



**HAL**  
open science

## Relative merits of different types of rest-frame optical observations to constrain galaxy physical parameters

Camilla Pacifici, Stéphane Charlot, Jérémy Blaizot, Jarle Brinchmann

► **To cite this version:**

Camilla Pacifici, Stéphane Charlot, Jérémy Blaizot, Jarle Brinchmann. Relative merits of different types of rest-frame optical observations to constrain galaxy physical parameters. *Monthly Notices of the Royal Astronomical Society*, 2012, 421, pp.2002-2024. 10.1111/j.1365-2966.2012.20431.x. hal-03645873

**HAL Id: hal-03645873**

**<https://hal.science/hal-03645873>**

Submitted on 5 Jun 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Relative merits of different types of rest-frame optical observations to constrain galaxy physical parameters

Camilla Pacifici,<sup>1,2\*</sup> Stéphane Charlot,<sup>1,2</sup> Jérémy Blaizot<sup>3,4,5</sup> and Jarle Brinchmann<sup>6</sup>

<sup>1</sup>UPMC Université Paris 6, UMR7095, Institut d'Astrophysique de Paris, F-75014 Paris, France

<sup>2</sup>CNRS, UMR7095, Institut d'Astrophysique de Paris, F-75014 Paris, France

<sup>3</sup>Université de Lyon, Lyon F-69003; Université de Lyon 1, Observatoire de Lyon, Saint-Genis Laval F-69230, France

<sup>4</sup>CNRS, UMR5574, Centre de Recherche Astrophysique de Lyon, Lyon F-69007, France

<sup>5</sup>Ecole Normale Supérieure de Lyon, Lyon F-69007, France

<sup>6</sup>Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, the Netherlands

Accepted 2011 December 21. Received 2011 December 15; in original form 2011 September 7

## ABSTRACT

We present a new approach to constrain galaxy physical parameters from the combined interpretation of stellar and nebular emission in wide ranges of observations. This approach relies on the Bayesian analysis of any type of galaxy spectral energy distribution using a comprehensive library of synthetic spectra assembled using state-of-the-art models of star formation and chemical enrichment histories, stellar population synthesis, nebular emission and attenuation by dust. We focus on the constraints set by five-band *ugriz* photometry and low- and medium-resolution spectroscopy at rest wavelengths  $\lambda = 3600\text{--}7400\text{ \AA}$  on a few physical parameters of galaxies: the observer-frame absolute *r*-band stellar mass-to-light ratio,  $M_*/L_r$ ; the fraction of a current galaxy stellar mass formed during the last 2.5 Gyr,  $f_{\text{SFH}}$ ; the specific star formation rate,  $\psi_S$ ; the gas-phase oxygen abundance,  $12 + \log(\text{O}/\text{H})$ ; the total effective *V*-band absorption optical depth of the dust,  $\hat{\tau}_V$ ; and the fraction of this arising from dust in the ambient interstellar medium,  $\mu$ . Since these parameters cannot be known a priori for any galaxy sample, we assess the accuracy to which they can be retrieved from observations by simulating ‘pseudo-observations’ using models with known parameters. Assuming that these models are good approximations of true galaxies, we find that the combined analysis of stellar and nebular emission in low-resolution [50 Å full width at half-maximum (FWHM)] galaxy spectra provides valuable constraints on all physical parameters. The typical uncertainties in high-quality spectra are about 0.13 dex for  $M_*/L_r$ , 0.23 for  $f_{\text{SFH}}$ , 0.24 dex for  $\psi_S$ , 0.28 for  $12 + \log(\text{O}/\text{H})$ , 0.64 for  $\hat{\tau}_V$  and 0.16 for  $\mu$ . The uncertainties in  $12 + \log(\text{O}/\text{H})$  and  $\hat{\tau}_V$  tighten by about 20 per cent for galaxies with detectable emission lines and by another 45 per cent when the spectral resolution is increased to 5 Å FWHM. At this spectral resolution, the analysis of the combined stellar and nebular emission in the high-quality spectra of 12 660 Sloan Digital Sky Survey (SDSS) star-forming galaxies using our approach yields likelihood distributions of  $M_*$ ,  $12 + \log(\text{O}/\text{H})$ ,  $\hat{\tau}_V$  and  $\psi_S$  similar to those obtained in previous separate analyses of the stellar and nebular emission at the original (twice higher) SDSS spectral resolution. Meanwhile, rest-frame *ugriz* photometry provides competitive constraints on  $M_*/L_r$ . We show that the constraints derived on galaxy physical parameters from these different types of observations depend sensitively on signal-to-noise ratio. Our approach can be extended to the analysis of any type of observation across the wavelength range covered by spectral evolution models.

**Key words:** galaxies: general – galaxies: ISM – galaxies: statistics – galaxies: stellar content.

## 1 INTRODUCTION

The light emitted by stars and the interstellar medium (ISM) in galaxies contains important clues about the physical processes that drive galaxy evolution. Different strategies have been adopted to survey the spectral energy distributions of large samples of

\*E-mail: pacifici@iap.fr

galaxies to gather these clues. High-quality photometry from surveys such as the Sloan Digital Sky Survey (SDSS; York et al. 2000), the Canada–France–Hawaii Telescope Legacy Survey (CFHTLS; Goranova et al. 2009) and the Cosmic Evolution Survey (COSMOS; Scoville et al. 2007) has brought important information about the rest-frame optical emission from millions of galaxies out to redshifts of a few. This is supplemented by medium-resolution spectroscopy of over a million local galaxies from the SDSS and the Two-degree Field Galaxy Redshift Survey (2dFGRS; Colless et al. 2001), and low- and medium-resolution spectroscopy of tens of thousands of galaxies at higher redshift from surveys such as the VIMOS VLT Deep Survey (VVDS; Le Fèvre et al. 2005) and the Deep Extragalactic Evolutionary Probe (DEEP2; Davis et al. 2003). The interpretation of the high-quality data gathered by these surveys has triggered the development of sophisticated models of galaxy spectral evolution (e.g. Bruzual & Charlot 2003; González Delgado et al. 2005; Maraston 2005; Coelho et al. 2007; Walcher et al. 2009; Vazdekis et al. 2010; Thomas, Maraston & Johansson 2011; see also Charlot & Longhetti 2001; Kewley & Dopita 2002). These models can help us constrain the star formation and chemical enrichment histories of galaxies through the analysis of different types of photometric and spectroscopic observations. So far, however, the relative merits of different observational approaches to constrain basic physical properties of galaxies have not been quantified in detail.

Recent studies have brought important insight into the efficiency of different types of observations to constrain physical quantities such as star formation history, metallicity and dust content from observations of the rest-frame optical emission of galaxies. Multi-band photometry has proved valuable to constrain galaxy redshifts and basic quantities such as stellar mass-to-light ratio (provided a stellar initial mass function; hereafter IMF) and, when combined with rest-frame ultraviolet and near-infrared constraints, star formation rate and attenuation by dust (e.g. Bell & de Jong 2001; Salim et al. 2005, 2007; Ilbert et al. 2006; Schaerer & de Barros 2010; but see Wuyts et al. 2009; Conroy, White & Gunn 2010). Medium-resolution spectroscopy allows separate analyses of nebular-emission and stellar-absorption features, and hence, more accurate constraints on the star formation history, stellar and interstellar metallicities, attenuation by dust and nuclear activity (e.g. Kauffmann et al. 2003a; Brinchmann et al. 2004; Heckman et al. 2004; Tremonti et al. 2004; Cid Fernandes et al. 2005; Gallazzi et al. 2005; Panter et al. 2007; Wild et al. 2007). At low spectral resolution, the difficulty of distinguishing between the signatures of nebular-emission and stellar-absorption features, especially around H-Balmer lines, renders analyses more complicated (e.g. Lamareille et al. 2006). This difficulty could be overcome with the ability to interpret simultaneously the stellar and nebular emission from galaxies. In this case, the relative merits of multi-band photometry and low- and medium-resolution spectroscopy to constrain galaxy physical parameters would need to be reexamined.

In this paper, we develop a new approach to interpret the combined stellar and nebular emission from galaxies in wide ranges of multi-wavelength observations. To achieve this, we appeal to a set of state-of-the-art models of galaxy star formation and chemical enrichment histories, stellar population synthesis, nebular emission and attenuation by dust. We use these models to generate a comprehensive library of several million galaxy spectral energy distributions. This library can be used to retrieve probability density functions of physical parameters from the Bayesian analysis of any observed spectral energy distribution. We focus here on the constraints set by five-band *ugriz* photometry and low- and medium-

resolution spectroscopy at rest wavelengths  $\lambda = 3600\text{--}7400 \text{ \AA}$  on a few basic physical parameters of galaxies: the observer-frame absolute *r*-band stellar mass-to-light ratio ( $M_*/L_r$ ); the fraction of a current galaxy stellar mass formed during the last 2.5 Gyr ( $f_{\text{SFH}}$ ); the specific star formation rate ( $\psi_S$ ); the gas-phase oxygen abundance [ $12 + \log(\text{O}/\text{H})$ ]; the total effective *V*-band absorption optical depth of the dust ( $\hat{\tau}_V$ ); and the fraction of this arising from dust in the ambient ISM ( $\mu$ ). These parameters cannot be known a priori for any galaxy sample. To assess the accuracy to which they can be retrieved from different types of observations, we therefore simulate ‘pseudo-observations’ by convolving the spectral energy distributions of models with known parameters with appropriate instrument responses and then applying artificial noise to mimic true observations. This approach allows us to make accurate predictions for the optimistic case in which the models we rely on are good approximations of true galaxies. A most notable outcome of our study is that the combined analysis of stellar and nebular emission in low-resolution [50 Å full width at half-maximum (FWHM)], high-quality (median signal-to-noise ratio per pixel  $\bar{S}/\bar{N} = 20$ ) galaxy spectra provides valuable constraints on all the physical parameters mentioned above. We also explore how the constraints from different types of photometric and spectroscopic observations depend on the signal-to-noise ratio (S/N) and the inclusion of nebular emission in the analysis. Our main results are summarized in Table 2 (Section 5).

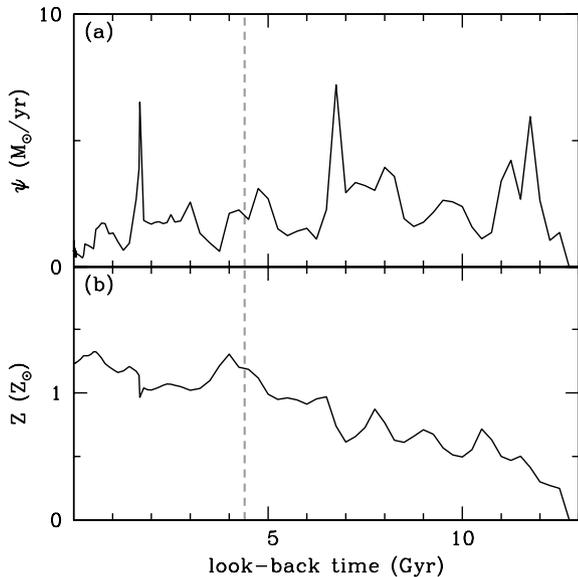
We describe our approach to constrain galaxy physical parameters from the combined analysis of stellar and nebular emission in Section 2. In Section 3, we investigate the accuracy to which the parameters  $M_*/L_r$ ,  $f_{\text{SFH}}$ ,  $\psi_S$ ,  $12 + \log(\text{O}/\text{H})$ ,  $\hat{\tau}_V$  and  $\mu$  can be constrained from different types of photometric and spectroscopic observations. Then, in Section 4, we compare the constraints derived from the analysis of the medium-resolution spectra of 12 660 SDSS star-forming galaxies using our approach with those obtained in previous separate analyses of the stellar and nebular emission. We also investigate the influence of the prior distributions of physical parameters on the retrieved probability density functions. Our conclusions are summarized in Section 5. Throughout this paper, we adopt the standard cosmology  $\Omega_M = 0.3$ ,  $\Omega_\Lambda = 0.7$  and  $h = 0.7$ .

## 2 MODELLING APPROACH

In this section, we describe our approach to assess the retrievability of galaxy physical parameters from the combined analysis of stellar and nebular emission in wide ranges of multi-wavelength observations. We first build a comprehensive library of star formation and chemical enrichment histories of galaxies (Section 2.1). We appeal to state-of-the-art models of spectral evolution to compute the emission from stars and gas and the attenuation by dust in these galaxies (Section 2.2). This allows us to assemble a large library of galaxy spectral energy distributions (Section 2.3), which we can use to retrieve likelihood distributions of physical parameters from the Bayesian analysis of any type of observed spectral energy distribution (Section 2.4).

