is needed to resolve the problem which is beyond the scope of this paper. However, we explain in the next paragraph how we modify the red continuum slope to better match the data.

The red SED continuum follows a function proportional to $\lambda^{-\beta'}$, heavily modified by absorption lines. Therefore, we can also change the steepness of the continuum by varying β'. This is done in such a way that we multiply the SED redwards of a wavelength λ_i with $a \cdot \lambda^{-\beta}$, where a is chosen such that the fluxes of the underlying model SED and the modified SED coincide at λ_i. In this way, we can change the continuum slope for $\lambda \geq \lambda_i$, wherefore the created SED on average follows $\lambda^{-(\beta+\beta')}$. We will term the such modified models '$\lambda^{-\beta}$ models' in the remainder of the text.

We exemplary show the resulting SEDs of an LRG model from G13 when modified by $a \cdot \lambda^{-\beta}$ for $\lambda \geq \lambda_i = 4000$ Å and $\beta = 0.5, 1.0, 1.5$ in Fig. 7. We also plot the SDSS filters as positioned in wavelength in the observed frame at $z = 0.5$ and $z = 0.9$, the approximate extreme redshift values of the CMASS sample. We can see that for $\lambda \geq 4000$ Å the change in slope mostly affects the i and z band at lower redshifts. For higher redshifts $\lambda_i = 4000$ Å has moved in the r band. Therefore, we create variations of the BC03 SSP and the CSP models with decreasing SFR explained above, with $\lambda_i \in [3000, 3500, 4000, 4500, 5000, 5500, 6000]$ Å such that the variation does affect the fluxes in different filters for similar redshifts. Furthermore, the β values span a range from 0.5 to 2.0 with a step size of 0.5. Larger ranges for values of λ_i and/or β do not improve the results further, since the location in colour spaces they would inhabit is already covered by the other models or not populated by the CMASS galaxies.

In Fig. 8, we show the colours of the CMASS data (grey). We furthermore plot the colours of an exemplary SED from G13 (red), and the colours of the same model when modified by $\lambda^{-\beta}$, for $\beta = 0.5$ and 1 (green and blue). The data in Fig. 8 are split into four redshift bins, and we plot the mean colour and corresponding root mean square values of each bin. The points are colour-coded in such a way that they darken with increasing redshift, which is shown by
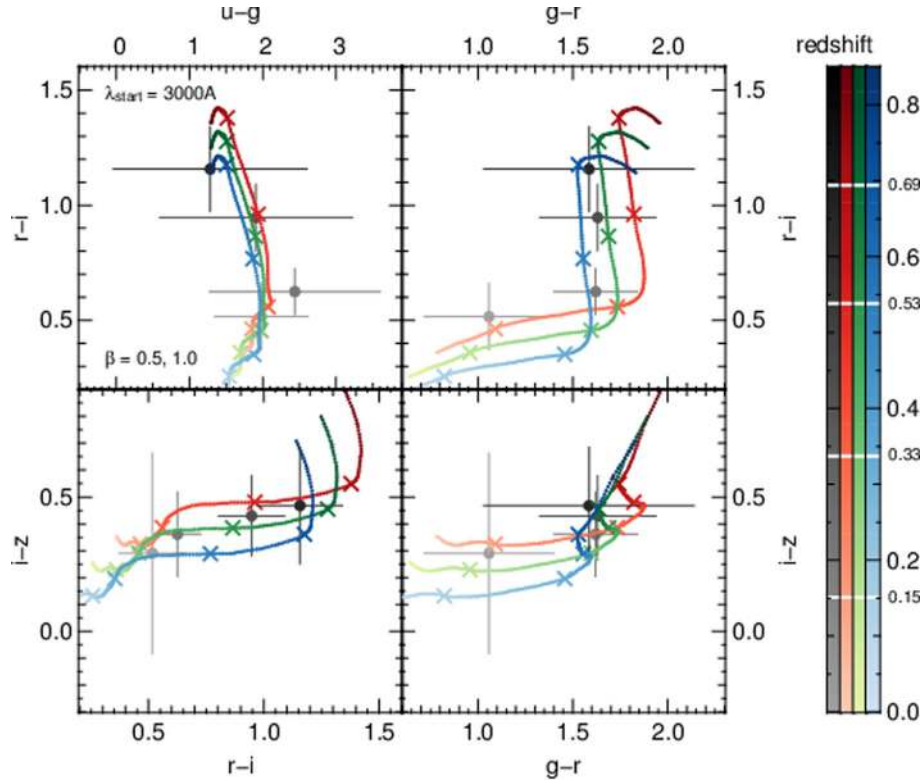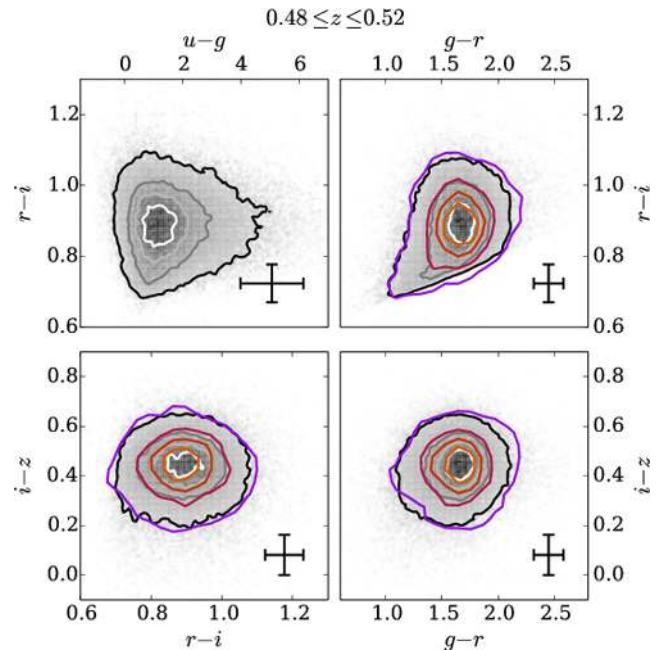
Figure 8. Colour–colour plots of the CMASS data and model SEDs. The BOSS data are split into four equally sized redshift bins. From these bins, we calculate the mean colour and plot it as points in the panel. The error bars denote the accompanying root mean square values of the colour in each bin. The points and the error bars are colour-coded with redshift (see the colour bar on the right, where white lines indicate the mean redshift in each bin). The red shaded track shows the colours of the model SED of Fig. 7 with varying redshift, where the crosses are at the same z as the data. Again, the redshift of the track is colour-coded and shown in the colour bar. The green and blue shaded tracks originate from the same model, when multiplied by $\lambda^{-\beta}$ with $\beta = 0.5, 1$ for $\lambda \geq 3000$ Å. A more detailed description is given in the text.

the grey colour bar on the right of Fig. 8, where white lines denote the mean redshifts within the bins. On top of the data, the colour tracks as functions of redshift are plotted for the models. The model colours at the z values of the data are highlighted by crosses, and the tracks in the colour spaces are again colour-coded as a function of redshift. The track of the original model is plotted by varying intensities of red, whereas those of the corresponding $\lambda^{-\beta}$ models are plotted in green shades for $\beta = 0.5$, and blue shades for $\beta = 1$, respectively. All plotted models have $\lambda_i = 3000$ Å. While the data colours (grey) are well matched by the original model (red track) at lower redshifts, the deviations from the mean colour increase with increasing z. The modification by $\lambda^{-\beta}$ produces a better agreement with the mean data colours. The value of β necessary to fit the data best is somewhat ambiguous in this plot, but has to be decided individually for each galaxy together with the best-fitting model.

4.1.3 SED fitting results with modified BC03 models

We now introduce the modified versions of the BC03 models together with the original SSPs and CSPs with decreasing SFHs from Section 4.1.1 in the SEDFIT code with the same fitting parameters as in Section 4.1.1. Figs 9–12 show the SED fitting results in colour spaces of the CMASS data in the four z bins. We see that the contours of the data and the best-fitting models (which are located at the same steps in frequency) almost coincide in Figs 9–11.

Comparing Figs 3–6 with Figs 9–12, we can observe an improvement on the match between the locations in colour space populated by the best-fitting models and the data. This is additionally

Figure 9. Colour–colour plots for CMASS galaxies at $0.48 \leq z \leq 0.52$ (analogous to Fig. 6). The BOSS data are shown by grey shades and accompanying grey contours. Error bars denote the median error in colour in this redshift bin. The density of the resulting colours of the best-fitting BC03 and BC03 $\lambda^{-\beta}$ models are shown by coloured contours. Both sets of contours are drawn at the same frequency levels, 5, 25, 45, and 65 per cent.
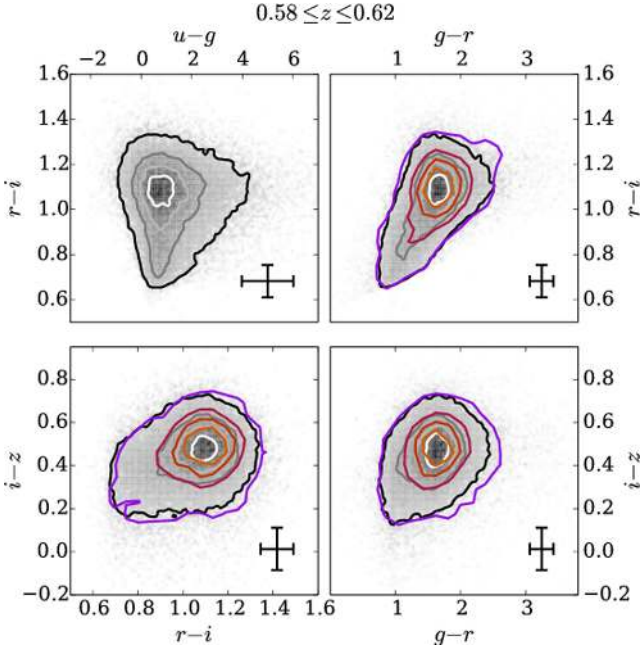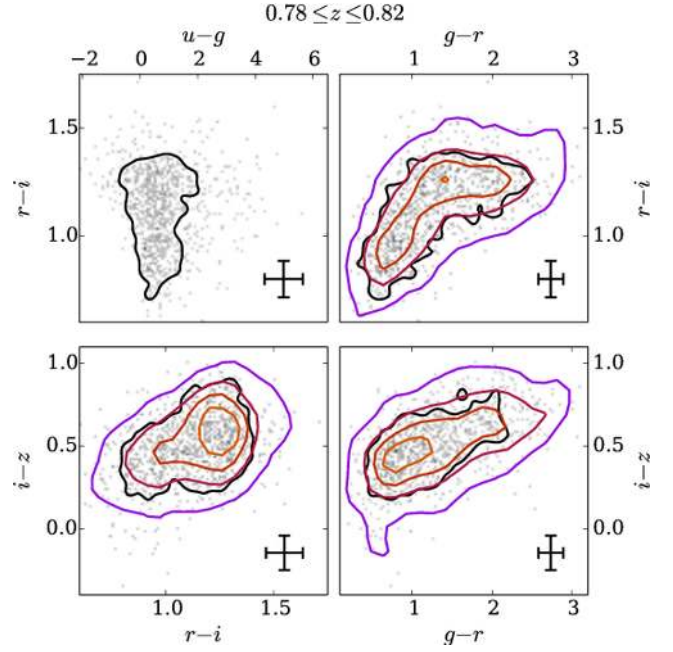
Figure 10. Colour–colour plots for CMASS galaxies and best-fitting BC03 $\lambda^{-\beta}$ models at $0.58 \leq z \leq 0.62$ (analogous to Fig. 4). See Fig. 9 for a detailed explanation of the plot.
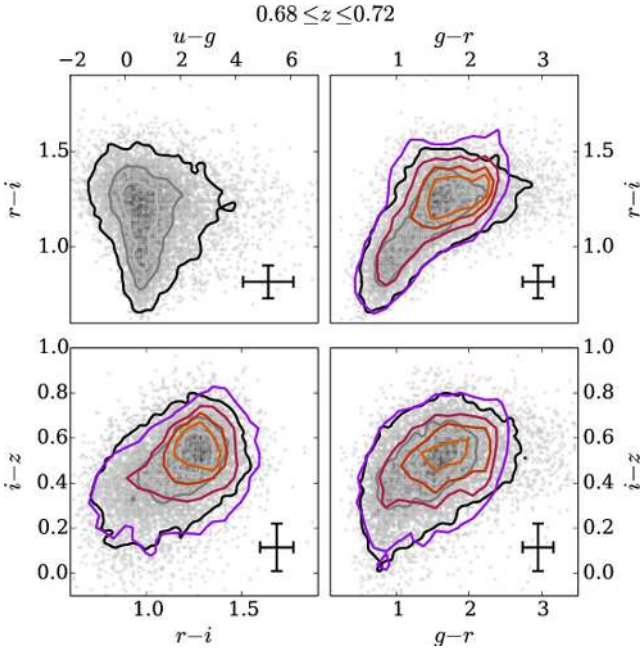