### 2.1 Library of star formation and chemical enrichment histories

We build a comprehensive library of galaxy star formation and chemical enrichment histories by performing a semi-analytic post-treatment of the Millennium cosmological simulation of Springel



**Figure 1.** Example of galaxy star formation and chemical enrichment histories inferred from the semi-analytic post-treatment of the Millennium cosmological simulation: (a) star formation rate,  $\psi$ , and (b) interstellar metallicity,  $Z$ , plotted as a function of look-back time, for a galaxy with a present-day stellar mass of  $1.5 \times 10^{10} M_{\odot}$ . The vertical dashed line marks the evolutionary stage at which the galaxy is looked at in Fig. 3 (corresponding to a galaxy age of 8.4 Gyr).

et al. (2005). This large-scale cosmological simulation follows the growth, interaction and merging history of dark matter haloes from redshift  $z = 127$  to the present time.<sup>1</sup> We adopt the simple semi-analytic recipes of De Lucia & Blaizot (2007, see also Croton et al. 2006) to follow star formation and the associated metal production by gas falling into these dark matter haloes. We compute in this way the star formation and chemical enrichment histories of 500 000 galaxies with present-day stellar masses of between  $M_{*} \approx 5 \times 10^9 h^{-1} M_{\odot}$  (corresponding roughly to the mass resolution of the simulation; Blaizot et al. 2004) and  $5 \times 10^{11} h^{-1} M_{\odot}$ . Fig. 1 shows an example of star formation and chemical enrichment histories of a galaxy with present-day stellar mass  $1.5 \times 10^{10} M_{\odot}$ . The spikes in both curves reflect the various episodes of interactions and mergers that shaped the evolution of the galaxy. As shown by Wild et al. (2008), star formation histories computed in this way account well for the observed properties of galaxies at different redshifts in the VVDS ( $0.5 \lesssim z \lesssim 1$ ) and SDSS ( $0.05 \lesssim z \lesssim 0.1$ ) samples.

About 35 per cent of the 500 000 galaxies generated from the semi-analytic post-treatment of the Millennium simulation do not form stars at  $z = 0$ . These contain a negligible amount of interstellar gas and correspond to the population of present-day early-type galaxies. The remaining galaxies span a relatively narrow range of star formation properties. Their specific star formation rates peak around  $\psi_S \approx 0.1 \text{ Gyr}^{-1}$ , with a dispersion of only a factor of 2. Furthermore, the interstellar metallicity  $Z$  in these galaxies correlates significantly with  $\psi_S$ , in the sense that the most metal-poor galaxies form stars most actively. The galaxies also have roughly similar recent histories of star formation, as quantified by the star

formation history parameter  $f_{\text{SFH}}$ , which we define as the fraction of the current galaxy stellar mass formed during the last 2.5 Gyr (this time lag corresponds to the main-sequence lifetime of A–F stars with strong Balmer absorption features.). This parameter has a narrow distribution centred on  $f_{\text{SFH}} \approx 0.15$ , with a dispersion of only 0.1. The distributions of  $\psi_S$  and  $f_{\text{SFH}}$  are compatible with roughly constant (overall) star formation over the past 2.5 Gyr in galaxies more massive than  $5 \times 10^9 h^{-1} M_{\odot}$ , given that 40 per cent of the stellar mass formed during this period is returned to the ISM in the form of winds and supernovae (e.g. Bruzual & Charlot 2003). The above distributions of  $\psi_S$ ,  $Z$  and  $f_{\text{SFH}}$  are consistent with the observed properties of nearby, star-forming SDSS galaxies (e.g. Kauffmann et al. 2003a; Brinchmann et al. 2004; Tremonti et al. 2004).

The relatively limited ranges of star formation histories and interstellar metallicities described above, while appropriate to interpret observations of nearby SDSS galaxies more massive than  $5 \times 10^9 h^{-1} M_{\odot}$ , could bias statistical estimates of  $\psi_S$ ,  $Z$  and  $f_{\text{SFH}}$  in galaxies drawn from other samples or in different mass ranges (e.g. Weinmann et al. 2006; see also Section 4.2). To avoid such biases, we broaden the original prior distributions of these parameters predicted by the semi-analytic recipes of De Lucia & Blaizot (2007) in two ways. First, we widen the range in  $f_{\text{SFH}}$  by randomly drawing the evolutionary stage (i.e. galaxy age) at which a galaxy is looked at in the library. In practice, we draw this stage uniformly in redshift between  $z = 0$  and 1.7, i.e. the upper limit beyond which  $f_{\text{SFH}}$  reaches unity for a substantial fraction of the galaxies (Fig. 4b; this resampling makes the proportion of early-type galaxies drop from 35 to about 10 per cent in the library). Secondly, for galaxies with ongoing star formation at the selected stage, we broaden the distributions of the current star formation rate and the current interstellar metallicity as follows (we define here as ‘current’ the average of a quantity over a period of 10 Myr before a galaxy is looked at). We redistribute the specific star formation rate over the interval  $-2 < \log(\psi_S/\text{Gyr}^{-1}) < 1$  using the probability density function  $p[\log(\psi_S/\text{Gyr}^{-1})] = 0.05[1.5^5 - |\log(\psi_S/\text{Gyr}^{-1}) + 0.5|^5]$ , which is approximately uniform over the logarithmic range from  $\psi_S = 0.03$  to  $3 \text{ Gyr}^{-1}$  and drops to zero at  $\psi_S = 0.01$  and  $10 \text{ Gyr}^{-1}$  (Fig. 4c). We redistribute the current interstellar metallicity  $Z$  logarithmically between 0.1 and 3.5 times solar ( $Z_{\odot} = 0.017$ ).

The random resampling of  $\psi_S$  and  $Z$  for star-forming galaxies in the library is justified in part by the stochastic nature of star formation and chemical evolution in hierarchical scenarios of galaxy formation. Also, since the changes impact only the last 10 Myr of evolution, they have a negligible influence on the global star formation and chemical enrichment histories derived from the semi-analytic post-treatment of the Millennium simulation. In return, the use of wider and uncorrelated prior distributions of  $\psi_S$  and  $Z$ , together with the broader sampling in galaxy age (and  $f_{\text{SFH}}$ ), greatly enhances the usefulness of the model library to explore and interpret the spectral properties of observed galaxy populations.

## 2.2 Galaxy spectral modelling

We now describe the way in which we compute the emission from stars and gas and the attenuation by dust in the model galaxies generated in the previous section. The luminosity per unit wavelength  $L_{\lambda}(t)$  emerging at time  $t$  from a galaxy can be expressed as (see, e.g. Charlot & Longhetti 2001)

$$L_{\lambda}(t) = \int_0^t dt' \psi(t-t') S_{\lambda}[t', Z(t-t')] T_{\lambda}(t, t'), \quad (2.1)$$

<sup>1</sup> In the Millennium simulation, dark matter is represented as a set of about  $10^{10}$  point particles, which interact gravitationally in a cubic region of side  $500 h^{-1} \text{ Mpc}$ . Individual particles have a mass of  $8.6 \times 10^8 h^{-1} M_{\odot}$ , and the detection threshold for a dark matter halo is 20 particles, i.e. about  $1.7 \times 10^{10} h^{-1} M_{\odot}$  (see Springel et al. 2005, for detail).

where  $\psi(t - t')$  is the star formation rate at time  $t - t'$ ,  $S_\lambda[t', Z(t - t')]$  is the luminosity emitted per unit wavelength and per unit mass by a stellar generation of age  $t'$  and metallicity  $Z(t - t')$ , and  $T_\lambda(t, t')$  is the transmission function of the ISM, defined as the fraction of the radiation produced at wavelength  $\lambda$  at time  $t$  by a generation of stars of age  $t'$  that is transferred by the ISM. Following Charlot & Longhetti (2001), we write  $T_\lambda(t, t')$  as the product of the transmission functions of the ionized gas,  $T_\lambda^+(t, t')$ , and the neutral ISM,  $T_\lambda^0(t, t')$ ,

$$T_\lambda(t, t') = T_\lambda^+(t, t')T_\lambda^0(t, t'). \quad (2.2)$$

In the next paragraphs, we describe the prescriptions we adopt for the functions  $S_\lambda$ ,  $T_\lambda^+$  and  $T_\lambda^0$ .

### 2.2.1 Stellar emission

We compute the luminosity  $S_\lambda(t, Z)$  emitted per unit wavelength and per unit mass by a stellar generation of age  $t$  and metallicity  $Z$  using the latest version of Bruzual & Charlot (2003) stellar population synthesis code (Charlot & Bruzual, in preparation). This code predicts the spectral evolution of stellar populations at wavelengths from 91 Å to 160 μm and at ages between  $1 \times 10^5$  and  $2 \times 10^{10}$  yr, for different metallicities (from about 0.005  $Z_\odot$  to 4  $Z_\odot$ ), IMFs and star formation histories. We use the most recent version of the code, which incorporates a new library of observed stellar spectra (Sánchez-Blázquez et al. 2006) and new prescriptions for the evolution of stars less massive than 20  $M_\odot$  (Bertelli et al. 2008, 2009) and for the thermally pulsating asymptotic giant branch (TP-AGB) evolution of low- and intermediate-mass stars (Marigo et al. 2008). The main effect of the revised TP-AGB prescription is to improve the predicted near-infrared colours of intermediate-age stellar populations (e.g. González-Lópezlira et al. 2010). In all applications throughout this paper, we adopt the Galactic-disc IMF of Chabrier (2003).

### 2.2.2 Nebular emission

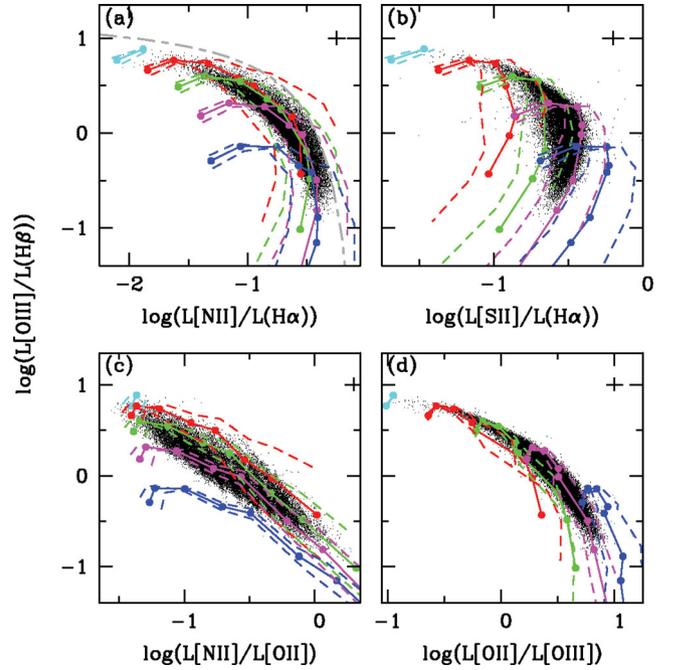
A main feature of our study is that we account for the contribution by ionized gas to the emission from galaxies. We follow the prescription of Charlot & Longhetti (2001) to compute the transmission function  $T_\lambda^+(t, t')$  of the gas photoionized at time  $t$  by stars of age  $t'$  in a galaxy. This consists in adopting effective (i.e. galaxy-wide) parameters to describe the ensemble of H II regions and the diffuse gas ionized by young stars throughout the galaxy. The main adjustable parameters are the interstellar metallicity,  $Z$ , the zero-age ionization parameter,  $U_0$  (which characterizes the volume-averaged ratio of ionizing-photon to gas densities at age  $t' = 0$ ), and the dust-to-metal (mass) ratio,  $\xi_d$  (which characterizes the depletion of metals on to dust grains) of the photoionized gas.<sup>2</sup>

Following Charlot & Longhetti (2001), we neglect the contribution by stars older than 10 Myr to nebular emission, i.e. we write

$$T_\lambda^+(t, t') = 1, \quad \text{for } t' > 10 \text{ Myr}. \quad (2.3)$$

We use the standard photoionization code CLOUDY (Ferland 1996) to compute the transmission function at earlier ages, assuming that galaxies are ionization bounded. For  $t' < 10^7$  yr, therefore,  $T_\lambda^+(t, t')$  is close to zero at wavelengths bluewards of the H-Lyman limit and greater than unity at wavelengths corresponding to emission lines

<sup>2</sup> These effective parameters were denoted by  $\tilde{Z}$ ,  $\langle \tilde{U}_0 \rangle$  and  $\tilde{\xi}_d$  in Charlot & Longhetti (2001).



**Figure 2.** Emission-line luminosities of galaxies computed using the models described in Section 2.2.2 (lines), compared with high-quality observations of a sample of 28 075 star-forming galaxies from the SDSS DR7 (black dots). The models assume constant star formation over the past 10 Myr. The data are corrected for attenuation by dust as described in Brinchmann et al. (2004). (a)  $L([\text{O III}]\lambda 5007)/L(\text{H}\beta)$  versus  $L([\text{N II}]\lambda 6584)/L(\text{H}\alpha)$ . (b)  $L([\text{O III}]\lambda 5007)/L(\text{H}\beta)$  versus  $L([\text{S II}]\lambda\lambda 6716, 6731)/L(\text{H}\alpha)$ . (c)  $L([\text{O III}]\lambda 5007)/L(\text{H}\beta)$  versus  $L([\text{N II}]\lambda 6584)/L([\text{O II}]\lambda 3727)$ . (d)  $L([\text{O III}]\lambda 5007)/L(\text{H}\beta)$  versus  $L([\text{O II}]\lambda 3727)/L([\text{O III}]\lambda 5007)$ . In each panel, lines of different colours refer to models with different zero-age ionization parameter  $\log U_0$  (cyan:  $-1.5$ ; red:  $-2.0$ ; green:  $-2.5$ ; magenta:  $-3.0$ ; blue:  $-3.5$ ). At fixed  $\log U_0$ , the lower dashed line corresponds to models with dust-to-metal ratio  $\xi_d = 0.1$ , the solid line to models with  $\xi_d = 0.3$  and the upper dashed line to models with  $\xi_d = 0.5$ . Along each line, dots mark the positions of models with different gas metallicity  $Z$  (0.10, 0.20, 0.50, 0.75, 1.00, 1.50, 2.00 and 3.00 times  $Z_\odot$ , from left to right; only low-metallicity models are considered for  $\log U_0 = -1.5$ ). In (a), the grey long-short-dashed line shows the Kauffmann et al. (2003b) criterion to separate star-forming galaxies from AGNs.

(we record the luminosities of the 109 most prominent emission lines at all wavelengths; we also include the recombination continuum radiation). When computing  $T_\lambda^+(t, t')$  in equations (2.1)–(2.2), we take the metallicity of the photoionized gas to be the current metallicity  $Z(t)$  of the galaxy. We randomly draw the dust-to-metal ratio from a uniform prior distribution between  $\xi_d = 0.1$  and 0.5, and the zero-age ionization parameter from a logarithmic prior distribution between  $\log U_0 = -3.5$  and  $-1.5$ . We fix all other parameters of the photoionization code at the standard values favoured by Charlot & Longhetti (2001, see their paper for detail).

In Fig. 2, we compare the luminosities of prominent optical emission lines obtained in this way for galaxies with different  $Z$ ,  $U_0$  and  $\xi_d$ , with high-quality observations of a sample of 28 075 star-forming galaxies from the SDSS Data Release 7 (DR7; Abazajian et al. 2009). The different panels show the relations between different luminosity ratios constructed with the  $[\text{O II}]\lambda 3726, 3729$  (hereafter  $[\text{O II}]\lambda 3727$ ),  $\text{H}\beta$ ,  $[\text{O III}]\lambda 5007$ ,  $\text{H}\alpha$ ,  $[\text{N II}]\lambda 6584$  and  $[\text{S II}]\lambda\lambda 6716, 6731$  lines. We selected the star-forming galaxies of Fig. 2 from the SDSS DR7 by requiring high S/N ( $> 10$ ) in all lines

and excluding galaxies in which the emission could be contaminated by an active galactic nucleus [AGN; we applied the conservative criterion of Kauffmann et al. (2003b) to reject AGNs from the standard Baldwin, Phillips & Terlevich (1981) line-diagnostic diagram reproduced in Fig. 2a]. The excellent agreement between models and SDSS observations in Fig. 2 is similar to that obtained by Brinchmann et al. (2004) using the original Charlot & Longhetti (2001) models. This allows the derivation of independent constraints on  $Z$ ,  $U_0$  and  $\xi_d$  from the different line luminosity ratios. For simplicity, in what follows, we adopt a fixed intrinsic line velocity dispersion of  $100 \text{ km s}^{-1}$  (typical of SDSS star-forming galaxies; e.g. Brinchmann et al. 2004) when computing the contribution by ionized gas to the emission from galaxies.

### 2.2.3 Attenuation by dust

The attenuation by dust of the emission from a galaxy depends on several factors, including the amount and structure of the interstellar gas, the metallicity and physical conditions of this gas and the orientation of the galaxy. Two galaxies with identical star formation and chemical enrichment histories but different sizes (and hence different total gas masses), ISM structures and orientations can have widely different dust attenuation optical depths, even though their dust-free spectral energy distributions are the same (modulo a mass scaling factor). We assume here for simplicity that the dust attenuation optical depth of a galaxy does not depend on other global galaxy parameters, such as the age  $t$ , the specific star formation rate  $\psi_S$  and the gas metallicity  $Z$ . We therefore write the transmission function of the neutral ISM in equations (2.1)–(2.2) as simply a function of stellar age  $t'$ ,

$$T_\lambda^0(t, t') = T_\lambda^0(t'). \quad (2.4)$$

In the ISM, stars are born in giant molecular clouds, which dissipate on a time-scale of the order of 10 Myr (e.g. Blitz & Shu 1980; Kawamura et al. 2009). Thus, the emission from stars younger than this time-scale and from the photoionized gas is more attenuated than that from older stars. We account for this effect using the simple dust model of Charlot & Fall (2000), who express the transmission function of the neutral gas as<sup>3</sup>

$$T_\lambda^0(t') = \exp[-\hat{\tau}_\lambda(t')]. \quad (2.5)$$

Here  $\hat{\tau}_\lambda(t')$  is the ‘effective absorption’ (or attenuation) optical depth of the dust seen by stars of age  $t'$ . This represents the average absorption of photons emitted in all directions by stars in all locations within the galaxy, including the effects of scattering on the path lengths of the photons before they are absorbed. Following Charlot & Fall (2000), we write

$$\hat{\tau}_\lambda(t') = \begin{cases} \hat{\tau}_\lambda^{\text{BC}} + \hat{\tau}_\lambda^{\text{ISM}} & \text{for } t' \leq 10 \text{ Myr,} \\ \hat{\tau}_\lambda^{\text{ISM}} & \text{for } t' > 10 \text{ Myr,} \end{cases} \quad (2.6)$$

where  $\hat{\tau}_\lambda^{\text{BC}}$  and  $\hat{\tau}_\lambda^{\text{ISM}}$  are the effective absorption optical depths of the stellar birth clouds and the ambient (i.e. diffuse) ISM, respectively. We note that, since the intrinsic spectral energy distributions of young and old stars evolve in time, equation (2.6) implies that the global attenuation curve of a galaxy containing several generations of stars is a complicated function of galaxy age, even though  $\hat{\tau}_\lambda$  is a simple function of  $t'$  (see e.g. fig. 5 of Charlot & Fall 2000).

<sup>3</sup> The function denoted here by  $T_\lambda^0(t')$  is equivalent to the function  $T_\lambda(t')$  defined by equation (6) of Charlot & Fall (2000), for  $f = 0.0$  in their notation.

We must also specify the dependence of  $\hat{\tau}_\lambda^{\text{BC}}$  and  $\hat{\tau}_\lambda^{\text{ISM}}$  on wavelength. For stellar birth clouds, we adopt a simple power law of index  $-1.3$  appropriate for optically thick clouds with dust properties in the middle range of the Milky Way, the Large and the Small Magellanic Clouds (da Cunha, Charlot & Elbaz 2008). In the ambient ISM, the shape of the effective absorption curve is more uncertain, as the spatial distribution of dust relative to the stars can vary widely from galaxy to galaxy. Also, while  $\hat{\tau}_\lambda^{\text{ISM}}$  represents an average optical depth along rays emanating in all directions from all stars, in practice, the attenuation will depend to some degree on the angle under which a galaxy is observed. Thus we write

$$\hat{\tau}_\lambda^{\text{BC}} = (1 - \mu)\hat{\tau}_V (\lambda/5500 \text{ \AA})^{-1.3}, \quad (2.7)$$

$$\hat{\tau}_\lambda^{\text{ISM}} = \mu\hat{\tau}_V (\lambda/5500 \text{ \AA})^{-n}. \quad (2.8)$$

In these expressions,  $\hat{\tau}_V = \hat{\tau}_V^{\text{BC}} + \hat{\tau}_V^{\text{ISM}}$  is the total effective  $V$ -band absorption optical depth of the dust seen by young stars inside birth clouds, and  $\mu = \hat{\tau}_V^{\text{ISM}}/(\hat{\tau}_V^{\text{BC}} + \hat{\tau}_V^{\text{ISM}})$  is the fraction of this contributed by dust in the ambient ISM. A value of  $\mu$  around 0.3 and a typical slope  $n = 0.7$  in equation (2.8) account remarkably well for several observed mean relations between the ultraviolet, optical and infrared properties of nearby galaxies (Charlot & Fall 2000; Kong et al. 2004; Salim et al. 2007; da Cunha et al. 2008; Wild et al. 2011a). More refined investigations using sophisticated dust models suggest that, in detail, the attenuation curve should flatten when the optical depth increases (Pierini et al. 2004; Tuffs et al. 2004; Rocha et al. 2008). This is supported by the typically greyer optical attenuation curves observed in nearby edge-on galaxies compared to face-on ones (Wild et al. 2011b). A recent analysis shows that the trend favoured by current models and observations can be reasonably well approximated by a simple expression of the slope of the attenuation curve of the form

$$n \approx \frac{2}{1 + \mu\hat{\tau}_V}, \quad (2.9)$$

with a scatter of about 20 per cent at fixed  $\mu\hat{\tau}_V$  (Chevallard et al., in preparation).

To compute  $T_\lambda^0(t, t')$  in equations (2.1)–(2.2), we therefore proceed as follows. We first draw  $\hat{\tau}_V$  between 0.01 and 4 from the probability density function  $p(\hat{\tau}_V) = 0.18 \arctan[5(4 - \hat{\tau}_V)]$ , which is approximately uniform over the range from  $\hat{\tau}_V = 0.01$  to 3 and drops to zero at  $\hat{\tau}_V = 4$  (see Fig. 4e). We further draw  $\mu$  from a uniform prior distribution between 0.1 and 0.7, and we draw  $n$  uniformly over an interval  $\Delta n = 0.3$  centred on the mean value predicted by equation (2.9). Then, we report the selected  $\hat{\tau}_V$ ,  $\mu$  and  $n$  in equations (2.4)–(2.8). This prescription accounts for a much broader range of plausible dust attenuation laws than the simple recipes usually adopted in galaxy spectral analyses (e.g. Kauffmann et al. 2003a; Brinchmann et al. 2004; Cid Fernandes et al. 2005; Panter et al. 2007; Wuyts et al. 2009). We therefore expect larger typical uncertainties arising from attenuation by dust than in these previous studies.

### 2.3 Library of galaxy spectral energy distributions

We use the library of star formation and chemical enrichment histories described in Section 2.1 and the spectral models described in Section 2.2 to simulate a large library of galaxy spectral energy distributions. This will be used in the remainder of this paper to assess the retrievability of galaxy physical parameters from various types of observations. To ensure that all kinds of spectral energy

**Table 1.** Prior distributions of the current (i.e. averaged over the last 10 Myr) physical properties of star-forming galaxies in the library of star formation and chemical enrichment histories assembled in Sections 2.1 and 2.2.

Parameter	Description	Range	Probability density function
$\psi_S$	Specific star formation rate	$-2 < \log(\psi_S/\text{Gyr}^{-1}) < 1$	$0.05[1.5^5 -  \log(\psi_S/\text{Gyr}^{-1}) + 0.5 ^5]$
$Z$	Metallicity of the star-forming gas	$-1 < \log(Z/Z_\odot) < 0.54$	Uniform
$\log U_0$	Zero-age ionization parameter	$-3.5 < \log U_0 < -1.5$	Uniform
$\xi_d$	Dust-to-metal ratio in the ionized gas	$0.1 < \xi_d < 0.5$	Uniform
$\hat{\tau}_V$	Total V-band attenuation optical depth of the dust	$0.01 < \hat{\tau}_V < 4$	$0.18 \arctan[5(4 - \hat{\tau}_V)]$
$\mu$	Fraction of $\hat{\tau}_V$ contributed by dust in diffuse ISM	$0.1 < \mu < 0.7$	Uniform
$n$	Slope of the attenuation curve in the diffuse ISM	$1.6 < n(1 + \mu\hat{\tau}_V) < 2.4$	Uniform

distributions are properly sampled, for each of the 500 000 star formation and chemical enrichment histories in the original library, we draw 10 different realizations of the evolutionary stage (and hence of the parameter  $f_{\text{SFH}}$ ) and, for star-forming galaxies, of the current physical properties (parameters  $\psi_S$ ,  $Z$ ,  $\log U_0$ ,  $\xi_d$ ,  $\hat{\tau}_V$  and  $\mu$ ), as summarized in Table 1. The final library therefore contains 5 million models, which we can use to generate catalogues of galaxy spectra.

In the present study, we focus primarily on the rest-frame optical properties of galaxies. Our main goal is to compare the constraints that can be derived on galaxy physical parameters from different types of observations in this wavelength range. At wavelengths between 3525 and 7500 Å, the stellar population spectra computed using the models presented in Section 2.2.1 have a native spectral resolution of 2.3 Å (FWHM), corresponding to a resolving power  $R \approx 2200$  at 5000 Å (the resolution is coarser outside this wavelength range). This allows us to investigate the retrievability of galaxy physical parameters from four different types of observations at optical wavelengths.

(i) *Medium-resolution spectroscopy* ( $\lambda = 3600\text{--}7400$  Å). We adopt a fiducial resolution of 5 Å FWHM, corresponding to  $R = 1000$  at  $\lambda = 5000$  Å, and a fixed pixel size of 2.5 Å.

(ii) *Low-resolution spectroscopy* ( $\lambda = 3600\text{--}7400$  Å). We adopt a fiducial resolution of 50 Å FWHM, corresponding to  $R = 100$  at  $\lambda = 5000$  Å, and a fixed pixel size of 25 Å.