Figure 11. Colour–colour plots for CMASS galaxies and best-fitting BC03 $\lambda^{-\beta}$ models at $0.68 \leq z \leq 0.72$ (analogous to Fig. 5). See Fig. 9 for a detailed explanation of the plot.
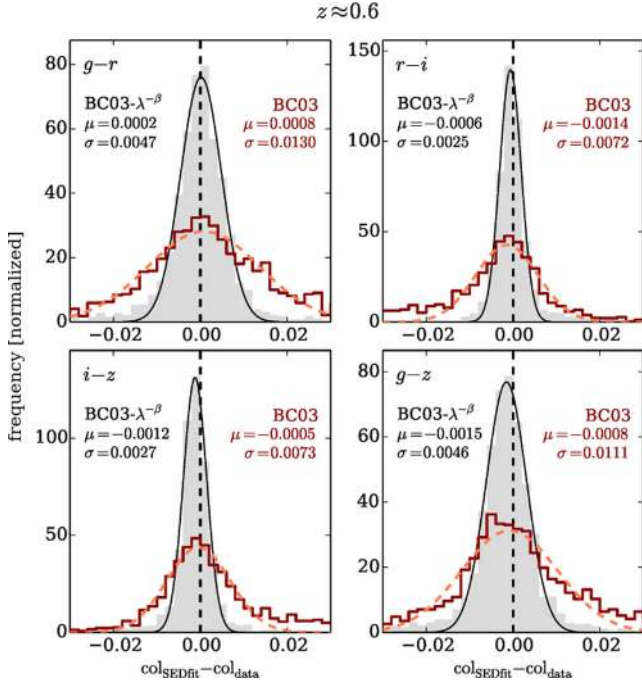

Figure 12. Colour–colour plots for CMASS galaxies and best-fitting BC03 $\lambda^{-\beta}$ models at $0.78 \leq z \leq 0.82$ (analogous to Fig. 6). See Fig. 9 for a detailed explanation of the plot. The contours had to be smoothed to make them visible due to the small sample size in this redshift bin. This is done in the same way as in Fig. 6.
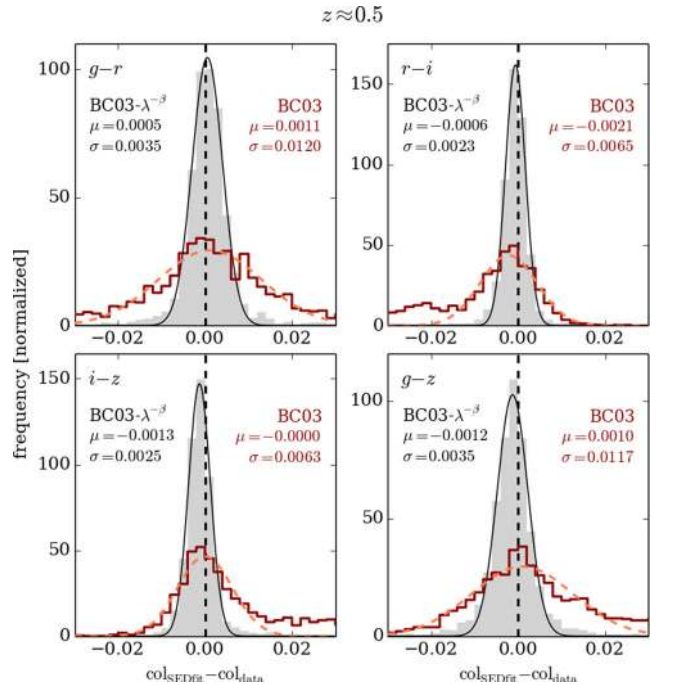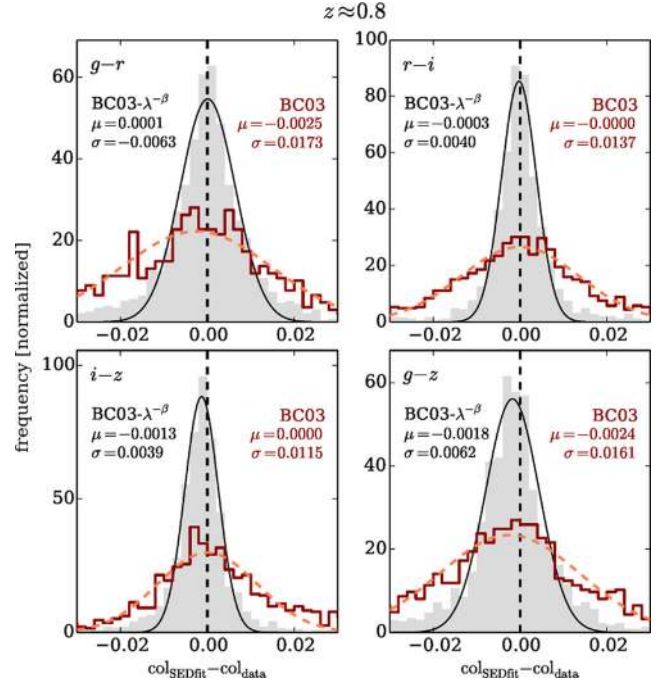

Figure 13. Deviations in $g - r$, $r - i$, $i - z$, and $g - z$ colours for BC03 models and BC03 $\lambda^{-\beta}$ models from the data within $0.48 \leq z \leq 0.52$. The Δcol distribution for the BC03 $\lambda^{-\beta}$ models is plotted in grey, and the best-fitting Gaussian is indicated by a solid black line. The same distribution, but for unmodified BC03 models is given by a dark red line, where we plot the best-fitting Gaussian by a dashed orange line. The black dashed vertical line highlights Δcol $= 0$. The best-fitting parameters of the Gaussian curves are given in the plots.
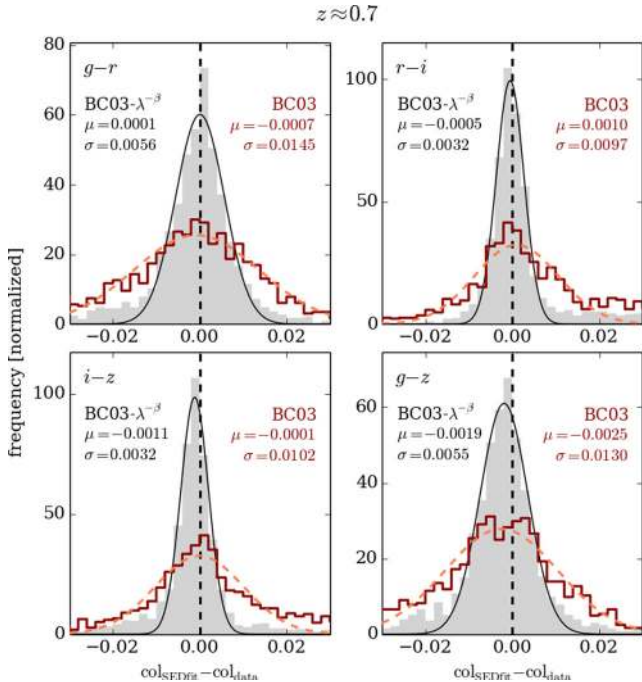
confirmed by Fig. A1 in Appendix A, where we see that the resulting χ^2 values of the fits are more frequently found at lower χ^2 values for the BC03 $\lambda^{-\beta}$ models than for the original ones, hence improving the goodness of fit. Here, we want to specifically analyse the offsets in the $g - r$, $r - i$, $i - z$, and $g - z$ colours of the best-fitting BC03 models and their modified variations from the data. We show in Figs 13–16 the deviations of the model colours, predicted by the best-fitting SEDs, to the data, Δcol $=$ col$_{model}$ $-$ col$_{data}$, for both

Figure 14. Deviations in $g - r$, $r - i$, $i - z$, and $g - z$ colours for BC03 models and BC03 $\lambda^{-\beta}$ models from the data within $0.58 \leq z \leq 0.62$. A detailed description of the plot is given in Fig. 13.

Figure 15. Deviations in $g - r$, $r - i$, $i - z$ and $g - z$ colours for BC03 models and BC03 $\lambda^{-\beta}$ models from the data within $0.68 \leq z \leq 0.72$. A detailed description of the plot is given in Fig. 13.

setups and in all four considered z bins. The distribution in Δcol of the original BC03 models (SSPs, as well as decreasing and increasing SFR CSPs) is presented by red histograms, where the Gaussian curve best-fitting the histogram is shown by a dashed orange line. The colour offsets of the BC03 $\lambda^{-\beta}$ models are plotted by a grey histogram, with an according Gaussian drawn in black. Δcol = 0 is

Figure 16. Deviations in $g - r$, $r - i$, $i - z$, and $g - z$ colours for BC03 models and BC03 $\lambda^{-\beta}$ models from the data within $0.78 \leq z \leq 0.82$. A detailed description of the plot is given in Fig. 13.

shown by a dashed black line, and we print the parameter values of the Gaussians in the plots.

While the mean deviations presented in Figs 13–16 are more or less the same in all considered redshift bins, the standard deviation of Δcol is about 2–3 times higher for the original BC03 SEDs. Furthermore, the BC03 $\lambda^{-\beta}$ distributions of Δcol resemble a Gaussian much more closely than that of the original BC03, especially in the $r - i$ and $i - z$ colours, but also in $g - z$. We already pointed out that the $i - z$ colour is overestimated by the BC03 models analysed in G13 for $z \sim 0.4$. This is still true for $z \sim 0.5$–0.7, observable in Figs 13–15 (cf. Figs 9–11). Concurrently, the predicted $r - i$ colours of the unmodified BC03 models are too blue up to $z \sim 0.6$, and too red in $z \sim 0.7$. In contrast to that, the colours of the BC03 $\lambda^{-\beta}$ SEDs yield very symmetric distributions, although with increased flanks in comparison to a Gaussian for higher redshifts.

In summary, we investigated two additional degrees of freedom, apart from the model SED parameters metallicity, SFH, age, burst, and dust, by modification of the red continuum slope through multiplying with $\lambda^{-\beta}$ at $\lambda \geq \lambda_i$ to match the colours of the CMASS galaxies to a better extent. When fitted to the data, the $\lambda^{-\beta}$ SEDs predict colours that deviate less from the data than the unmodified models they originate from, and yield lower χ^2 values. We will use the variety of best-fitting SEDs (including the additional burst and dust components) as a basis to select from when we define a model set for photometric redshift template fitting codes.

4.2 Selection of best-fitting SEDs for the new template set

We want to select a set of model SEDs from the best-fitting models of the previous section that represent the data in terms of colours for each redshift bin. The space we can construct from the SDSS colours is many-dimensional, and we want to reduce the dimensions for simplicity of the selection process. Therefore, we decided that the target space should have two dimensions, a compromise
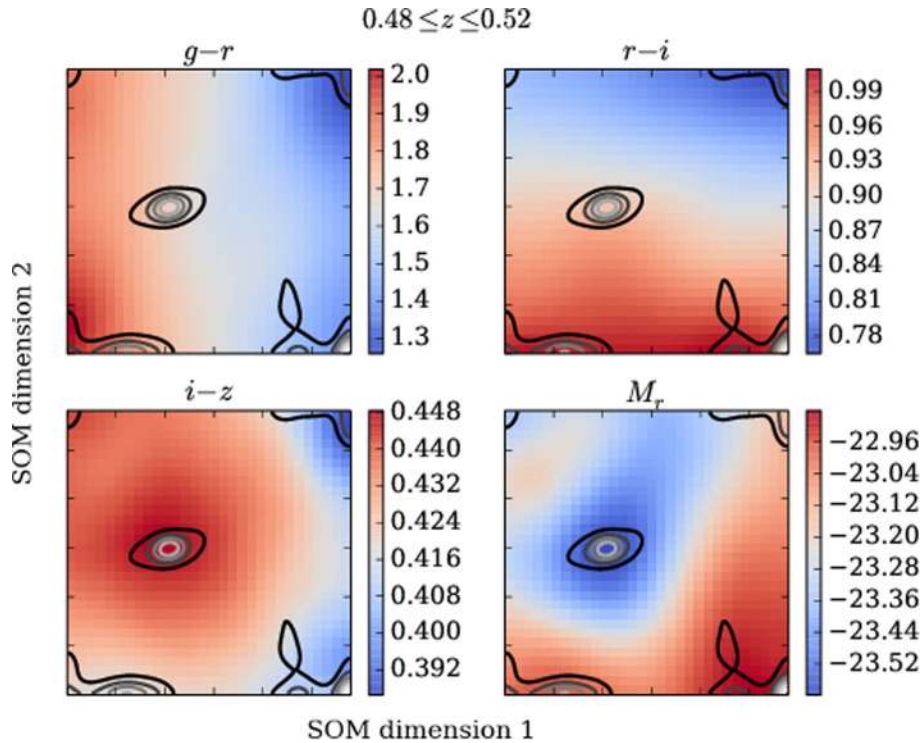
Figure 17. SOM with 30 × 30 bins for $g - r$, $r - i$, $i - z$, and M_r for galaxies within $0.48 \leq z \leq 0.52$, where the parameter values are colour-encoded (see colour bars on the right of each panel). M_r is computed by fitting templates from G13 to the spectroscopic redshift. Contours show the number density of the data in the map (note that the contours are smoothed to improve the clarity of the figure) and are drawn at six equally distributed frequency levels from ∼ 15 (black) to ∼ 85 per cent (white).

between the loss of information (which is greater for less dimensions) and the simplification through the reduction of dimensions. We could, in principle, perform a principal component analysis(PCA), and reduce the dimensions by concentrating on the space which is spanned by the first two eigenvectors of the PCA that have the highest variance. Another possibility for the reduction of dimensions is a self-organizing map (SOM or Kohonen-map; Kohonen 1982, 2001). An SOM is an artificial neural network that provides a discrete representation of a set of higher dimensional data values in a lower dimensional space (most often two-dimensional). The network is trained using unsupervised learning to map the data on to the surface in such a way that data points with similar properties (i.e. data values) are located in close neighbourhoods. Unlike a PCA, where the neglection of the third and later components yields a complete loss of the information carried by them, the two-dimensional SOM still entails this information in its points. This is why we chose an SOM over a PCA for the mapping on to a two-dimensional plane in the model selection below.

We create maps of the galaxy catalogues from the $g - r$, $r - i$, and $i - z$ colours of the four considered redshift bins. Furthermore, we can easily introduce also the absolute magnitudes M of the galaxies as a forth quantity to be mapped, since M can also hold information about the colour (e.g. Baldry et al. 2004). We have to take care of the errors in the data by normalizing the colours to their mean value and dividing through the colour errors determined by the uncertainties in the photometry. The absolute magnitudes of the galaxies are calculated by fitting the LRG model SEDs of G13 to the data at their spectroscopic redshifts. From the best-fitting models, we extract the absolute magnitude in the SDSS r band, which has to be normalized to a standard normal distribution to be comparable

to the colour values. Once this is done we train the SOM and create a surface with 30 × 30 discrete x- and y-values with assigned input values (i.e. colours and M_r). The positions of data points in the SOM are then determined by performing a nearest neighbour search.

We use an SOM algorithm implemented in python and provided in the PYMVPA package (Hanke et al. 2009). Fig. 17 presents the SOM trained on the catalogue with $0.48 \leq z \leq 0.52$ on the $g - r$, $r - i$, and $i - z$ colours and on M_r. The four panels are representations of the same SOM, but with the values of the four mapped parameters $g - r$, $r - i$, and $i - z$ colours and M_r encoded in colour. The density of the mapped underlying data is shown by contours. We can see from Fig. 17 that the extreme values of $g - r$ and $r - i$ have the greatest separation in the map, which is due to these colours having the largest spread in values. This is not a result of measurement errors, since we took these into account through the normalization of the colours explained in the last paragraph. We can see for example that the dense region in the middle of the panels corresponds to a higher concentration of the data at the respective colour values. These are $g - r \sim 1.7$, $r - i \sim 0.9$, and $i - z \sim 0.45$, which represent also the areas of highest density in the colour–colour plots of Fig. 9.

As previously mentioned, we aim to select a sample of model SEDs from the best-fitting models of the last section, where we want to take these density variations in the population of the data into account. Therefore, we want to identify clusters in the mapped data to select a model from each cluster that should represent the galaxies within the same cluster cell in terms of colour and absolute r-band magnitude. To perform the cluster search we employ a k-means clustering algorithm (Steinhaus 1957) that partitions the two-dimensional maps into cluster cells. Each cell is a Voronoi bin
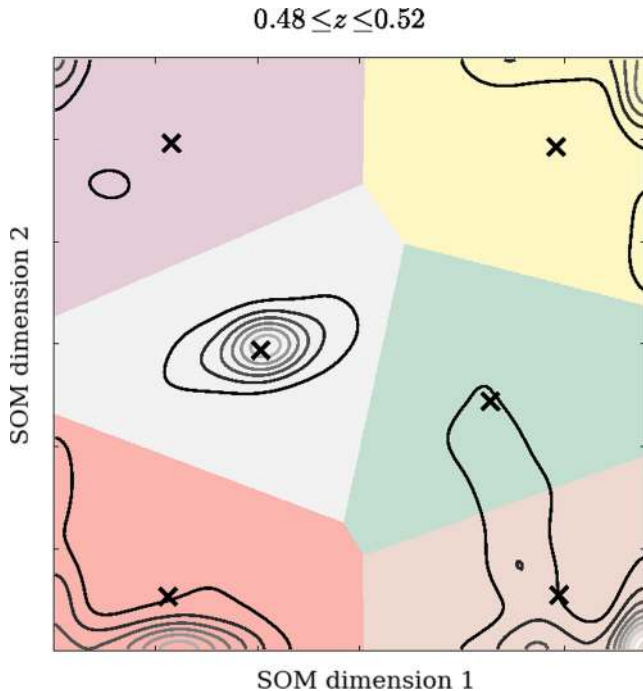
$$0.48 \leq z \leq 0.52$$

Figure 18. *k*-means clustering of the SOM for $0.48 \leq z \leq 0.52$ from Fig. 17. The Voronoi bins are highlighted by different colours, and cluster centres are marked by black crosses. Contours show the number frequency of the data (note that the contours are smoothed to increase the clarity of the figure) and are drawn at 10 per cent intervals between 10 per cent (black) and 90 per cent (white).

(Voronoi 1908), where two bins or clusters are separated from one another by a border orthogonal to the line connecting the cluster centres. Every data point belongs to the cluster with the nearest centre. Specifically, we use the python *k*-means clustering algorithm included in the SCIKIT-LEARN package (Pedregosa et al. 2011).

Fig. 18 presents the bins of the *k*-means clustering algorithm with $k = 6$ clusters applied to the SOM of Fig. 17, where the cluster centres are indicated by black crosses. The density of the data in the SOM is shown by contours in the plot. The data exhibit a number of clusters in all redshift bins which does not exceed six, wherefore $k = 6$ was chosen.

We want to select a set of SEDs from the best-fitting model SEDs which we will use in the following as template set for photo-*z*s. Therefore, for each cluster bin we take all models into account whose corresponding data points also lie in the same bin. Then, for each of these models separately, we estimate photometric redshifts on all objects within the cell and determine resulting quality parameters, i.e. mean error, scatter and catastrophic failures of the photo-*z*s. For each cluster, we chose a model that yields the best photoz results in terms of these parameters. Thereby, we have for each of the four redshift bins six models, hence 24 models in total. The such selected SEDs will be used in the following to estimate photo-*z*s.

5 PHOTOMETRIC REDSHIFTS

In this section, we analyse the photometric redshift results we get using our novel templates in combination with the PHOTOZ code on the whole CMASS sample detailed in Section 3. Afterwards, we compare our photo-*z*s with the photometric redshifts provided by the SDSS data base.

## 5.1 Photometric redshifts with the novel template SEDs

To introduce the models created in the previous section into the PHOTOZ code, we define the *z* prior such that μ_z of a model SED is the central value of the *z* bin of the catalogue from which the model originates, wherefore $\mu_z = 0.5, 0.6, 0.7, 0.8$. As σ_z, we set a default value of 0.2 which leads to a prior $P_z(z|T)$ function wide enough to avoid focusing effects at certain redshift bin centres, while ensuring a smooth transition between them. The resolution in redshift of a photoz run is 0.002 in the range of $z \in [0.0, 3.0]$. The allowed redshift range is much wider than that which is populated by the CMASS galaxies, such that we can analyse if the photoz accuracy is diminished by z_{phot} values that are highly over- or underestimated. This is done because we would like to be able to run photoz codes with the new models on catalogues with galaxies from larger redshift ranges and with more variations in SED type. We therefore want to make sure that small errors in z_{phot} are not due to a restriction in the redshift range.

To improve the priors by adapting them iteratively, we analyse the outcome for subsamples of objects that are fitted best by a specific model. Thereby we can adjust the redshift and luminosity priors for each model SED in order to reduce outliers and bias. Essentially, we decrease the value of σ_z whenever a model yields lower z_{phot} accuracies for redshifts further away from its bin centre. Furthermore, we allow μ_z to vary if the photoz performance of a specific model can be enhanced. If we observe that a specific template provides very bad redshifts which cannot be resolved by adjusting the respective prior, we omit these models completely in following runs. This is mostly the case for models created from the highest two redshift bins. Since we chose the models only on account of their photoz performance on a redshift bin, they might still yield a bad estimate in redshift ranges outside the bin. The resulting model set then consists of nine SEDs with adapted redshift priors. The luminosity priors where set initially to $\mu_M = 20.0$, $\sigma_M = 6.0$, and $p_M = 6.0$, to allow for a wide range of higher luminosities. The high exponent leads to a very flat functional behaviour within σ_M, and to steep decreases in $P_M(M)$ at $\mu_M \pm \sigma_M$. Adjusting $P_M(M)$ in the iteration is not necessary, since we cannot detect outliers which could be avoided through a different luminosity prior.