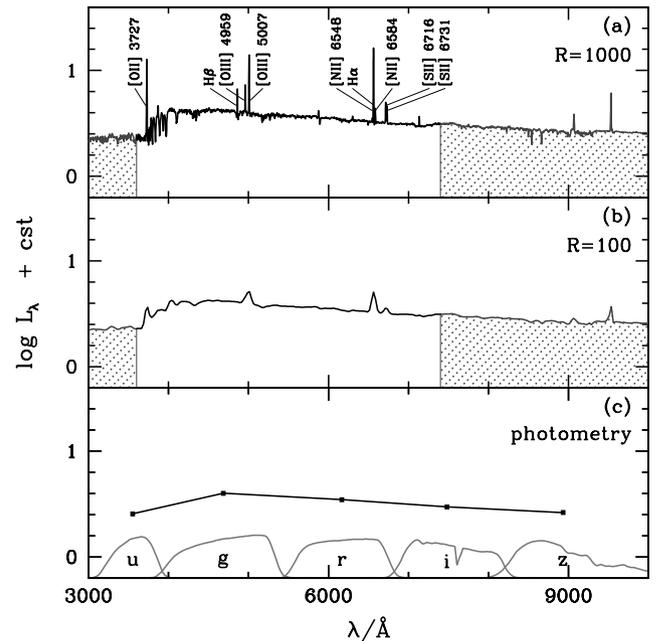
(iii) *Equivalent-width measurements of strong emission lines.* We also investigate the constraints set purely by the equivalent widths of the following emission lines, at both low and medium spectral resolution: [O II]λ3727; Hβ; [O III]λ4959, [O III]λ5007; [N II]λ6584; Hα; [N II]λ6584; [S II]λ6716 and [S II]λ6731. An advantage of equivalent widths is that they can provide useful information even in spectra which are not flux-calibrated.

(iv) *Broad-band photometry.* We adopt the SDSS *ugriz* filter response functions.

As an example, Fig. 3 shows the spectral energy distribution of the same model galaxy as in Fig. 1 at the age of 8.4 Gyr (with current physical parameters resampled using the distributions in Table 1), computed alternatively at medium (panel a) and low (panel b) spectral resolution and convolved with the SDSS *ugriz* filter response functions (panel c).

## 2.4 Retrieval of galaxy physical parameters

The galaxy spectral library generated in the previous section can be used to assess the retrievability of physical parameters from different types of observations. To achieve this, we first draw a random subsample of 10 000 spectral energy distributions from the library,



**Figure 3.** Spectral energy distribution of the same model galaxy as in Fig. 1 at the age of 8.4 Gyr, computed using the models described in Sections 2.1 and 2.2. (a) At a spectral resolution of 5 Å FWHM ( $R = 1000$  at 5000 Å). The black portion of the spectrum ( $\lambda = 3600\text{--}7400$  Å) is that used to retrieve galaxy physical parameters at this resolution. (b) Same as in (a), but at a spectral resolution of 50 Å FWHM ( $R = 100$  at 5000 Å). (c) Same as in (a), but convolved with the SDSS *ugriz* filter response functions (shown at the bottom). The model galaxy has  $f_{\text{SFH}} = 0.3$  and current parameters (resampled using the distributions in Table 1)  $\psi_S = 0.08 \text{ Gyr}^{-1}$ ,  $Z = 0.4 Z_\odot$ ,  $\log U_0 = -2.8$ ,  $\xi_d = 0.3$ ,  $\hat{\tau}_V = 1.0$ ,  $\mu = 0.6$  and  $n = 0.7$ . In (a) and (b), the adopted pixel size is half the size of a resolution element.

to which we apply artificial noise to mimic true observations. For low- and medium-resolution spectroscopy, we apply noise independent of wavelength (i.e. white noise), parametrized in terms of the median S/N per pixel,  $\overline{S/N}$ . We consider the cases  $\overline{S/N} = 5$  and 20 (for reference, the stellar population spectra computed using the models presented in Section 2.2.1 have a native  $\overline{S/N}$  of about 95). For multi-band photometry, instead, we take the S/N to be the same in all wavebands ( $S/N = 10$  or 30). In the following, we refer to this subsample of 10 000 noisy spectral energy distributions as ‘pseudo-observations’, and to the 4990 000 remaining spectral energy distributions in the library as simply ‘models’.

We use a Bayesian approach to quantify the accuracy to which several physical parameters of these pseudo-galaxies can be retrieved from different types of observables. The likelihood of the

$j$ th model, noted as  $\mathcal{M}_j$ , given the spectral energy distribution of a pseudo-galaxy can be written as

$$\ln P(\mathcal{M}_j|\{\mathcal{O}_i\}) = -\frac{1}{2} \sum_i \left( \frac{\mathcal{O}_i - w_j \mathcal{M}_{j,i}}{\sigma_i} \right)^2, \quad (2.10)$$

where  $\mathcal{O}_i$  and  $\mathcal{M}_{j,i}$  are the spectral energy distributions (i.e. either the spectral flux density at the  $i$ th wavelength, or the equivalent width of the  $i$ th emission line, or the flux in the  $i$ th photometric band; see Section 2.3) of the pseudo-galaxy and the  $j$ th model, respectively,  $\sigma_i$  is the uncertainty in  $\mathcal{O}_i$  corresponding to the adopted S/N, and

$$w_j = \left( \sum_i \frac{\mathcal{O}_i \mathcal{M}_{j,i}}{\sigma_i^2} \right) \left[ \sum_i \left( \frac{\mathcal{M}_{j,i}}{\sigma_i} \right)^2 \right]^{-1} \quad (2.11)$$

is the model scaling factor that maximizes  $P(\mathcal{M}_j|\{\mathcal{O}_i\})$  (this parameter is unity in the case of scale-free equivalent widths). In practice, in the case of medium-resolution spectroscopy, the summation on the right-hand side of expression (2.10) extends over 1520 wavelength points and hence generally yields values greater than 2000, even for a good-fit model.<sup>4</sup> The absolute likelihood in this case cannot be evaluated numerically, as it is always rounded to 0. For medium-resolution spectroscopy, therefore, we rather compute the probability of the  $j$ th model relative to the best-fitting model, noted  $\mathcal{M}_0$ ,

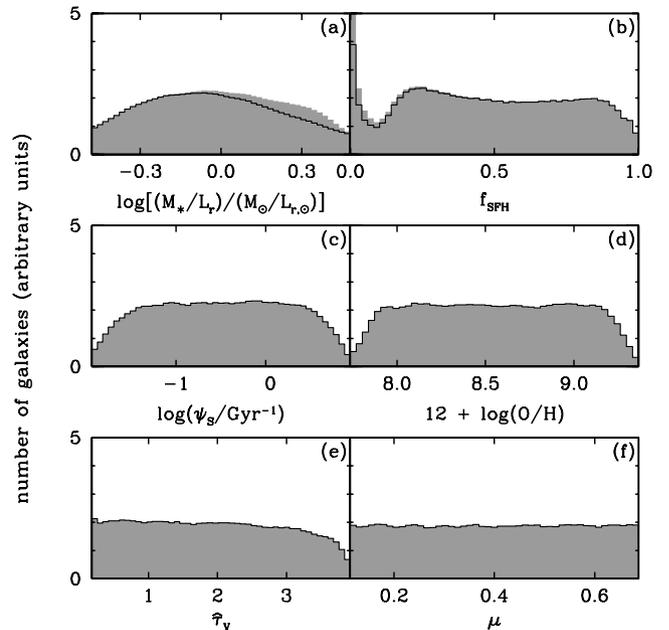
$$\ln \tilde{P}(\mathcal{M}_j|\{\mathcal{O}_i\}) = \ln P(\mathcal{M}_j|\{\mathcal{O}_i\}) - \ln P(\mathcal{M}_0|\{\mathcal{O}_i\}). \quad (2.12)$$

We have checked that in the case of low-resolution spectroscopy, where both  $P(\mathcal{M}_j|\{\mathcal{O}_i\})$  and  $\tilde{P}(\mathcal{M}_j|\{\mathcal{O}_i\})$  can be evaluated numerically, the likelihood distributions of galaxy physical parameters obtained using equations (2.10) and (2.12) are similar.

In this paper, we focus on the retrievability of six physical parameters from the observed spectral energy distributions of galaxies:

- (i) the observer-frame absolute  $r$ -band (stellar) mass-to-light ratio,  $M_*/L_r$ . This is the ratio of the galaxy stellar mass to the absolute luminosity in the observed  $r$  band;
- (ii) the fraction of the current galaxy stellar mass formed during the last 2.5 Gyr,  $f_{\text{SFH}}$ ;
- (iii) the specific star formation rate,  $\psi_S$ ;
- (iv) the gas-phase oxygen abundance in units of  $12 + \log(\text{O}/\text{H})$ , where O/H is the abundance by number of oxygen relative to hydrogen. In the models of nebular emission adopted in Section 2.2.2, the gas-phase oxygen abundance corresponding to solar metallicity ( $Z = Z_\odot$ ) is 8.81, 8.73 and 8.63 for dust-to-metal ratios  $\xi_d = 0.1, 0.3$  and  $0.5$ , respectively.
- (v) the total effective  $V$ -band absorption optical depth of the dust,  $\hat{\tau}_V$  (Section 2.2.3);
- (vi) the fraction of  $\hat{\tau}_V$  arising from dust in the ambient ISM,  $\mu$ .

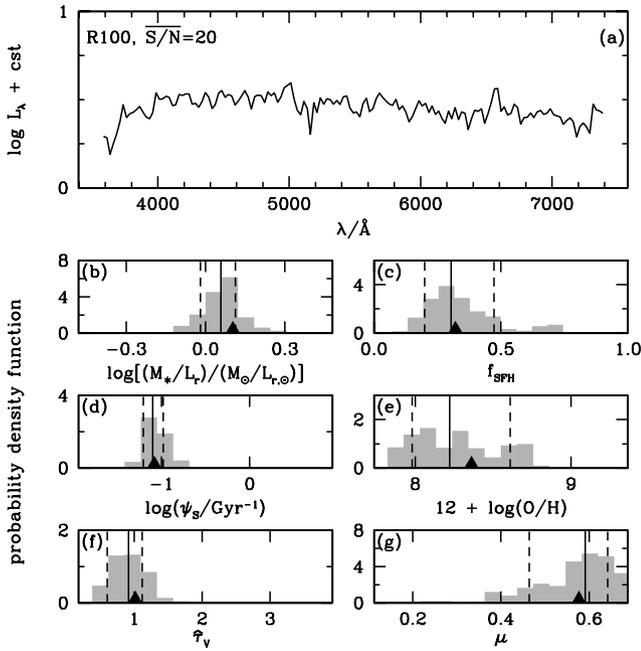
<sup>4</sup> We note that only a small contribution to these high values arises from our neglect of intrinsic correlations between spectral pixels in equation (2.10). A detailed investigation reveals that such correlations are significant only for pixels associated with emission lines, which reflects the fact that all emission lines arise from the same ionizing radiation. We have used a test sample of 3000 pseudo-galaxies to check the effect of accounting for pixel-to-pixel correlations on the calculation of model likelihoods. This requires the introduction of the covariance matrix in equation (2.10) and is computationally very demanding. The result is a negligible effect on the computed model likelihoods. Hence equation (2.10) should be appropriate to compute model likelihoods.



**Figure 4.** Prior distributions of selected physical parameters of the 5 million galaxies in the spectral library generated in Section 2.3: (a) observer-frame absolute  $r$ -band stellar mass-to-light ratio,  $M_*/L_r$ ; (b) fraction of current stellar mass formed during the last 2.5 Gyr,  $f_{\text{SFH}}$ ; (c) specific star formation rate,  $\psi_S$ ; (d) gas-phase oxygen abundance,  $12 + \log(\text{O}/\text{H})$ ; (e) total effective  $V$ -band absorption optical depth of the dust,  $\hat{\tau}_V$ ; (f) fraction of  $\hat{\tau}_V$  arising from dust in the ambient ISM,  $\mu$ . In each panel, the shaded histogram shows the distribution for all galaxies, while the solid histogram shows the contribution by star-forming galaxies alone. Non-star-forming galaxies are off scale (at  $\log \psi_S = -\infty$ ) in panel (c) and do not contribute to the distributions of interstellar parameters in panels (d)–(f).

Fig. 4 shows the prior distributions of these parameters in the galaxy spectral library generated in Section 2.3. In each panel, the shaded histogram shows the distribution including all galaxies, while the solid histogram shows the contribution by star-forming galaxies alone. Galaxies without star formation have typically high  $M_*/L_r$  in Fig. 4(a) and  $f_{\text{SFH}} \approx 0$  in Fig. 4(b). These galaxies are off scale (at  $\log \psi_S = -\infty$ ) in Fig. 4(c). They do not contribute to the histograms in Figs 4(d)–(f), which pertain to interstellar parameters. As emphasized in Sections 2.1 and 2.2, the prior distributions of Fig. 4 are as flat as possible to avoid unwanted biases in the retrieval of physical parameters from observed spectral energy distributions of galaxies. This choice is motivated by the desire to make the spectral library appropriate for the analysis of galaxies in wide ranges of physical properties (we note that the finite extent of the prior distributions in Fig. 4 could still bias analyses of galaxies with extreme parameters). Alternative prior distributions can be designed to help reduce the uncertainties in the parameters derived for galaxies with known specific physical properties (e.g. actively star-forming, early-type, metal-poor). We return to the influence of prior distributions on parameter determinations in Section 4.2, where we investigate the results obtained using the library of galaxy star formation and chemical enrichment histories derived from the original semi-analytic recipes of De Lucia & Blaizot (2007).

In Fig. 5, we show examples of the probability density functions (likelihood distributions) of physical parameters retrieved, using equations (2.10)–(2.11) and the prior distributions of Fig. 4, from



**Figure 5.** Probability density functions of the same physical parameters as in Fig. 4 retrieved, using the Bayesian approach described in Section 2.4, from the pseudo-galaxy spectrum shown at the top. The spectrum was obtained by adding artificial noise with  $S/N = 20$  to the  $R = 100$  model spectrum of Fig. 3(b). In each panel, the black triangle indicates the true value of the parameter of the pseudo-galaxy, the solid line the best estimate (50th percentile of the retrieved likelihood distribution) and the dashed lines the associated 68 per cent confidence interval (16th and 84th percentiles).

the low-resolution, high-S/N spectrum of a pseudo-galaxy in our library. This pseudo-observation was obtained by adding artificial noise ( $S/N = 20$ ) to the model spectrum of Fig. 3(b). We take the ‘best estimate’ of each parameter in Fig. 5 to be the median of the retrieved probability density function and the associated ‘uncertainty’ to be half the 16th–84th percentile range (this would be equivalent to the  $\pm 1\sigma$  range in the case of a Gaussian distribution). To quantify potential biases in the retrieved likelihood distribution, we further define the ‘accuracy’ as the absolute difference between the retrieved best estimate and true parameter value. In this example, all the parameters are well retrieved. The success in recovering the fraction  $f_{\text{SFH}}$  of intermediate-age stars (Fig. 5c) and the interstellar parameters  $12 + \log(\text{O}/\text{H})$ ,  $\hat{\tau}_V$  and  $\mu$  (Figs 5e–g) is remarkable. It is the first illustration of the power of our approach to exploit strong absorption- and emission-line signatures in noisy, low-resolution spectra. We examine this property in detail in Section 3.2.

It is important to stress that the approach adopted here to assess the retrievability of galaxy physical parameters from different types of observations, while idealized, is as realistic as can be achieved using modern techniques. Pseudo-observations are required to test the retrievability of physical parameters from observables, because the physical parameters of observed galaxies cannot be known a priori. The fact that we draw pseudo-observations from the same spectral library as the models used to analyse them make the constraints derived on the retrievability of physical parameters optimistic: our results correspond to the best possible case in which the models are a realistic representation of true galaxies. In Section 4, we also illustrate the constraints that can be derived by applying our approach to interpret a sample of observed galaxy spectra.

### 3 CONSTRAINTS ON GALAXY PHYSICAL PARAMETERS FROM DIFFERENT TYPES OF OBSERVATIONS

In this section, we use the modelling approach described in the previous section to quantify the accuracy to which physical parameters of galaxies can be retrieved from different types of observations. We focus on the retrievability of the six parameters introduced in Section 2.4: the observer-frame absolute  $r$ -band stellar mass-to-light ratio,  $M_*/L_r$ ; the fraction of the current galaxy stellar mass formed during the last 2.5 Gyr,  $f_{\text{SFH}}$ ; the specific star formation rate,  $\psi_S$ ; the gas-phase oxygen abundance,  $12 + \log(\text{O}/\text{H})$ ; the total effective  $V$ -band absorption optical depth of the dust,  $\hat{\tau}_V$ ; and the fraction of  $\hat{\tau}_V$  arising from dust in the ambient ISM,  $\mu$ . In Section 3.1, we first examine the constraints set on these parameters by multi-band photometric observations. We consider the influence of several factors, such as the S/N, the number of photometric bands, the contamination of broad-band fluxes by nebular emission lines and the uncertainty introduced by the lack of redshift information. We find that photometric observations can provide useful constraints mainly on  $M_*/L_r$  and  $\psi_S$ . Then, in Section 3.2, we quantify the improvement that can be achieved in the constraints on all parameters by appealing to different types of spectroscopic observations.

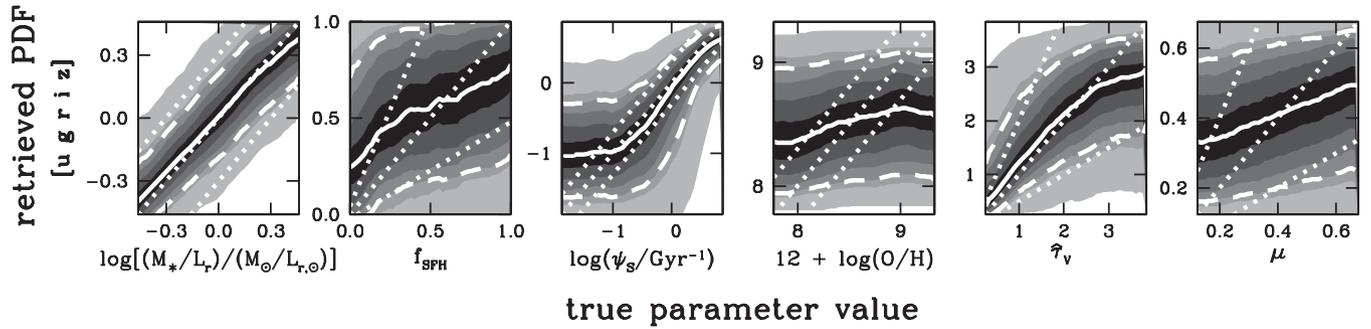
#### 3.1 Constraints from multi-band photometry

##### 3.1.1 Nearby galaxies

We first explore the extent to which physical parameters of nearby galaxies can be constrained using five-band  $ugriz$  photometry. We extract 10 000 spectral energy distributions from the library assembled in Section 2.3 to compute pseudo-observations of galaxies in these bands, as described in Section 2.4. We adopt for the moment an  $S/N = 30$  in all bands and place all spectral energy distributions at  $z = 0$ . The rest of the 5 million models in the library allow us to retrieve likelihood distributions of  $M_*/L_r$ ,  $f_{\text{SFH}}$ ,  $\psi_S$ ,  $12 + \log(\text{O}/\text{H})$ ,  $\hat{\tau}_V$  and  $\mu$  for each of these 10 000 pseudo-galaxies, by means of equations (2.10)–(2.11). In the following, we refer to this setting as the ‘standard’ case.

Fig. 6 summarizes the constraints that can be set on the physical parameters of nearby galaxies using high-quality, five-band  $ugriz$  photometry. For each parameter, we plot the average probability density function retrieved from the above analysis as a function of true parameter value. This average was obtained by co-adding and then renormalizing the probability density functions retrieved in narrow bins of true parameter value for the 10 000 pseudo-galaxies in our sample. The solid line in each panel of Fig. 6 locates the median (i.e. best estimate; Section 2.4) of the average likelihood distribution in each bin, the dashed lines the 16th and 84th percentiles (68 per cent confidence interval) and the outer edges of the lightest grey level the 2.5th and 97.5th percentiles (these would correspond to the  $\pm 2\sigma$  levels in the case of a Gaussian distribution). For reference, three dotted lines indicate the identity relation and deviations by a factor of 2 ( $\pm 0.3$  dex) between the retrieved and true parameters.

Fig. 6 shows that the  $r$ -band stellar mass-to-light ratio of nearby galaxies can be reasonably well constrained using five-band  $ugriz$  photometry. The retrieved best estimate is in excellent agreement with the true parameter value over most of the explored range, and the associated uncertainty, defined as half the 16th–84th percentile range (Section 2.4), is only about 0.19 dex (this number would be larger if we allowed for IMF variations.). At the extremities of the

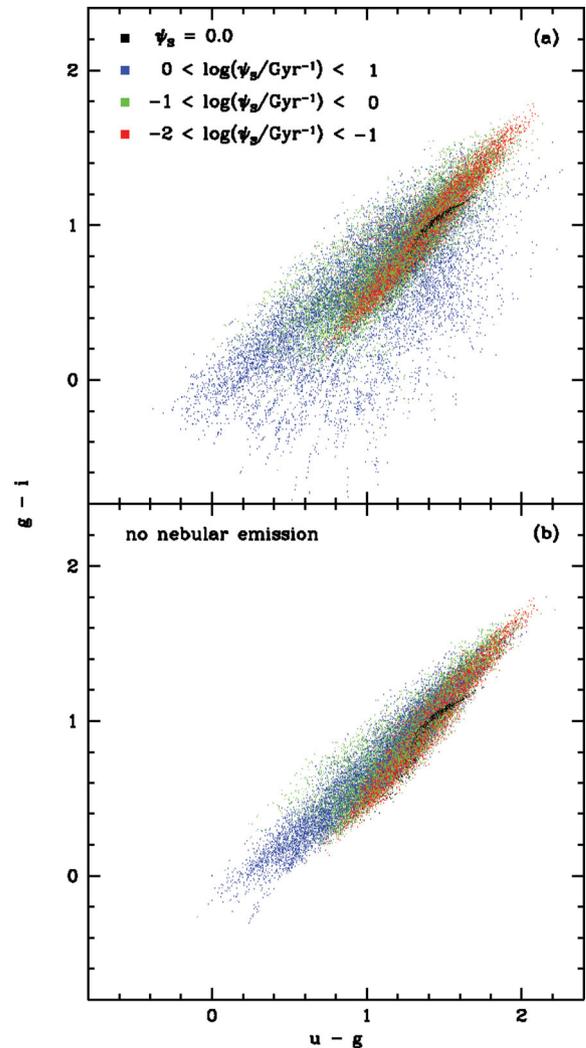


**Figure 6.** Average probability density functions of the same six physical parameters as in Fig. 4 retrieved, using the Bayesian approach described in Section 2.4, from five-band *ugriz* photometry with  $S/N = 30$  of a sample of 10 000 pseudo-galaxies (standard case). Each panel corresponds to a different parameter [from left to right:  $M_*/L_r$ ,  $f_{\text{SFH}}$ ,  $\psi_s$ ,  $12 + \log(O/H)$ ,  $\tau_v$  and  $\mu$ ]. In each case, average probability density functions in 50 narrow bins of true parameter value were obtained by co-adding and then renormalizing the probability density functions retrieved for the 10 000 pseudo-galaxies (see Section 3.1.1 for detail). Grey levels locate the 2.5th, 12th, 16th, 22nd, 30th, 40th, 60th, 70th, 78th, 84th, 88th and 97.5th percentiles of the average likelihood distribution in each bin. The solid line locates the associated median (best estimate) and the two dashed lines the 16th and 84th percentiles (68 per cent confidence interval). For reference, three dotted lines indicate the identity relation and deviations by a factor of 2 ( $\pm 0.3$  dex) between the retrieved and true parameter values.

range, the relative deficiency of galaxies with very small and very large  $M_*/L_r$  in the spectral library (Fig. 4a) slightly biases the retrieved likelihood distributions towards intermediate  $M_*/L_r$  values. Fig. 6 also shows that, in contrast to  $M_*/L_r$ , the fraction  $f_{\text{SFH}}$  of stellar mass formed in the past 2.5 Gyr cannot be well retrieved from five-band *ugriz* photometry. This is because the strong absorption features of intermediate-age stars have a negligible influence on broad-band fluxes. The signatures of  $f_{\text{SFH}}$  on broad-band colours are otherwise difficult to disentangle from those of stellar metallicity and attenuation by dust.

The constraints on specific star formation rate from five-band *ugriz* photometry in Fig. 6 are meaningful only for actively star-forming galaxies, i.e. for  $\psi_s > 1 \text{ Gyr}^{-1}$ , and the associated uncertainty is at best 0.30 dex. In such galaxies, the contamination of broad-band fluxes by strong emission lines produces identifiable signatures ( $[\text{O II}]$  in the *u* band;  $\text{H}\beta$  and  $[\text{O III}]$  in the *g* band;  $\text{H}\alpha$ ,  $[\text{N II}]$  and  $[\text{S II}]$  in the *r* band). Even then, the similar colours of galaxies with both low  $\psi_s$  and  $\tau_v$  and those with both high  $\psi_s$  and  $\tau_v$  give rise to long tails towards low  $\psi_s$  in the retrieved likelihood distributions. This degeneracy is illustrated by the overlap of galaxies with different  $\psi_s$  in Fig. 7(a), which shows the  $g - i$  versus  $u - g$  colours of a subset of 50 000 models in the spectral library. Together with the fact that, even in galaxies with strong emission lines, broad-band colours do not allow the distinction between different line ratios, this also implies poor constraints on the gas-phase oxygen abundance and the dust parameters  $\tau_v$  and  $\mu$  in Fig. 6. The retrieved likelihood distributions of these parameters are largely driven by the prior distributions of Fig. 4.

The results of Fig. 6 are consistent with those of Wuyts et al. (2009), who explored the recovery of physical parameters from broad-band spectral energy distributions of model galaxies taken from hydrodynamical merger simulations. Our study differs from theirs in that we consider much wider ranges of star formation and chemical enrichment histories and dust properties to retrieve likelihood distributions of physical parameters. Also, Wuyts et al. (2009) neglect the contribution by nebular emission to broad-band fluxes and include 11 bands to sample the rest-frame spectral energy distributions of galaxies over a wide range of wavelengths, between roughly 0.15 and  $3 \mu\text{m}$ . Finally, they analyse galaxy simulations in the redshift range  $1.5 \leq z \leq 2.9$ , which span a much narrower range of stellar population ages than the pseudo-galaxies in our sample.



**Figure 7.** (a)  $g - i$  colour plotted against  $u - g$  colour for a subset of 50 000 models from the galaxy spectral library assembled in Section 2.3. The models are colour-coded according to specific star formation rate, as indicated. (b) Same as (a), but without including the contribution by nebular emission to broad-band fluxes.