The model SEDs are shown in Fig. 19, where we plot them in the wavelength range covered by SDSS at the redshifts of our catalogue $0.45 \leq z \leq 0.9$. In the lower panel, we present the redshift prior parameters μ_z and σ_z, which correspond to the model SEDs. Fig. 20 shows the colours of the nine models as a function of redshift. With the nine models, we can account for the large spread of the data in most cases. Furthermore, we cover also the bluer parts in $g - r$, $r - i$, and $i - z$ (cf. Fig. 2). For lower redshifts, our models produce colours that cover only the bluer $r - i$ and $i - z$ ranges of the data. This is not because the BC03 $\lambda^{-\beta}$ do not fit the colours of the data (cf. Figs 9–12), but accidental, since the models were selected (from within their cluster bins) on account of their photoz performance in Section 4.2. The model SEDs with designations 7–9 cover only peculiar blue colours of the data. These models were created on the basis of the higher *z* samples, which is why they match the data better at higher redshifts (especially the $i - z$ colour). The bluer colours of the models can only be observed from a small number of galaxies. Therefore, when all models are fitted to the spectroscopic redshift (but also in the photoz estimation below), the models 7–9 are best fitting only for $\sim$1–5 per cent of galaxies.

The physical parameters and the λ_i and β values of the nine model SEDs are summarized in Table 1. They explain the behaviour of the SEDs. From Fig. 19, we see that the model SEDs roughly follow
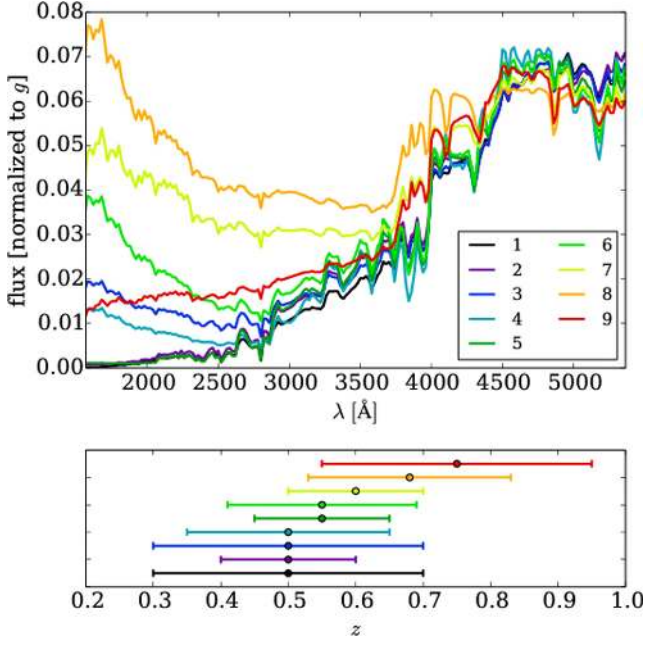
Figure 19. Upper panel: the nine surviving model SEDs within the range of the SDSS *ugriz* filter system at redshifts $0.45 \leq z \leq 0.9$ (the z range of our galaxy sample). Lower panel: corresponding z prior parameters, μ_z and σ_z. The colour code is the same as in the upper panel.

a trend and become bluer with increasing redshift (except for the red coloured SED with $\mu_z = 0.8$), which is mirrored in the values of Table 1. The e-folding time-scale τ roughly increases, as well as the burst fraction, making the resulting SEDs bluer on average. The red highlighted SED has lower fluxes than the orange and the yellow-green SEDs in the UV part of the spectrum, not continuing the sequence. It is redder because of the high extinction values and because of its (relatively) high metallicity, $Z = Z_{\odot}$.

In summary, although we can see a qualitative trend in the SEDs as a function of redshift prior (which originates from the spectroscopic redshifts of the underlying z bin with small adaptation), the trends in the physical parameters are not that evident. This is because they are degenerate and changes in one parameter can yield similar results in the SEDs as a variation in another parameter (e.g. the well-known age–metallicity degeneracy).

Before analysing the photoz performance of the new models, we want to introduce several metrics which we use to assess the photometric redshift quality. We decide to provide a large number of metrics to enable the reader to compare with other publications. The photometric redshift error is $\Delta z = z_{\mathrm{phot}} - z_{\mathrm{spec}}$, and $\Delta z_{\mathrm{rf}} = \Delta z / (1 + z_{\mathrm{spec}})$ in the rest frame. Catastrophic outliers are defined such that $|\Delta z_{\mathrm{rf}}| \geq 0.15$ (cf. Ilbert et al. 2006). The mean errors are characterized by the bias $\langle \Delta z \rangle$ and the mean absolute error $\langle |\Delta z| \rangle$, as well as by their corresponding values in the rest frame, $\langle \Delta z_{\mathrm{rf}} \rangle$ and $\langle |\Delta z_{\mathrm{rf}}| \rangle$. The root of the sample variance is denoted $\sigma(\Delta z_{\mathrm{rf}})$, and $\sigma_{68,\mathrm{rf}}$ is the half of the width of the Δz_{rf} distribution where 68.27 per cent of the sample is located,
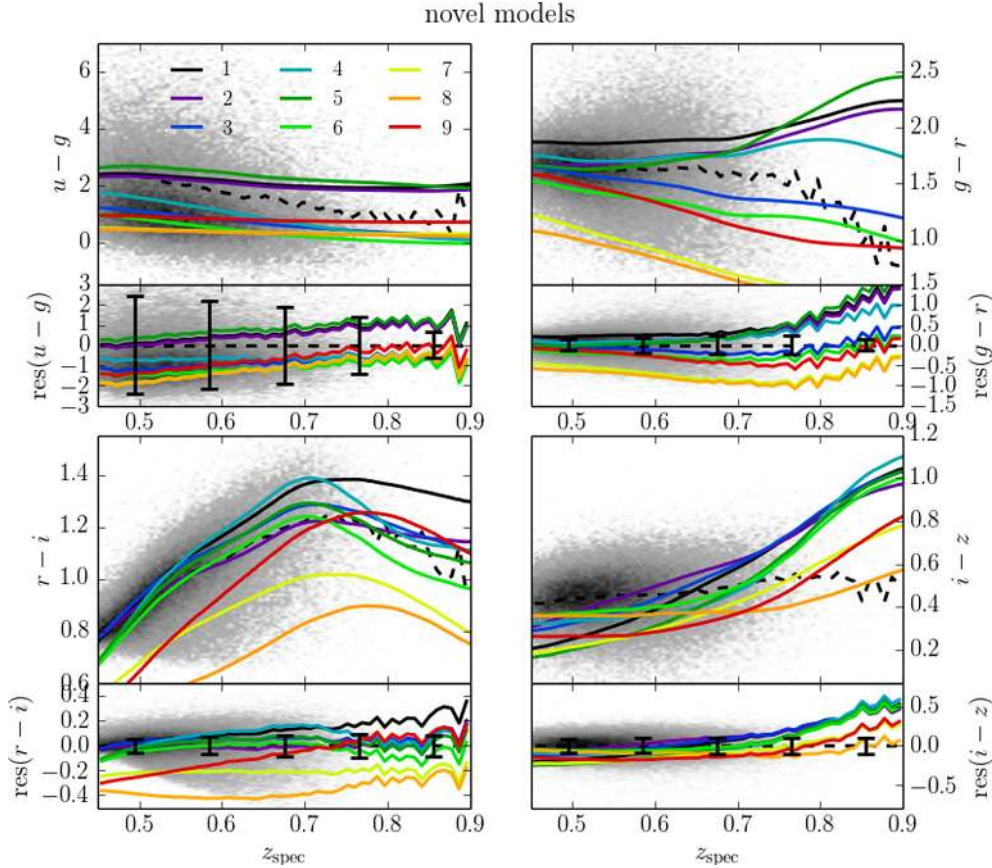
Figure 20. In the upper panels of the four major panels, the CMASS colours are plotted as a function of redshift in grey, and the median is shown by a black dashed line. The lower panels are normalized to the median colour and show the residuals. Error bars present the median data errors in five redshift bins. On top of that the predicted colours of the nine selected models are drawn. The colour code matches that of Fig. 19.

Table 1. Physical parameters of the nine surviving templates. The first column gives the numbering of the models and the plot colour according to Figs 19 and 20. The column 'burst' is the mass fraction of the burst, whereas $A_{V,\,\text{burst}}$ is the burst extinction.

Model #	Z	τ (Gyr)	Age(Gyr)	A_V	Burst	$A_{V,\,\text{burst}}$	λ_i (Å)	β
1 (Black)	0.008	SSP	2.0	1.2	–	–	5000	2.0
2 (Violet)	0.004	SSP	6.0	0.6	1.0 %	2.0	5500	1.5
3 (Blue)	0.008	1.0	6.0	1.0	1.0 %	0.0	5000	1.5
4 (Turquoise)	0.05	SSP	8.0	0.7	1.0 %	0.0	3500	2.0
5 (Dark green)	0.02	SSP	4.0	0.0	1.0 %	1.0	5000	1.5
6 (Green)	0.05	SSP	3.0	1.3	2.0 %	0.0	3500	2.0
7 (Yellow-green)	0.05	3.0	5.0	1.4	1.0 %	0.0	5000	2.0
8 (Orange)	0.004	50.0	4.0	2.7	1.0 %	0.0	3000	1.5
9 (Red)	0.02	3.0	4.0	2.2	2.0 %	2.0	4500	2.0

corresponding to a 1σ confidence interval. Finally, in terms of scatter we also calculate the normalized median absolute deviation $\sigma_{\text{NMAD}} \equiv 1.48 \cdot \text{median}(|\Delta z_{\text{rf}}|)$ (Ilbert et al. 2006) which is calculated for non-outliers only, and gives a clue about the width of the Δz_{rf} distribution without regarding the flanks. We calculate the fractions $\eta_{2\sigma}$, $\eta_{3\sigma}$ of galaxies within $\langle \Delta z_{\text{rf}} \rangle \pm i \cdot \sigma(\Delta z_{\text{rf}})$, where $i = 2$, 3 (cf. Carrasco Kind & Brunner 2014). If Δz_{rf} behaved as a perfect Gaussian, then $2 \cdot \sigma(\Delta z_{\text{rf}})$ and $3 \cdot \sigma(\Delta z_{\text{rf}})$ would be the 2σ and 3σ confidence intervals and therefore hold 95.45 and 99.73 per cent of the objects. But the distribution of Δz_{rf} can be non-Gaussian and still yield reasonable values for $\eta_{2\sigma}$ and $\eta_{3\sigma}$ (compare Figs 22 and 27 later on).

To evaluate the precision of the photoz errors δz estimated by the code (equation 1) from the PDF (cf. Section 2), we introduce $\Delta z' \equiv \Delta z / \delta z$. If the errors are estimated correctly, the distribution of $\Delta z'$ should resemble a standard normal distribution (cf. Sánchez et al. 2014). We will therefore analyse the values of the mean $\langle \Delta z' \rangle$ and the standard deviation $\sigma_{\Delta z'}$ of the distribution of $\Delta z'$.

Lastly, we perform Kolmogorov–Smirnov (KS) tests on the distributions $N(z_{\text{phot}})$ and $P(z_{\text{phot}})$ (cf. Carrasco Kind & Brunner 2014), where the latter is the z_{phot} distribution derived by stacking the PDFs for all objects. The KS test value is the maximum absolute difference between the cumulative distribution functions $\int_{-\infty}^{z} P(\hat{z}) \mathrm{d}\hat{z}$ of the probability density of $N(z_{\text{phot}})$, or the stacked $P(z_{\text{phot}})$, to the cumulative distribution of $N(z_{\text{spec}})$.