For all these reasons, Wuyts et al. (2009) find smaller typical uncertainties in the retrieved stellar mass ( $\sim 0.10$  dex) and star formation rate ( $\sim 0.31$  dex) than can be inferred from the constraints on  $M_*/L_r$  and  $\psi_S$  in Fig. 6. They also find much tighter constraints on the dust attenuation optical depth ( $\sim 0.31$ ). Wuyts et al. (2009) note that using simple (single-burst, constant and exponentially declining) star formation histories to fit the results of more realistic hydrodynamical merger simulations can induce significant deviations between the retrieved maximum-likelihood and true parameter values. Our adoption of the sophisticated galaxy spectral library assembled in Section 2.3 should minimize this type of limitation when applying our approach to the interpretation of real observations (Section 4).

We now return to our investigation and examine the sensitivity of the results of Fig. 6 to the adopted S/N, the photometric wavelength coverage and the inclusion of nebular emission when interpreting broad-band observations of galaxies. To quantify potential differences in the retrieved likelihood distributions relative to the standard case, we define the improvement factor as the ratio

$$I_\sigma = \frac{\text{uncertainty in the standard case}}{\text{uncertainty}}. \quad (3.13)$$

This quantity can be computed for any true parameter value. It exceeds unity when the constraints are tighter (i.e. the uncertainty smaller) than in the standard case. We further define the gain in accuracy as the difference

$$\Delta = (\text{accuracy in the standard case}) - (\text{accuracy}). \quad (3.14)$$

This quantity is positive when the retrieved best estimate is more accurate (i.e. the absolute difference between the retrieved best estimate and true parameter value smaller) than in the standard case. In Fig. 8, we show the analogue of Fig. 6 for three distinct alternatives to the standard case: adopting an S/N = 10 instead of 30 (top row); removing the constraints from the reddest two photometric bands,  $i$  and  $z$  (middle row); and not including nebular emission in the model library used to analyse pseudo-observations (bottom row). In all cases, we show at the bottom of each panel the improvement factor  $I_\sigma$  and the gain in accuracy  $\Delta$  as a function of true parameter value.

Fig. 8 (top row) shows that reducing the S/N from 30 to 10 makes the uncertainty in the retrieved  $M_*/L_r$  larger by about 15 per cent and, at the extremities of the explored range, the accuracy between 10 and 25 per cent worse. A more dramatic difference relative to the standard case arises from the loss of the emission-line signal for actively star-forming galaxies. This causes a sharp drop in  $I_\sigma$  at high  $\psi_S$ , implying poor constraints on the specific star formation rate over the entire explored range. Interestingly, Fig. 8 (middle row) shows that the effect of removing the  $i$ - and  $z$ -band constraints at S/N = 30 is very similar to that of lowering the S/N of five-band *ugriz* photometry. This is because the signatures of dust attenuation can be less well delineated from those of star formation from the *ugr* colours alone, even if the contamination of these broad-band fluxes by strong emission lines can in principle be detected at high S/N (Fig. 6). As expected, the other parameters,  $f_{\text{SFH}}$ ,  $12 + \log(\text{O}/\text{H})$ ,  $\hat{\tau}_V$  and  $\mu$ , can be even less well retrieved than in the standard case when the S/N is lowered and when the  $i$ - and  $z$ -band constraints are removed.

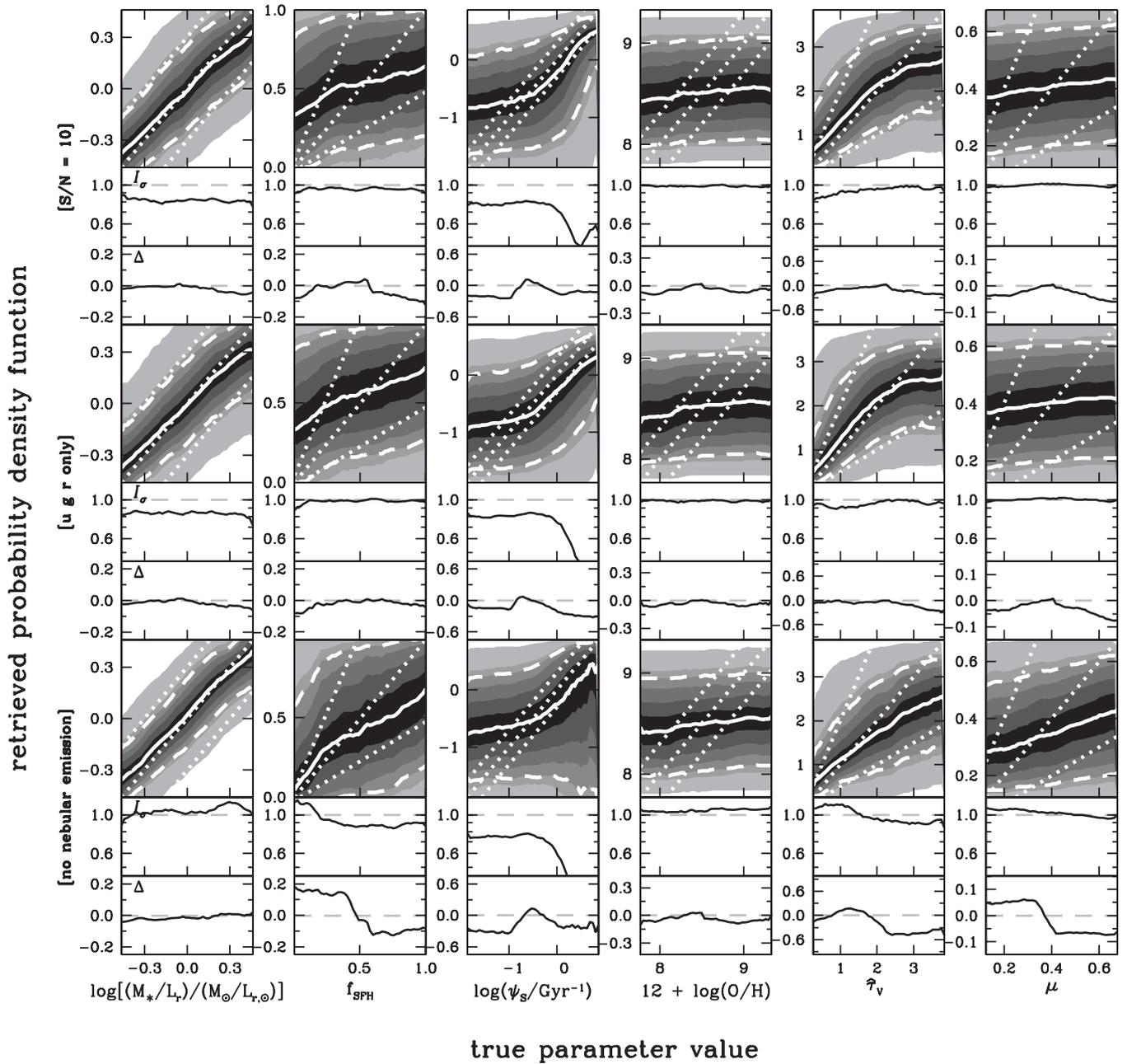
It is also of interest to quantify the importance of including nebular emission when interpreting broad-band observations of galaxies. A comparison between the top and bottom panels of Fig. 7 shows that the contribution by [O II] emission to the  $u$ -band flux and that of H $\beta$  and [O III] emission to the  $g$ -band flux can change the broad-band  $u - g$  and  $g - i$  colours of galaxies with  $\psi_S > 1 \text{ Gyr}^{-1}$  by

several tenths of a magnitude. In Fig. 8 (bottom row), we show the results of attempting to interpret the same pseudo-observations as in Fig. 6 (which include the contribution by nebular emission to the *ugriz* fluxes) using a spectral library which neglects nebular emission. This spectral library was constructed in the same way as that assembled in Section 2.3, but ignoring the prescription for nebular emission in Section 2.2.2. The retrieved  $M_*/L_r$  in this case is very similar to that obtained in the standard case, except for the slightly worse accuracy ( $\Delta < 0$ ) at very low  $M_*/L_r$ , consistent with the contamination of the  $r$ -band flux by H $\alpha$ , [N II] and [S II] emission in actively star-forming galaxies. As anticipated from Fig. 7, models without nebular emission do not sample broad-band colours well enough to allow the retrieval of meaningful constraints on  $\psi_S$  for actively star-forming galaxies. The other parameters,  $f_{\text{SFH}}$ ,  $12 + \log(\text{O}/\text{H})$ ,  $\hat{\tau}_V$  and  $\mu$ , are as poorly constrained as in the standard case. We conclude from Fig. 8 that high S/N, good wavelength coverage and accounting for nebular emission are all important to tighten the constraints on specific star formation rate from broad-band observations of galaxies, while the inclusion of nebular emission is less crucial to constrain stellar mass-to-light ratio.

### 3.1.2 Galaxies at unknown redshift

We have tested so far the retrievability of physical parameters from five-band *ugriz* photometry only for nearby galaxies. We now wish to explore the constraints achievable for galaxies at unknown redshift. Ilbert et al. (2006) have shown that redshift can be best estimated from five-band *ugriz* photometry out to  $z \sim 1.5$ . Photometric redshift measurements of more distant galaxies require observations at longer wavelengths (e.g. Ilbert et al. 2009), which are beyond the scope of the present paper (see Pacifici et al., in preparation). We therefore restrict ourselves here to the retrievability of physical parameters of galaxies from five-band *ugriz* photometry at redshifts in the range  $0 < z < 1$ .

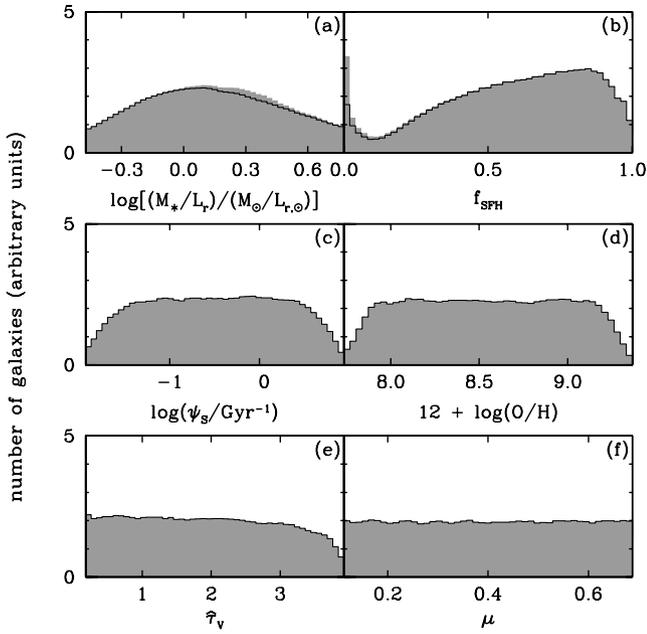
We use an approach similar to that described in Section 2.3 to generate a new library of 5 million spectral energy distributions of galaxies at random redshifts between 0 and 1. For each of the 500 000 star formation and chemical enrichment histories in the original library of Section 2.1, we first draw an observation redshift  $z_{\text{obs}}$  between 0 and 1. Then, we proceed as outlined in Section 2.3 and draw 10 different realizations of the evolutionary stage (i.e. galaxy age) at which the galaxy is looked at (uniformly in redshift between  $z_{\text{obs}}$  and  $z = 1.7$ , to better sample the parameter  $f_{\text{SFH}}$ ; see Section 2.1) and, for star-forming galaxies, of the current parameters  $\psi_S$ ,  $Z$ ,  $\log U_0$ ,  $\xi_d$ ,  $\hat{\tau}_V$  and  $\mu$  (using the distributions in Table 1). The resulting spectral library contains about 8 per cent early-type galaxies, i.e. slightly less than the 10 per cent in the  $z = 0$  library assembled in Section 2.3. Fig. 9 shows the prior distributions of the physical parameters  $M_*/L_r$ ,  $f_{\text{SFH}}$ ,  $\psi_S$ ,  $12 + \log(\text{O}/\text{H})$ ,  $\hat{\tau}_V$  and  $\mu$  in this library. As in Fig. 4, in each panel, the shaded histogram shows the distribution including all galaxies, while the solid histogram shows the contribution by star-forming galaxies alone. The prior distribution of the observer-frame absolute  $r$ -band stellar mass-to-light ratio in Fig. 9(a) reaches larger values than in Fig. 4(a) because of the contribution by (dusty) high-redshift galaxies with large  $K$ -corrections. The early evolutionary stage of high-redshift galaxies also implies a larger typical  $f_{\text{SFH}}$  in Fig. 9(b) than in Fig. 4(b). By construction, the distributions of the current physical parameters  $\psi_S$ ,  $12 + \log(\text{O}/\text{H})$ ,  $\hat{\tau}_V$  and  $\mu$  are the same in Figs 4 and 9 (from Table 1).