After this introduction of the photoz quality metrics, we present the photometric redshift results of the CMASS sample with the novel template set and priors of Section 4.2 and Fig. 19 in Fig. 21. The upper left panel shows z_{spec} versus z_{phot}, the middle left panel shows the photometric redshift restframe error as a function of spec-z, and the lower left panel presents z_{spec} versus $|\Delta z_{\text{rf}}|$. We indicate the median Δz_{rf} and $|\Delta z_{\text{rf}}|$ values in the middle left panel by solid and dashed black lines, whereas in the lower panel we highlight the $\sigma_{68,\text{rf}}$ and σ_{NMAD} values as a function of the spectroscopic redshift also by solid and dashed black lines. Finally, the right-hand panel presents the normalized redshift distributions for z_{spec} (grey filled histogram), the distribution derived from the single-value photometric redshifts z_{mode} (grey line) derived by the mode of $P(z)$, and the stacked PDFs (black line). The photometric redshift quality metrics discussed above are printed in the plot as well.

The bias has a small positive value for the lowest considered redshifts, then is close to zero, and is decreasing to negative values for higher redshifts $z \sim 0.7$. The overall mean value is still positive due to the small sample size at higher redshifts, visible by the number density in Fig. 1 or the right-hand panel of Fig. 21. The scatter (lower left panel in Fig. 21) increases slightly with increas-

ing spec-z, while the value of σ_{NMAD} stays more or less the same. This means, that the outliers (which are not considered in the calculation of σ_{NMAD}) are predominantly responsible for an increase in $\sigma(\Delta z_{\text{rf}})$. The fraction of catastrophic outliers however is very small $\eta = 0.22$ per cent. In the right-hand panel of Fig. 21 we observe that the photoz predictions from the mode of the PDF yield deviations from $N(z_{\text{spec}})$ larger than for the case where the whole PDFs are considered. The excesses observed in $N(z_{\text{phot}})$ are due to the overestimation mentioned previously. They are not anymore visible when we use the posterior $P(z)$ distributions in the $N(z)$ reconstruction, which means that the PDF should be favoured for science analyses (cf. e.g. Mandelbaum et al. 2008; Hildebrandt et al. 2012).

As for the error estimation of the PHOTOZ code, we see from the values of $\langle \Delta z' \rangle$ and $\sigma_{\Delta z'}$, that $\Delta z'$ behaves very close to a standard normal distribution, which is the goal in the photometric redshift error predictions. This means that not only the approach for the calculation of δz is legitimate, but also that the models and priors create a reasonable $P(z)$ (from which δz is extracted). Table 2 in Section 5.2 holds a summary of the derived quality metrics, together with results of public photo-zs from the SDSS, which we will analyse in Section 5.2.

To investigate if the estimated errors are reliable in identifying outliers, we assume the null hypothesis that an object is a photoz outlier. The probability of an outlier being falsely classified as a non-outlier on account of δz is then the type I error α. Additionally, the type II error β gives the probability of a non-outlier being misclassified as an outlier by the estimated error. For the photo-zs of this work, we get $\alpha = 0.93$ and $\beta = 3.5 \times 10^{-4}$. This means that an outlier is falsely classified as a non-outlier with probability α, but also that a non-outlier is almost never classified as an outlier (β). Although $\sigma(\Delta z')$ is close to, but slightly smaller than, 1, the deviation is probably caused by to shallow peaks around the mode of $P(z)$ for outliers, such that their errors are underestimated. The α and β value are summarized together with the results of the SDSS (Section 5.2) in Table 3.

The distribution of the photometric redshift errors in the rest frame is shown in Fig. 22 by a grey filled histogram. In red we indicate the bias and the ranges of $\langle \Delta z_{\text{rf}} \rangle \pm i \cdot \sigma(\Delta z_{\text{rf}})$, where $i = 1$, 2, 3. Furthermore, we calculate the real interval widths where 95.45, and 99.73 per cent of the galaxies are located and introduce them in the plot (black dash–dotted and dotted vertical lines). Finally, we calculate the Δz_{rf} distribution from the stacked PDFs and fit a Gaussian to the histogram (green lines).

The peak of the Δz_{rf} distribution is slightly shifted to the right, which is again due to the overestimated photo-zs at the lowest redshifts. Concerning the resemblance to a Gaussian, we observe
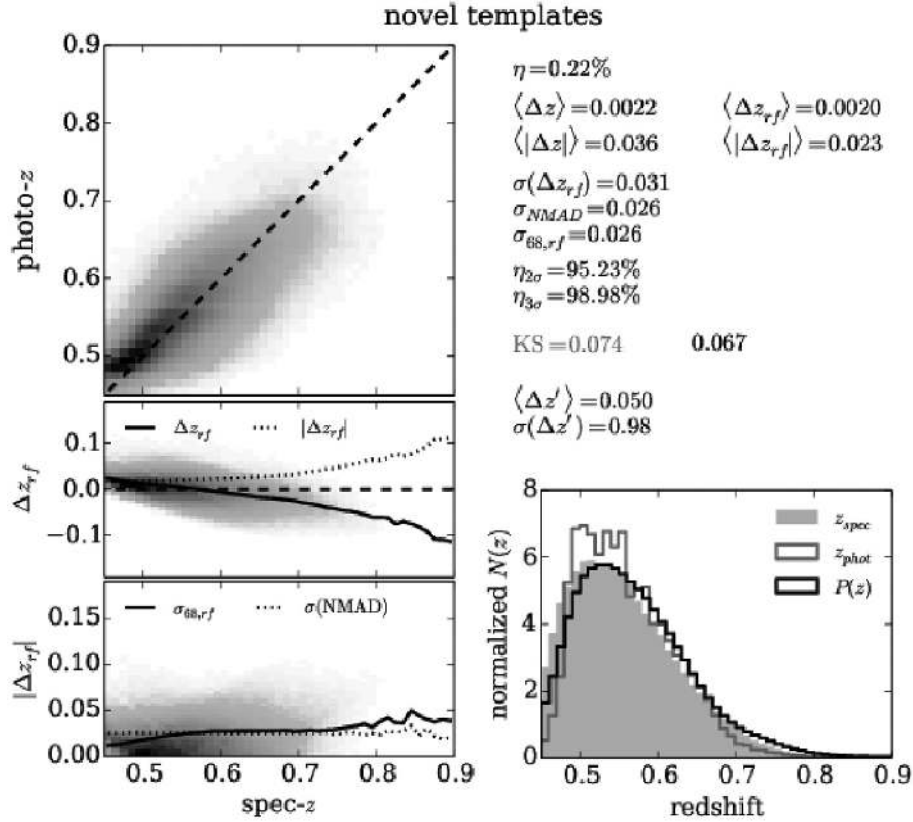
Figure 21. Photo-z results with novel templates, based on the BC03 $\lambda^{-\beta}$ models and selected through an SOM and a k-means clustering algorithm. Left-hand panels: the upper panel shows the spectroscopic versus the photometric redshift, where the dashed line is at $z_{phot} = z_{spec}$. The middle panel is the distribution of the restframe photoz error as a function of the spectroscopic redshift, where we plot the bias (solid line) as well as the mean absolute error (dashed line). Finally, the lower panel shows z_{spec} versus the absolute restframe error, where we plot σ_{68} (solid) and σ_{NMAD} (dashed) on top. Right-hand panel: the relative frequency distributions of z_{spec} (grey filled histogram), z_{phot} (dark grey line), and the stacked PDF of all objects (black line). Additionally, the photoz quality metrics are printed in the figure. We provide two values for the KS test, for the predictions of single z_{phot} values (from the mode of the PDF, grey), and for the PDF (black).

Table 2. Summary of photometric redshift quality metrics of the template fitting results with the novel templates and priors used with the PHOTOZ code, the SDSS photo-zs of the KF code, and the random forest code (SDSS-RF). Values for *KS* in the case of this work are derived from $P(z)$, whereas the values in brackets are calculated using only the z_{mode} predictions.

Setup	η (per cent)	$\langle\Delta z_{rf}\rangle$	$\langle|\Delta z_{rf}|\rangle$	$\sigma(\Delta z_{rf})$	σ_{NMAD}	$\sigma_{68,\,rf}$	$\langle\Delta z'\rangle$	$\sigma_{\Delta z'}$	KS
This work	0.22	0.0020	0.023	0.031	0.026	0.026	0.050	0.98	0.067 (0.074)
SDSS-KF	0.91	-0.011	0.027	0.038	0.027	0.029	-0.54	1.5	0.082
SDSS-RF	3.0	-0.023	0.035	0.051	0.032	0.036	-0.27	0.61	0.14

from the values given in the plot that $\sigma_{95,rf} \approx 2\sigma(\Delta z_{rf})$ (which is why we cannot discern the lines in Fig. 22) and $\sigma_{99,rf} > 3\sigma(\Delta z_{rf})$. The latter is due to the outliers which can be scattered far from the spec-z value due to the allowed fitting range. The fractions $\eta_{2\sigma}$

Table 3. Significance test for outlier classification. The null hypothesis is that an object is an outlier, and α and β are the type I and II errors, respectively (rounded to two significance figures).

|	This work	SDSS-KF	SDSS-RF
α	0.93	1.0	0.86
β	$3.5 \cdot 10^{-4}$	$7.9 \cdot 10^{-4}$	$1.4 \cdot 10^{-2}$

and $\eta_{3\sigma}$ of objects within $2\sigma(\Delta z_{rf})$ and $3\sigma(\Delta z_{rf})$ have reasonable values. While $\eta_{2\sigma} = 95.23$ per cent is very close to the desired value of 95.45 per cent, $\eta_{3\sigma} = 98.98$ per cent is slightly smaller by only ~ 0.8 per cent. The Δz_{rf} distribution derived from the PDFs is slightly broader than the distribution of the individual photoz results. This is a consequence of the asymmetry of the PDFs that often have higher probabilities for redshifts greater than the most probable z.

Fig. 23 presents the deviations of magnitudes, predicted by the models in a photoz run, and the data. A Gaussian is fitted to the histograms whose best-fitting parameters are printed in the panels. For the r- and i-band magnitudes the Δmag distributions are very narrow, while the expectation values of the Gaussians are near zero. This is thanks to the depth of the photometry in these filters, which is greatest in i and r, and the photometry has the smallest measurement
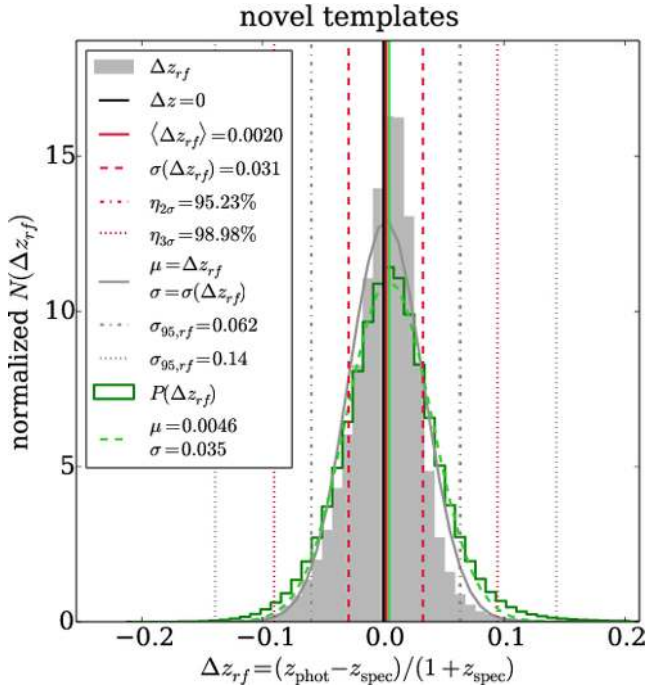
Figure 22. Photometric redshift restframe error distribution (grey filled histogram) estimated with the novel templates and priors. $\Delta z = 0$ is highlighted by a solid black line, whereas the bias is shown by a solid red line. The steps at $\Delta z_{rf} \pm \sigma(\Delta z_{rf})$, $2\sigma(\Delta z_{rf})$, and $3\sigma(\Delta z_{rf})$ are represented by dashed, dash–dotted, and dotted red lines, and the corresponding number fractions are given in the legend. We also plot a Gaussian with $\mu = \langle \Delta z_{rf} \rangle$, and $\sigma = \sigma(\Delta z_{rf})$ in grey. Furthermore, dash–dotted and dotted grey vertical lines show the real $\sigma_{95,rf}$ and $\sigma_{99,rf}$ intervals. Since $\sigma_{95,rf} \approx 2\sigma(\Delta z_{rf})$ we cannot discern the red and the grey dash–dotted lines. The Δz_{rf} derived from the stacked PDFs is shown by a green solid histogram, to which we fit a Gaussian highlighted by a dashed green curve.
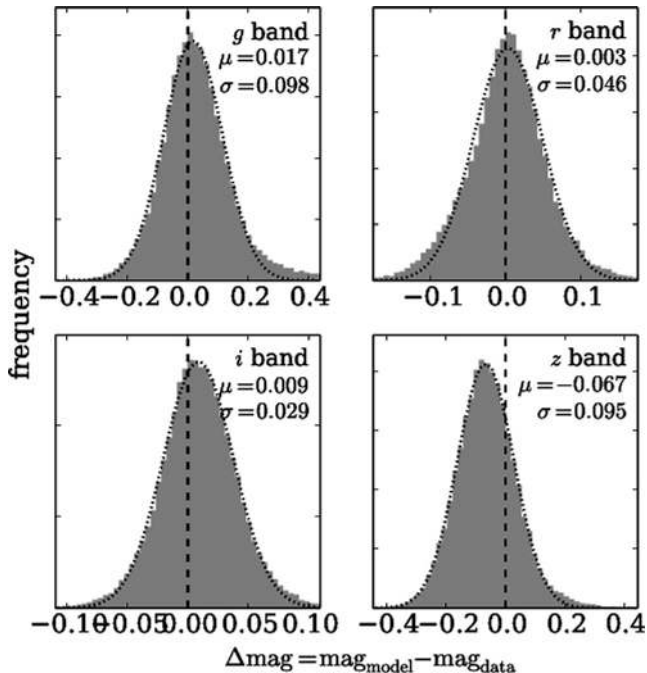
Figure 23. Deviations in magnitudes of the predicted magnitudes of the novel model SEDs to the data.

uncertainties. For the g- and the z-band magnitudes however, the broadness of the Δmag distributions increases, which is mostly an effect of the more shallow photometry in these bands, but also of the respective magnitude ranges of the galaxies. Furthermore, the mean magnitude deviations in these two bands are higher than in r and i. While μ of the best-fitting Gaussian is still small in comparison to σ for the g band, it is relatively high in the z band. We selected the models for the photoz estimation only on account of their photoz performance, and not on how well they match the data in terms of magnitudes. However, this does not necessarily imply that the BC03 $\lambda^{-\beta}$ models are not able to fit the data in terms of magnitudes. Indeed, calculating the average magnitude offsets of the SED fitting of Section 4.1.3, yield maximum values of $[0.6, 0.9, 0.3, 1.1] \cdot 10^{-3}$ in g, r, i, and z for the four redshift bins.

In the next section, we want to compare the photometric redshift results of this section with the photo-zs provided by SDSS.

5.2 Comparison to SDSS photometric redshifts

We present the photometric redshifts from the SDSS-III data base in comparison to the results of our code–template combination. The SDSS data base provides photo-zs from two different empirical methods. One code uses a k-d tree nearest neighbour fit to derive the redshifts (hereafter KF after Csabai et al. 2007). The SDSS-KF results for our catalogue are shown in Fig. 24, which is equivalent to Fig. 21 and also shows the quality metrics. From the middle panel of Fig. 24, we see that the location of the densest part of the population lies close to the $\Delta z = 0$ line for low redshifts $z \lesssim 0.6$. In all higher z_{spec} regions the photometric redshift is systematically underestimated which leads to the low negative value in the bias. An underestimation is also present in the results from Section 5.1, but it is slightly higher in the SDSS-KF case. From the lower panel of Fig. 24, we see that also here the rise in $\sigma_{68,rf}$ is mostly caused by outliers, which are not considered in the calculation of σ_{NMAD} (which stays more or less constant). Taking the estimated photometric redshift errors into account, the relatively high bias does not decrease, which is mirrored in the higher absolute value of $\langle \Delta z' \rangle$, but even still increases. Hence, the bad estimates of the photo-zs are not recognized by the error estimation. This can also be seen in the value of $\sigma_{\Delta z'}$ which is well beyond 1 and means that the errors are on average underestimated. This also affects the results of a significance test on outlier classification (cf. Section 5.1). The errors type I and II read (at two significant figures) $\alpha = 1.0$ and $\beta = 7.9 \times 10^{-4}$ (see also Table 3 for comparison with the other results), confirming that the errors are underestimated.

In terms of $N(z_{phot})$, the low photoz values at higher z_{spec} do not change the shape significantly, which is due to the low sample size in these regions, and results in the low value of KS. We cannot compare to the $N(z)$ results for PDFs, since they are not provided.

As before, we show in Fig. 25 the Δz_{rf} distribution. Comparing with our case from Fig. 22, we see that here the results are slightly shifted to the left, an effect due to the underestimated photo-zs at higher redshifts. Also in the case of the SDSS-KF redshifts, the $\sigma_{95,rf}$ and $\sigma_{99,rf}$ values do not perfectly coincide with $2\sigma(\Delta z_{rf})$ and $3\sigma(\Delta z_{rf})$, but are elevated. $\eta_{2\sigma} = 95.25$ per cent is very close to the desired value, while $\eta_{3\sigma} = 98.43$ per cent deviates by ~ 1.3 per cent.

In comparison to the redshifts, we get with our novel templates and code, the quality metrics for bias, error, and scatter are all higher in the case of the SDSS-KF photo-zs. Furthermore, the estimated redshift errors are more reliable in our code–template combination. The KS test yields better results in the case of this work, since the SDSS-KF photo-zs are systematically underestimated.
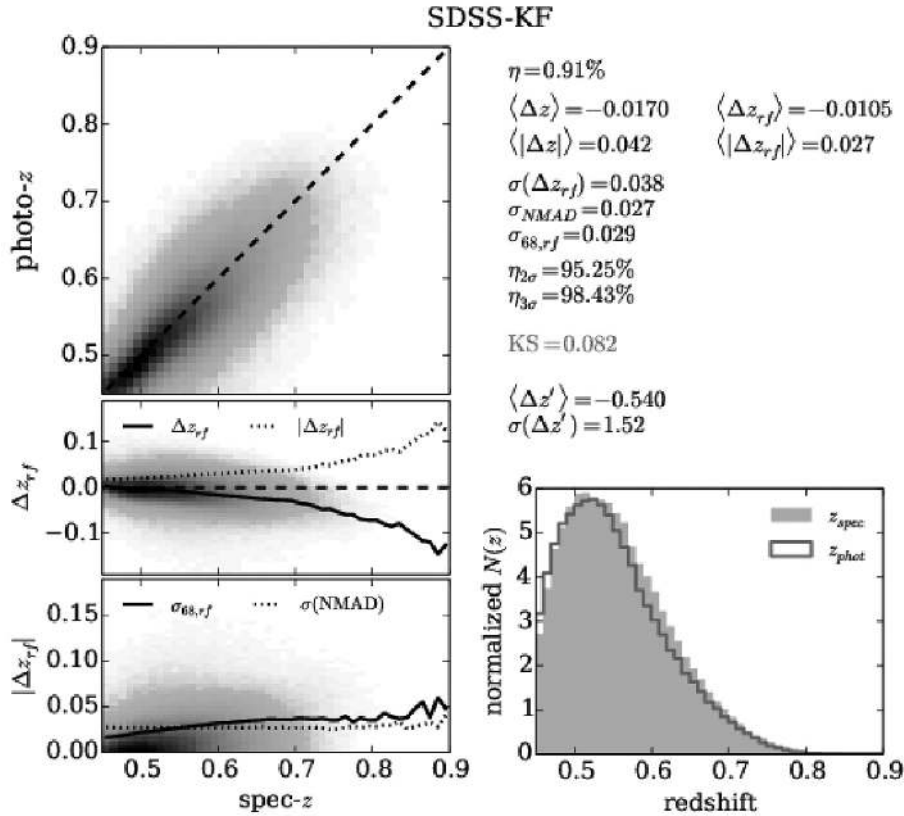
Figure 24. Photo-z results from the KF approach published by SDSS. See Fig. 21 for a detailed description.
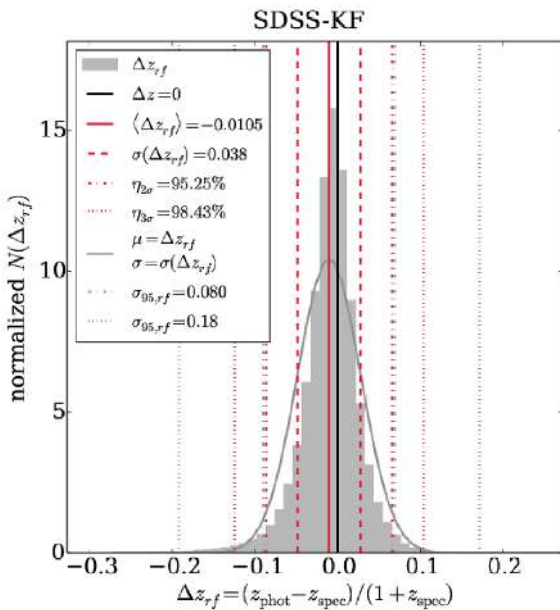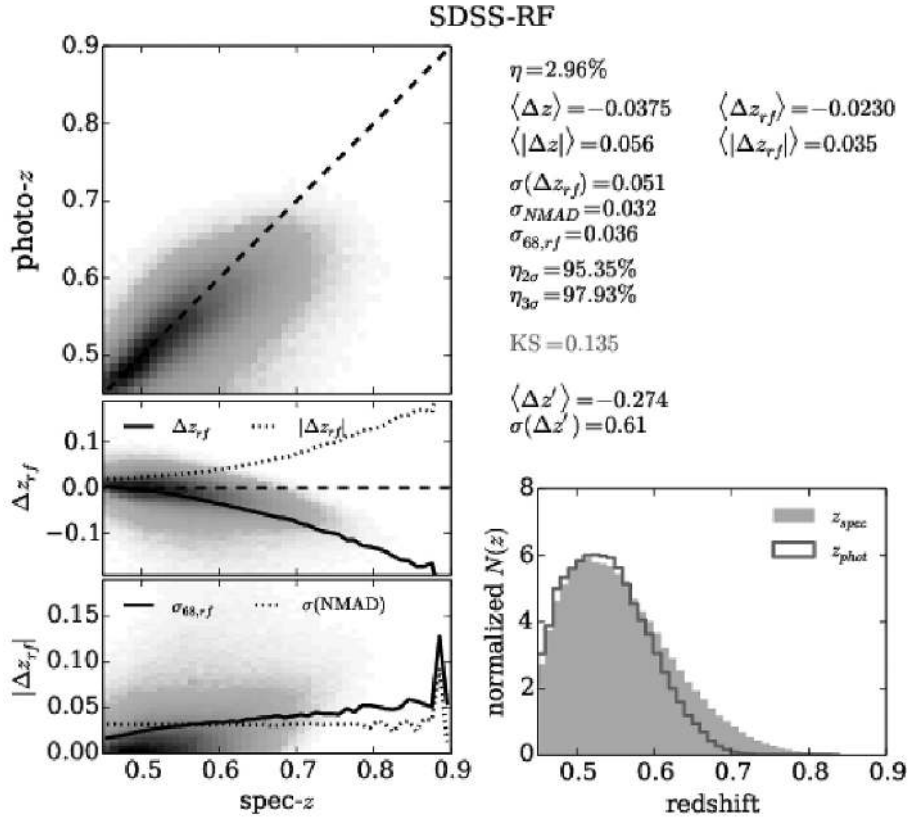
Figure 25. Photometric redshift restframe error distribution (grey filled histogram) estimated by the SDSS-KF method. A detailed explanation is given in Fig. 22.