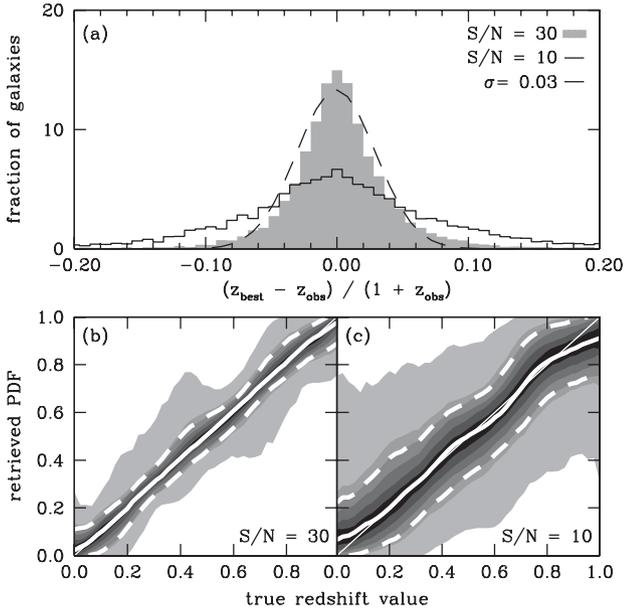
**Figure 8.** Same as Fig. 6, but for three distinct alternatives to the standard case: (top row) adopting an  $S/N = 10$  instead of 30; (middle row) using constraints from only the  $ugr$  photometric bands instead of  $ugriz$ ; and (bottom row) not including nebular emission in the model library used to analyse the sample of 10 000 pseudo-galaxies. In all cases, the improvement factor  $I_\sigma$  and the gain in accuracy  $\Delta$  (equations 3.13–3.14) are shown as a function of true parameter value at the bottom of each panel to quantify differences in the retrieved likelihood distributions relative to the standard case of Fig. 6.

By analogy with Section 3.1.1, we generate pseudo-observations in the  $ugriz$  photometric bands of 10 000 galaxies in the new library. These pseudo-observations sample different wavelength ranges of the rest-frame spectral energy distributions of galaxies at different redshifts, from roughly 3000–10 000 Å at  $z_{\text{obs}} = 0$  to 1500–5000 Å at  $z_{\text{obs}} = 1$ . We consider both high-quality ( $S/N = 30$ ) and low-quality ( $S/N = 10$ ) observations. Using the rest of the 5 million models in the library, we can retrieve likelihood distributions of  $z_{\text{obs}}$  for each of the 10 000 pseudo-galaxies, as well as likelihood distributions of  $M^*/L_r$ ,  $f_{\text{SPH}}$ ,  $\psi_s$ ,  $12 + \log(O/H)$ ,  $\hat{\tau}_V$  and  $\mu$ .

In Fig. 10(a), we plot the distribution of the difference between the retrieved best estimate (i.e. median) and true value of  $z_{\text{obs}}$ , in units of  $1 + z_{\text{obs}}$ , for the 10 000 pseudo-galaxies in our sample. The results for  $S/N = 30$  (shaded histogram) are in remarkable agreement with the  $\sigma = 0.03$  Gaussian uncertainty in this quantity quoted by Ilbert et al. (2006), who analysed CFHTLS ( $S/N \gtrsim 30$ )  $u^*g^*r^*i^*z^*$  photometry of about 3000 galaxies with spectroscopic redshifts between 0.2 and 1.5 (dashed line). For  $S/N = 10$ , the retrieved best estimate of  $z_{\text{obs}}$  in Fig. 10(a) remains in good general agreement with the true value (solid histogram). To better illustrate



**Figure 9.** Prior distributions of the same physical parameters as in Fig. 4 for the 5 million galaxies at redshifts between 0 and 1 in the spectral library assembled in Section 3.1.2. In each panel, the shaded histogram shows the distribution for all galaxies, while the solid histogram shows the contribution by star-forming galaxies alone. Non-star-forming galaxies are off scale (at  $\log \psi_s = -\infty$ ) in panel (c) and do not contribute to the distributions of interstellar parameters in panels (d)–(f).



**Figure 10.** (a) Distribution of the difference between the retrieved best estimate (i.e. median) and true value of  $z_{\text{obs}}$ , in units of  $1 + z_{\text{obs}}$ , for a sample of 10 000 pseudo-galaxies at redshifts between 0 and 1 observed in the *ugriz* photometric bands. The shaded and solid histograms show the distributions obtained for  $S/N = 30$  and  $S/N = 10$ , respectively. For reference, a dashed line indicates the  $\sigma = 0.03$  Gaussian accuracy quoted by Ilbert et al. (2006), who analysed CFHTLS ( $S/N \gtrsim 30$ )  $u^*g'r'iz'$  photometry of about 3000 galaxies with spectroscopic redshifts between 0.2 and 1.5. (b) Detail of the average retrieved probability density function plotted against  $z_{\text{obs}}$ , for  $S/N = 30$ . The lines and shading have the same meaning as in Fig. 6. (c) Same as (b), but for  $S/N = 10$ .

the origin of the histograms in Fig. 10(a), we show in Figs 10(b) and (c) the detail of the average retrieved probability density functions as a function of  $z_{\text{obs}}$  for  $S/N = 30$  and 10, respectively.

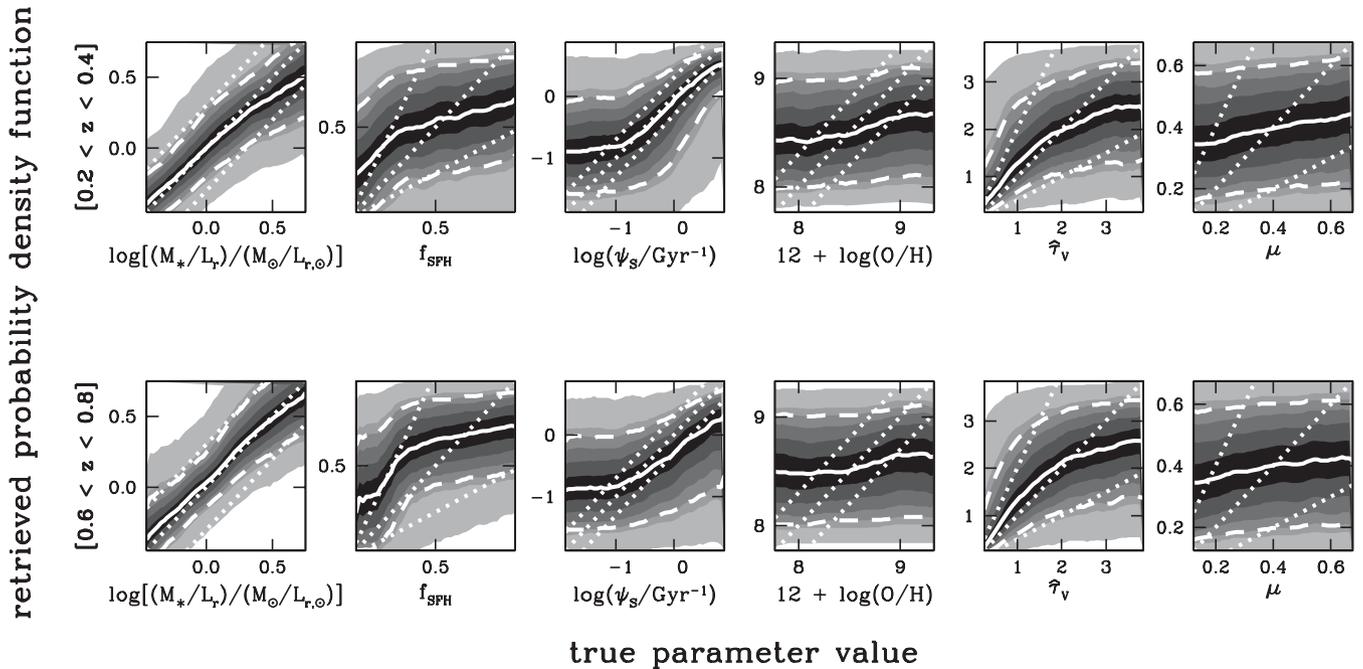
The retrieved probability density functions of the other physical parameters of interest to us are difficult to interpret when averaged over galaxies at all redshifts between 0 and 1, because of the different effects impacting the recovery of parameters in different redshifts ranges. For this reason, it is more instructive to explore the retrievability of  $M_*/L_r$ ,  $f_{\text{SFH}}$ ,  $\psi_s$ ,  $12 + \log(\text{O}/\text{H})$ ,  $\hat{\tau}_v$  and  $\mu$  for galaxies in different redshift ranges, for example  $0.2 < z_{\text{obs}} < 0.4$  and  $0.6 < z_{\text{obs}} < 0.8$ . We redraw 10 000 spectral energy distributions in each of these redshift ranges from the new galaxy spectral library, which we use to generate pseudo-observations in the *ugriz* photometric bands, for  $S/N = 30$ . As before, we assume that the redshift  $z_{\text{obs}}$  of a galaxy is unknown a priori and use all models at redshifts between 0 and 1 in the spectral library to retrieve likelihood distributions of physical parameters. According to Fig. 10(b),  $z_{\text{obs}}$  can be retrieved with an uncertainty of roughly 0.05 in both redshift ranges at the adopted  $S/N$ .

Fig. 11 shows the average probability density functions of physical parameters retrieved in this way for galaxies at redshifts  $0.2 < z_{\text{obs}} < 0.4$  (top row) and  $0.6 < z_{\text{obs}} < 0.8$  (bottom row). At low redshift, the constraints are similar to those obtained in Fig. 6 for galaxies at  $z = 0$ . The main difference is the extension of the  $M_*/L_r$  probability density functions to larger values, induced by the tail of high-redshift galaxies with large K corrections in the prior distribution (Fig. 9a). This tends to worsen the typical uncertainty in  $M_*/L_r$  (from 0.19 to 0.22 dex) over the range sampled by Fig. 6, i.e. for  $\log[(M_*/L_r)/(M_\odot/L_{r,\odot})] < 0.4$ , while at larger  $M_*/L_r$ , the smaller number of models implies a significant bias in the retrieved likelihood distribution. Another noticeable difference caused by the absence of redshift information between Fig. 11 (top row) and Fig. 6 is the slightly worse uncertainty and bias in the retrieved specific star formation rate of actively star-forming galaxies. Fig. 11 (bottom row) also shows that, at higher redshift, the loss of H $\alpha$  information further worsens  $\psi_s$  estimates in these galaxies. The constraints on  $M_*/L_r$  are similar to those for low-redshift galaxies. In both redshift ranges in Fig. 11, the parameters  $f_{\text{SFH}}$ ,  $12 + \log(\text{O}/\text{H})$ ,  $\hat{\tau}_v$  and  $\mu$  are as poorly constrained as for  $z = 0$  galaxies in Fig. 6.

We conclude that high-quality photometry in the *ugriz* bands can provide useful constraints on the mass-to-light ratio and specific star formation rate of galaxies, especially at low redshift, but less so on the recent history of star formation and the enrichment in metals and dust of the ISM. The lack of redshift information can significantly weaken these constraints.

### 3.2 Spectroscopic constraints

We now investigate how the constraints obtained above from multi-band photometry on the parameters  $M_*/L_r$ ,  $f_{\text{SFH}}$ ,  $\psi_s$ ,  $12 + \log(\text{O}/\text{H})$ ,  $\hat{\tau}_v$  and  $\mu$  can be improved by appealing to different types of spectroscopic observations. In Section 3.2.1, we first show that the equivalent widths of strong optical emission lines alone can provide useful clues about the specific star formation rate and enrichment of the ISM in star-forming galaxies. Then, in Sections 3.2.2 and 3.2.3, we explore in detail the constraints retrievable from low- and medium-resolution spectroscopy, exploiting the ability with our approach to interpret simultaneously the stellar and nebular emission from galaxies. Throughout this section, we assume that the galaxies have well-determined spectroscopic redshifts and perform all calculations at rest-frame wavelengths in the range between 3600 and 7400 Å (Fig. 3).



**Figure 11.** Average probability density functions of the same six physical parameters as in Fig. 9 retrieved, using the Bayesian approach described in Section 2.4, from five-band *ugriz* photometry with  $S/N = 30$  of samples of 10000 pseudo-galaxies at random redshifts: (top row) drawn in the redshift range  $0.2 < z_{\text{obs}} < 0.4$ ; and (bottom row) drawn in the redshift range  $0.6 < z_{\text{obs}} < 0.8$ . In each case, the redshift  $z_{\text{obs}}$  of a galaxy is assumed not to be known a priori, and the probability density functions are computed using models at all redshifts between 0 and 1 in the spectral library assembled in Section 3.1.2. In all panels, the lines and shading have the same meaning as in Fig. 6.

### 3.2.1 Equivalent widths of strong optical emission lines

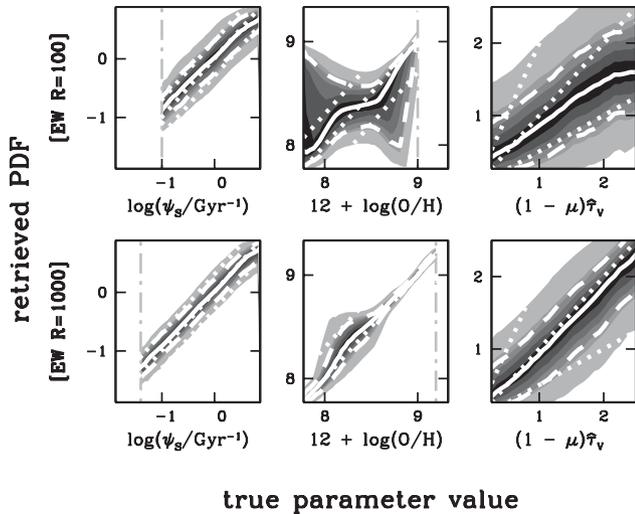
The equivalent widths of optical emission lines reflect the ratio of young stars, which produce ionizing radiation, to older stars, which dominate the optical continuum radiation. These equivalent widths should therefore provide valuable information about the specific star formation rate and, if several lines are available, the gas-phase oxygen abundance and dust attenuation optical depth in star-forming galaxies. Since line equivalent widths do not provide any information about the absolute stellar continuum emission, they cannot be used to constrain the parameters  $M_*/L_r$  or  $f_{\text{SPH}}$ . Similarly, line equivalent widths should not provide any useful constraint on the dust attenuation optical depth in the ambient ISM,  $\mu\hat{\tau}_V$ , which affects in similar ways emission lines emerging from stellar birth clouds and the continuum radiation from old stars (equation 2.8). We therefore focus here on the retrievability of the parameters  $\psi_S$ ,  $12 + \log(O/H)$  and the dust attenuation optical depth in stellar birth clouds,  $(1 - \mu)\hat{\tau}_V$ , from observations of the equivalent widths of strong optical emission lines in star-forming galaxies.