The second photometric redshift results published by the SDSS are from another empirical code, which uses random forests to predict photo-zs (Carliles et al. 2010, hereafter SDSS-RF). Fig. 26 presents the results of the SDSS-RF code. In this case, the outlier

rate is significantly higher than in both previous cases, as well as the mean and the absolute restframe errors. Above $z \sim 0.55$ the photo-zs are systematically underestimated to a greater extent than in the former two cases, yielding the low bias. Furthermore, the scatter values $\sigma(\Delta z_{\mathrm{rf}})$, $\sigma_{68,\mathrm{rf}}$, and σ_{NMAD} have also increased in comparison to the results of this work, as well as compared to the SDSS-KF redshifts.

In the distribution of $N(z_{\mathrm{phot}})$, we can observe the lack of higher photoz values which accumulate around $z \sim 0.55$. Because of that, the KS test value is higher. In terms of their estimated photometric redshift errors, the SDSS-RF method outperforms the SDSS-KF approach, but still produces a larger bias in $\langle \Delta z' \rangle$. The errors are furthermore overestimated, visible in the $\sigma_{\Delta z'}$ value which is lower than the desired value of 1. Again, this affects also the outlier classification derived from δz. In the case of SDSS-RF photo-zs, the null hypothesis of an object being an outlier yields $\alpha = 0.86$ and $\beta = 0.015$ (shown together with the previous results in Table 3). The higher values of δz yield a smaller probability of an object being misclassified as a non-outlier, but also raises β (although only to a small value). So although the errors are overestimated on average, outliers are still misclassified at a significance level of 0.86.

In Fig. 27, we plot the Δz_{rf} distribution. The shift to the left is present to an even greater extent than in the previous case, and strongly deviating from a Gaussian with $\mu = \langle \Delta z_{\mathrm{rf}} \rangle$ and $\sigma = \sigma(\Delta z_{\mathrm{rf}})$ also plotted in Fig. 27. This is a result of the underestimation of the photo-zs. Furthermore, as in the previous cases, the value of $\sigma_{95,\mathrm{rf}}$ is marginally higher than $2\sigma(\Delta z_{\mathrm{rf}})$, while the deviation is greater in $\sigma_{99,\mathrm{rf}}$. On the other hand $\eta_{2\sigma} = 95.35$ per cent is even closer to 95.45 per cent than in both previous cases, while $\eta_{3\sigma} = 97.93$ per cent deviates more strongly, ~ 1.8 per cent.

Figure 26. Photo-z results from the RF approach published by the SDSS. See Fig. 21 for a detailed description.
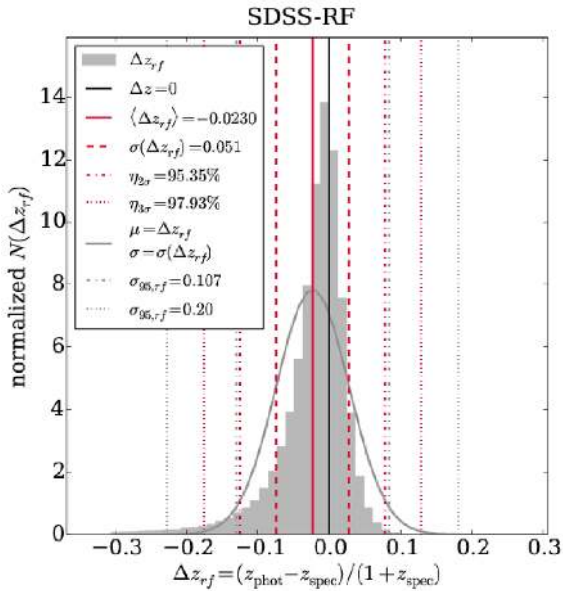
Figure 27. Photometric redshift restframe error distribution of the SDSS-RF published by the SDSS. A detailed explanation is given in Fig. 22.

We present a summary of the photometric redshift quality metrics of the results from the template fitting of this work and the two SDSS photo-zs in Table 2. The model-prior combination of this work produces the lowest outlier fraction, bias, and mean absolute error compared to the SDSS photoz results. When considered as a function of spec-z the photometric redshifts derived in this work

are slightly biased to lower z_{phot} values for higher z_{spec}, but not to the same extent of the SDSS codes. Furthermore, the different scatter values calculated (i.e. $\sigma(\Delta z_{rf})$, $\sigma_{68,rf}$, and σ_{NMAD}) are also lower for photo-zs of this work. The distribution of redshift errors is more similar to a Gaussian for the models, priors, and code of this work, which can be observed in Figs 22, 25, and 27. Considering the number fractions within $2\sigma(\Delta z_{rf})$ and $3\sigma(\Delta z_{rf})$ all three codes perform similarly well. However, this does not mean that the Δz_{rf} distributions are necessarily good approximations of Gaussian distributions with $\mu = \Delta z_{rf}$ and $\sigma = \sigma(\Delta z_{rf})$ which can be observed in Figs 22, 25, and 27.

When we evaluate the similarity between the redshift distributions, from which spectroscopic and photometric redshifts are sampled, through a KS test, the photo-zs of this work yield the best results for predictions of the mode which are improved when considering the whole $P(z)$ distribution.

The results of a significance test with the null hypothesis being that an object is a photometric redshift outlier is shown in Table 3. Not one of the codes produces errors that can reliably predict outliers. Albeit the SDSS-RF produces the best α, this is mostly due to the overestimated errors (cf. Table 2).

5.3 Deviations in colour predictions

In the last section, we analysed the photometric redshift results with our code in comparison to the SDSS photo-zs. The redshifts from the SDSS-RF method were outperformed by those derived with the SDSS-KF code. We want to analyse if the photo-zs of this work can produce the CMASS galaxy colours to a better extent than if we
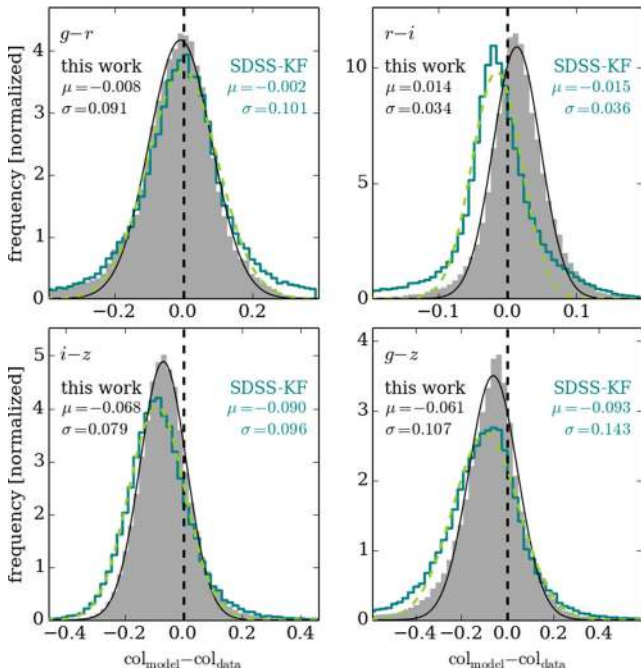
Figure 28. Deviation in $g - r$, $r - i$, $i - z$, and $g - z$ colours for photometric redshift results with the novel templates (grey filled histogram) and model predictions when fitting the same templates to the SDSS-KF photo-zs (blue histogram). Gaussians are fitted to both distributions, and are shown by a solid black line and a dashed green line, respectively. The best-fitting parameters of the Gaussians are printed in the panels.

fit the same models to the data at the photometric redshifts of the SDSS-KF method.

Fig. 28 presents the differences between the model predictions and the data in the $g - r$, $r - i$, $i - z$, and $g - z$ colours. The prediction of the photo-zs of this work is plotted in grey, the best fitting Gaussians are presented by black curves, and their parameters are given in the plot. When fitted to the photometric redshifts of the SDSS-KF approach, the novel BC03 $\lambda^{-\beta}$ models from Section 4.2 predict colours that are represented by blue histograms in Fig. 28, to which we also fit Gaussian curves. In all four colours, the standard deviation of the Δcol distribution is larger for SDSS-KF. Furthermore, the mean deviation from the data colours is greater for SDSS-KF results in the $i - z$ and $g - z$ colours. μ is of the same order in $r - i$ for both cases. It is smaller for SDSS-KF photo-zs in $g - r$, but there both values are very low. The fact that the SDSS-KF $r - i$, $i - z$, and $g - z$ colours are shifted bluewards from the distribution in colours of this work is a result of the SDSS-KF photo-zs being more heavily underestimated. All considered Δcol distributions resemble their best-fitting Gaussians well in the inner parts, with increased flanks. However, this increase is more significant in for the SDSS-KF photo-zs in $r - i$.

Therefore, the photometric redshifts derived in this paper not only generally have better quality metrics than the SDSS photo-zs, but can furthermore recreate the colours of the CMASS galaxies to a better extent.

6 SUMMARY AND CONCLUSIONS

In this work, we created a set of model SEDs that are designed to match the colours of the BOSS CMASS sample and provide

accurate photometric redshifts. We first analysed the colours of the LRG model of Maraston et al. (2009) which was created to match the $g - r$ and $r - i$ colours of the LRG sample of SDSS-II (Eisenstein et al. 2001), and found that we cannot use a singular age or evolution configuration that matches the data in all colours. Therefore, we created models for four redshift bins of widths $\Delta z = 0.04$ centred on $z_{spec} = 0.5, 0.6, 0.7, 0.8$ with the stellar population synthesis code of BC03. We generated SSPs and CSPs with decreasing SFHs at various metallicities, and sampled the models from a wide age range. These models were then fitted in superposition to a burst component and dust extinction to the data at known spectroscopic redshifts.

We observe a mismatch in the colours of the data and the models which is due to a too shallow red continuum slope in the model SEDs. Large variations in the model parameters (i.e. metallicity, SFH, age, burst, dust extinction, IMF) do not produce models that better fit the data. We speculate that inaccuracies in the modelling of the AGB phase could be the reason for the observed colour mismatch. In order to better recreate the colours of the CMASS galaxies, we introduce additional degrees of freedom by modifying the model SEDs by multiplication with $\lambda^{-\beta}$ for $\lambda > \lambda_i$ with several values for λ_i and β. We showed that the BC03 models modified in this way indeed are a better match to the colours of the CMASS galaxies (Figs 3–6 and 9–16) and also yield better χ^2 values from the fitting (Fig. A1).

From these best-fitting SEDs, we selected a small subset that should cover the region in colour space and absolute magnitude in r of the CMASS sample. We therefore projected the CMASS galaxy colours and M_r of the four z bins on to two-dimensional planes using a SOM. Afterwards, we partition the plane in six clusters for each redshift bin with a k-means clustering algorithm and select one model SED per cluster cell that produces the best photo-zs for galaxies within the same cell. We estimated photometric redshifts with a template fitting code and with the selected models and analyse their individual performance. Thereby, we modified their redshift priors to improve on the photo-zs, but also decided to omit some of the templates that do not yield accurate photo-zs on the whole sample, regardless how the priors are modified. We then compared the photometric redshift results with the photo-zs of two empirical methods published by SDSS and calculated several metrics that assess the quality of the photo-zs, their estimated errors, and their distribution. We found that the photo-zs with the generated models of this work produce better values in all quality metrics. Furthermore, we observed that including the stacked PDFs yields better results in the reconstruction of $N(z_{spec})$ mirrored in the results of a KS test. Concerning the estimated errors δz, a significance test shows that none of the three considered δz results provides a reliable classification of outliers. However, the probabilities of a non-outlier being misclassified is very small for all three considered cases.

Finally, we compared the predicted colours of the novel model SEDs when fitted to the photo-zs of this work and to the better of the two SDSS redshifts, to the data. We found that the deviations from the data are smaller for photo-zs of this work.

The photo-z catalog, as well as the models of this work and G13, can be downloaded at http://www.cefca.es/people/~ngreisel/

REFERENCES

Abazajian K. N. et al., 2009, ApJS, 182, 543
Ahn C. P. et al., 2014, ApJS, 211, 17
Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, MNRAS, 310, 540
Baldry I. K., Glazebrook K., Brinkmann J., Ivezić Ž., Lupton R. H., Nichol R. C., Szalay A. S., 2004, ApJ, 600, 681
Bender R. et al., 2001, in Cristiani S., Renzini A., Williams R. E., eds, Deep Fields, Springer-Verlag, p. 96
Benítez N., 2000, ApJ, 536, 571
Benítez N. et al., 2009, ApJ, 691, 241
Bolzonella M., Miralles J.-M., Pelló R., 2000, A&A, 363, 476
Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503
Brimioulle F., Lerchster M., Seitz S., Bender R., Snigula J., 2008, preprint (arXiv e-prints)
Brimioulle F., Seitz S., Lerchster M., Bender R., Snigula J., 2013, MNRAS, 432, 1046
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000 (BC03)
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Carliles S., Budavári T., Heinis S., Priebe C., Szalay A. S., 2010, ApJ, 712, 511
Carrasco Kind M., Brunner R. J., 2013, MNRAS, 432, 1483
Carrasco Kind M., Brunner R. J., 2014, MNRAS, 438, 3409
Cassarà L. P., Piovan L., Weiss A., Salaris M., Chiosi C., 2013, MNRAS, 436, 2824
Cenarro A. J., Gorgas J., Vazdekis A., Cardiel N., Peletier R. F., 2003, MNRAS, 339, L12
Chabrier G., 2003, PASP, 115, 763
Collister A. A., Lahav O., 2004, PASP, 116, 345
Conroy C., van Dokkum P. G., 2012, ApJ, 760, 71
Csabai I. et al., 2003, AJ, 125, 580
Csabai I., Dobos L., Trencséni M., Herczegh G., Józsa P., Purger N., Budavári T., Szalay A. S., 2007, Astron. Nachr., 328, 852
Dawson K. S. et al., 2013, AJ, 145, 10
Drory N., Feulner G., Bender R., Botzler C. S., Hopp U., Maraston C., Mendes de Oliveira C., Snigula J., 2001, MNRAS, 325, 550
Drory N., Bender R., Hopp U., 2004, ApJ, 616, L103
Drory N., Salvato M., Gabasch A., Bender R., Hopp U., Feulner G., Pannella M., 2005, ApJ, 619, L131
Eisenstein D. J. et al., 2001, AJ, 122, 2267

Eisenstein D. J. et al., 2011, AJ, 142, 72
Feldmann R. et al., 2006, MNRAS, 372, 565
Feulner G., Gabasch A., Salvato M., Drory N., Hopp U., Bender R., 2005, ApJ, 633, L9
Gabasch A. et al., 2004, A&A, 421, 41
Gabasch A., Goranova Y., Hopp U., Noll S., Pannella M., 2008, MNRAS, 383, 1319
Gerdes D. W., 2009, BAAS, 41, 483.03
Greisel N., Seitz S., Drory N., Bender R., Saglia R. P., Snigula J., 2013, ApJ, 768, 117 (G13)
Gruen D. et al., 2013, MNRAS, 432, 1455
Gruen D. et al., 2014, MNRAS, 442, 1507
Hanke M., Halchenko Y. O., Sederberg P. B., Hanson S. J., Haxby J. V., Pollmann S., 2009, Neuroinformatics, 7, 37
Hildebrandt H. et al., 2012, MNRAS, 421, 2355
Huterer D., Takada M., Bernstein G., Jain B., 2006, MNRAS, 366, 101
Ilbert O. et al., 2006, A&A, 457, 841
Kohonen T., 1982, Biol. Cybernetics, 43, 59
Kohonen T., 2001, Self-Organizing Maps, Springer, Berlin
Kroupa P., 2001, MNRAS, 322, 231
Longhetti M., Saracco P., 2009, MNRAS, 394, 774
Ma Z., Hu W., Huterer D., 2006, ApJ, 636, 21
Mancone C. L., Gonzalez A. H., 2012, PASP, 124, 606
Mandelbaum R. et al., 2008, MNRAS, 386, 781
Maraston C., 1998, MNRAS, 300, 872
Maraston C., 2005, MNRAS, 362, 799 (M05)
Maraston C., Strömbäck G., 2011, MNRAS, 418, 2785 (M11)
Maraston C., Strömbäck G., Thomas D., Wake D. A., Nichol R. C., 2009, MNRAS, 394, L107 (M09)
Maraston C., Pforr J., Renzini A., Daddi E., Dickinson M., Cimatti A., Tonini C., 2010, MNRAS, 407, 830
Marigo P., Girardi L., Bressan A., Groenewegen M. A. T., Silva L., Granato G. L., 2008, A&A, 482, 883
Meidt S. E. et al., 2012, ApJ, 744, 17
Monna A. et al., 2014, MNRAS, 438, 1417
Oke J. B., Gunn J. E., 1983, ApJ, 266, 713
Padmanabhan N. et al., 2005, MNRAS, 359, 237
Pedregosa F. et al., 2011, J. Machine Learn. Res., 12, 2825
Pforr J., Maraston C., Tonini C., 2012, MNRAS, 422, 3285
Pickles A. J., 1998, PASP, 110, 863
Saglia R. P. et al., 2012, ApJ, 746, 128
Salaris M., Weiss A., Cassarà L. P., Piovan L., Chiosi C., 2014, A&A, 565, A9
Salpeter E. E., 1955, ApJ, 121, 161
Sánchez E. et al., 2011, MNRAS, 411, 277
Sánchez C. et al., 2014, MNRAS, 445, 1482
Shapley A. E., Steidel C. C., Erb D. K., Reddy N. A., Adelberger K. L., Pettini M., Barmby P., Huang J., 2005, ApJ, 626, 698
Smee S. A. et al., 2013, AJ, 146, 32
Steinhaus H., 1957, Bull. Acad. Pol. Sci., Cl. III, 4, 801
Voronoi G., 1908, J. Reine Angewandte Math., 133, 97178
York D. G. et al., 2000, AJ, 120, 1579
Zibetti S., Gallazzi A., Charlot S., Pierini D., Pasquali A., 2013, MNRAS, 428, 1479

APPENDIX A: SED FITTING RESULTS

In this paper, we create models that should match the BOSS CMASS galaxies in terms of their colours in redshift bins with widths 0.04 centred on $z = 0.5, 0.6, 0.7, 0.8$. We do so by fitting sets of model SEDs with different properties like metallicity, SFHs, and ages to the data (Section 4.1.3). In superposition to the SEDs, a burst model and additional dust extinctions for the burst and the main component are further degrees of freedom. In Fig. A1, we show the distributions of the χ^2 values returned by the SED fitting code, when we fit the BC03, M05, and M11 models in equal setups to the data. The setup
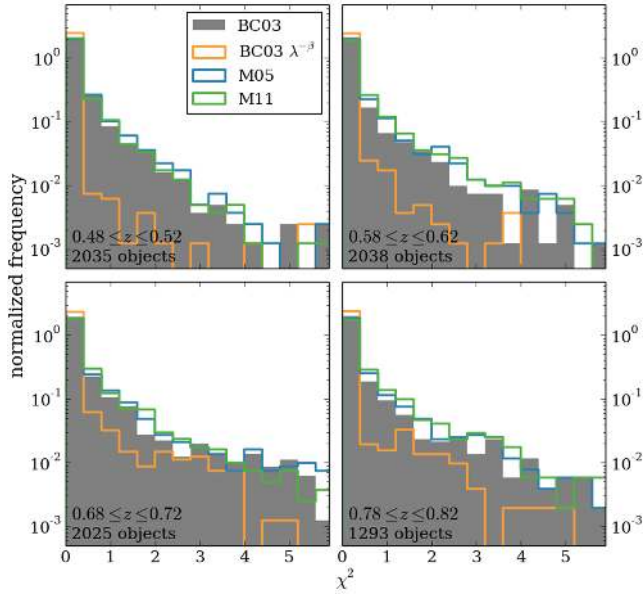
Figure A1. χ^2 distributions of SED fitting runs (at $z_{\rm spec}$) detailed with BC03 (grey), BC03 $\lambda^{-\beta}$ (orange), M05 (blue), and M11 (green) models for the four considered redshift regions. The frequency is normalized and plotted logarithmically. We can see that although the distributions for BC03, M05, and M11 are similar, the χ^2 values are improved by including the BC03 $\lambda^{-\beta}$ models.

of the SED fitting and the code is detailed in Section 4.1 and is equal for all cases.

In Section 4.1.3, we modify the BC03 models, changing their continuum slope by multiplying with $a \cdot \lambda^{-\beta}$ for $\lambda \geq \lambda_i$ with various β and λ_i values. This is done to introduce further degrees of freedom and to create models that match the colours of the data to a better extent. We fit the $\lambda^{-\beta}$ modifications of BC03 SSPs and CSPs with decreasing SFRs to the CMASS galaxy colours with the same setup as mentioned above and as detailed in Section 4.1.3. The resulting χ^2 values of the fit are shown together with the previous results of the BC03, M05, and M11 models in Fig. A1, where they are highlighted by green histograms.

While the results of the original BC03 models, M05, and M11 produce similar results in terms of χ^2, the fits are significantly improved by using the BC03 $\lambda^{-\beta}$ models. This is also shown in terms of colours in Figs 13–16 of Section 4.1.3.

APPENDIX B: QUERY FOR CMASS SAMPLE

We present here the SQL query executed to download the data from the SDSS-III CasJobs website[3] used in this work and detailed in Section 3. In the following, <fil> denotes any of the five Sloan filters and can be replaced by u, g, r, i, and z.

```
select p.ObjID, s.z as zSpec,
   s.CModelMag_<fil>, s.CModelMagErr_<fil>,
   s.Extinction_<fil>,
   p.CModelFlux_<fil>, p.CModelFluxIvar_<fil>,
   pz.z as zPhot, pz.zErr as zPhotErr,
   pzrf.z as zPhotRF, pzrf.zErr as zPhotRFErr
from specPhotoall s, spa a, Photoobjall p,
   Photoz pz, PhotozRF pzrf
where s.SpecObjID=a.SpecObjID and
   s.ObjID=p.ObjID and pz.ObjID=p.ObjID and
   pzrf.ObjID=p.ObjID and
   s.zWarning=0 and s.z>0 and
   (a.BOSS_TARGET1 & 2>0 or
   (a.BOSS_TARGET1 & 1>0 and s.Tile>=10324)) and
   a.BOSSPrimary=1 and a.zWarning_NoQSO=0 and
   a.Chunk!='BOSS1' and a.Chunk!='BOSS2' and
   p.Fiber2Mag_i<21.5 and p.Clean=1 and
   (p.CalibStatus_<fil> & 1)!=0 and
   s.z>=0.45 and s.z <=0.9
```

[3] http://skyserver.sdss3.org/CasJobs/.

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