To proceed, we select a sample of 10000 spectral energy distributions from the library assembled in Section 2.3 to compute pseudo-observations of the net emission equivalent widths of the following emission lines:  $[\text{O II}]\lambda 3727$ ;  $\text{H}\beta$ ;  $[\text{O III}]\lambda 4959$ ;  $[\text{O III}]\lambda 5007$ ;  $[\text{N II}]\lambda 6548$ ;  $\text{H}\alpha$ ;  $[\text{N II}]\lambda 6584$ ;  $[\text{S II}]\lambda 6716$  and  $[\text{S II}]\lambda 6731$ . We consider both the cases of low ( $50 \text{ \AA}$  FWHM;  $R = 100$  at  $\lambda = 5000 \text{ \AA}$ ) and medium ( $5 \text{ \AA}$  FWHM;  $R = 1000$  at  $\lambda = 5000 \text{ \AA}$ ) spectral resolution and fix the median  $S/N$  per pixel at  $\overline{S/N} = 20$ . In the case of low spectral resolution, we treat as single emission features the line blends:  $[\text{O III}]\lambda\lambda 4959, 5007$ ;  $[\text{N II}]\lambda 6548 + \text{H}\alpha$  +  $[\text{N II}]\lambda 6584$

and  $[\text{S II}]\lambda\lambda 6716, 6731$ . We also require that, at a given spectral resolution, the net emission equivalent widths of all features be measured at better than the  $3\sigma$  level.<sup>5</sup> We apply the same requirement to the model spectra used to compute likelihood distributions of  $\psi_S$ ,  $12 + \log(O/H)$  and  $(1 - \mu)\hat{\tau}_V$  for each pseudo-galaxy. These requirements are filled by roughly 2.7 million out of the 5 million original models at low resolution, and 1.7 million models at medium resolution (where lines are not blended).

Fig. 12 shows the average retrieved probability density functions of  $\psi_S$ ,  $12 + \log(O/H)$  and  $(1 - \mu)\hat{\tau}_V$  for the 10000 star-forming pseudo-galaxies in our sample, for low (top row) and medium (bottom row) spectral resolution. We find that, as expected, the equivalent widths of strong optical emission lines set tight constraints on the specific star formation rate,  $\psi_S$ . The uncertainty in the retrieved best estimate is only slightly worse at low (0.23 dex) than at medium (0.18 dex) spectral resolution. An advantage of medium spectral resolution is the possibility to constrain  $\psi_S$  in more quiescent star-forming galaxies than at low spectral resolution, for fixed  $\overline{S/N}$ . According to Fig. 12, for  $\overline{S/N} = 20$ , the limit under which the equivalent widths of all lines of interest cannot be measured at the  $3\sigma$  level is  $\psi_S \approx 0.1 \text{ Gyr}^{-1}$  at  $R = 100$  and  $\psi_S \approx 0.04 \text{ Gyr}^{-1}$  at  $R = 1000$ .

<sup>5</sup> To compute net emission equivalent widths in low-resolution (medium-resolution) spectra, we measure the continuum level as follows. We first de-spoke the spectrum with a 9-pixel (11-pixel) median smoothing. We then mask 5-pixel-wide (9-pixel-wide) regions centred on the emission lines, through which we interpolate the spectrum. Finally, we smooth the spectrum again and adopt the result as the continuum level.



**Figure 12.** Average probability density functions of the specific star formation rate,  $\psi_s$ , gas-phase oxygen abundance,  $12 + \log(\text{O}/\text{H})$ , and dust attenuation optical depth in stellar birth clouds,  $(1 - \mu)\hat{\tau}_v$ , retrieved, using the Bayesian approach described in Section 2.4, from the equivalent widths of optical emission lines in a sample of 10 000 star-forming pseudo-galaxies. The sample was extracted from the library of galaxy spectral energy distributions assembled in Section 2.3, assuming a median S/N per pixel  $\bar{S}/\bar{N} = 20$  and requiring  $3\sigma$  measurements of all line equivalent widths. Top row: using the equivalent widths of [O II] $\lambda$ 3727; H $\beta$ ; [O III] $\lambda$ 4959, 5007; [N II] $\lambda$ 6548 + H $\alpha$  + [N II] $\lambda$ 6584 and [S II] $\lambda$ 6716, 6731 at a spectral resolution of 50 Å FWHM ( $R = 100$  at  $\lambda = 5000$  Å). Bottom row: using the equivalent widths of [O II] $\lambda$ 3727; H $\beta$ ; [O III] $\lambda$ 4959; [O III] $\lambda$ 5007; [N II] $\lambda$ 6548; H $\alpha$ ; [N II] $\lambda$ 6584; [S II] $\lambda$ 6716 and [S II] $\lambda$ 6731 at a spectral resolution of 5 Å FWHM ( $R = 1000$  at  $\lambda = 5000$  Å). In all panels, the lines and shading have the same meaning as in Fig. 6.

Fig. 12 (bottom row) further shows that, at medium spectral resolution, the equivalent widths of optical emission lines provide tight constraints on the gas-phase oxygen abundance in star-forming galaxies. The uncertainty in the retrieved best estimate of  $12 + \log(\text{O}/\text{H})$  is only 0.02 at the high-metallicity end. It reaches 0.17 at low  $12 + \log(\text{O}/\text{H})$ , because the rise in intrinsic emission-line intensity with decreasing metallicity (e.g. fig. 4d of Charlot & Longhetti 2001) allows the inclusion of more dusty galaxies in the likelihood distributions. At low spectral resolution, the blending of the H $\alpha$  and [N II] lines worsens considerably these constraints (Fig. 12, top row). Moreover, the drop in emission-line intensity with increasing metallicity implies that  $12 + \log(\text{O}/\text{H})$  can be retrieved in more metal-rich galaxies at medium than at low spectral resolution, for fixed  $\bar{S}/\bar{N}$ . The constraints on the dust attenuation optical depth in stellar birth clouds in Fig. 12 are also much tighter at medium than at low spectral resolution. At low spectral resolution, the median retrieved uncertainty in  $(1 - \mu)\hat{\tau}_v$  is only about 20 per cent larger than at medium spectral resolution, but the best estimate is much more severely biased towards the prior expectation.

### 3.2.2 Low-resolution spectroscopy

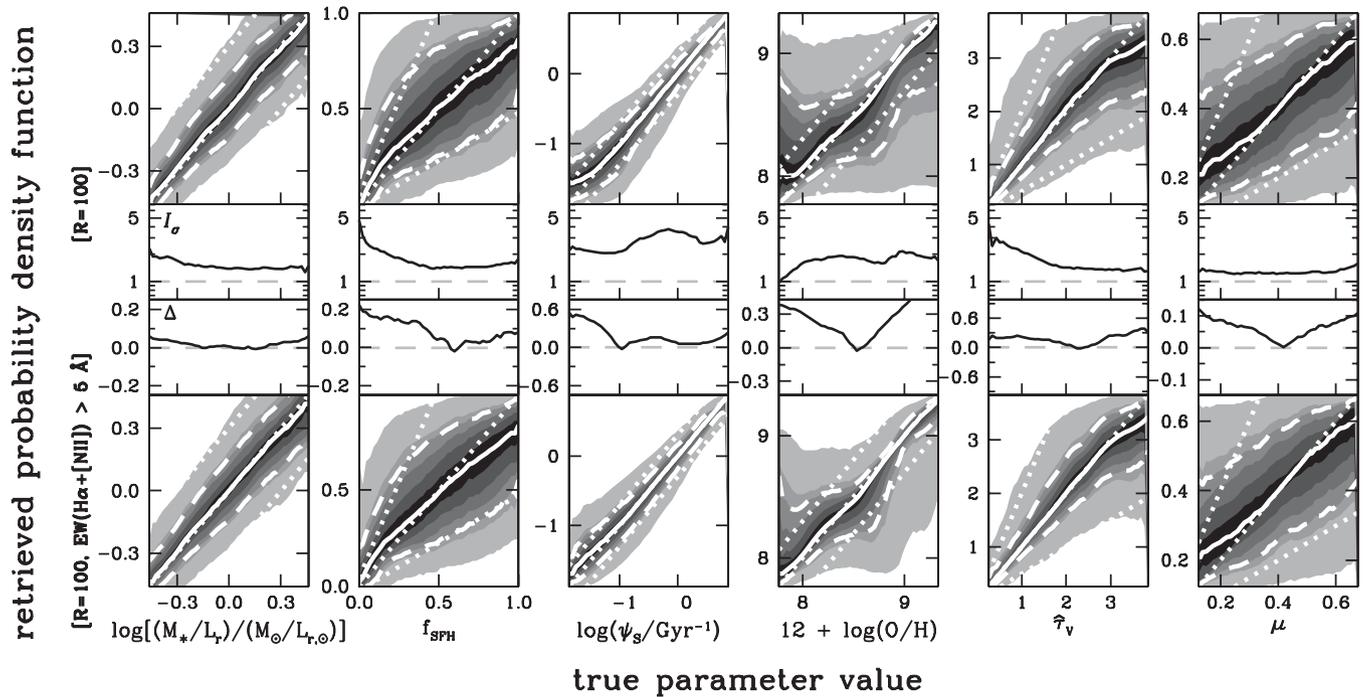
We now turn to one of the most innovative parts of the present study enabled by the approach developed in Section 2: the retrieval of physical parameters from the simultaneous interpretation of strong absorption and emission features in low-resolution spectral energy distribution of galaxies over the entire rest-wavelength range from

$\lambda = 3600$  to  $7400$  Å. As before, we extract 10 000 spectral energy distributions from the library assembled in Section 2.3 to compute pseudo-observations of galaxies at the resolution of 50 Å FWHM ( $R = 100$  at  $\lambda = 5000$  Å) in this wavelength range. We adopt for the moment a median S/N per pixel  $\bar{S}/\bar{N} = 20$ . We use the rest of the 5 million models in the library to retrieve likelihood distributions of physical parameters for each of these 10 000 pseudo-galaxies.

Fig. 13 (top row) shows the average retrieved probability density functions of  $M_*/L_r$ ,  $f_{\text{SFH}}$ ,  $\psi_s$ ,  $12 + \log(\text{O}/\text{H})$ ,  $\hat{\tau}_v$  and  $\mu$  for the 10 000 pseudo-galaxies in our sample. Also shown at the bottom of each panel are the factor of improvement in the retrieved uncertainty,  $I_\sigma$ , and the gain in accuracy,  $\Delta$ , relative to the standard case of Fig. 6 (equations 3.13–3.14). Fig. 13 (top row) reveals that a major improvement of low-resolution spectroscopy over rest-frame *ugriz* photometry is the ability to retrieve meaningful (i.e. unbiased) constraints not only on the stellar mass-to-light ratio and the specific star formation rate, but also on the recent star formation history ( $f_{\text{SFH}}$ ) and the enrichment of metals and dust in the ISM of galaxies [ $12 + \log(\text{O}/\text{H})$  and, to a lesser extent,  $\mu$ ]. The gain in accuracy ( $\Delta$ ) reaches up to 30 per cent of the explored dynamic range for  $12 + \log(\text{O}/\text{H})$ , 15 per cent for  $\hat{\tau}_v$  and  $\mu$  and up to 20 per cent for  $f_{\text{SFH}}$ . This results from the ability to detect and interpret, even at low resolution, the strong Balmer absorption features of intermediate-age stars and the emission-line signatures of interstellar gas. The median uncertainty in the retrieved fraction of current stellar mass formed in the last 2.5 Gyr amounts to about 0.23, that in  $12 + \log(\text{O}/\text{H})$  about 0.28 and that in the total dust attenuation optical depth,  $\hat{\tau}_v$ , about 0.64. For  $M_*/L_r$  and  $\psi_s$ , already constrained in a meaningful way by multi-band photometry, the improvement in uncertainty relative to the standard case amounts to a factor ( $I_\sigma$ ) of between 1.5 and 3.

A more in-depth investigation reveals that, in all panels in Fig. 13 (top row), the largest uncertainties pertain to galaxies with emission lines too weak to allow a distinction between the signatures of star formation history, metallicity and dust in the spectra. These are mainly galaxies with low specific star formation rate, high metallicity and strong attenuation of emission lines by dust in stellar birth clouds. In this context, it is interesting to explore how much the constraints tighten if we restrict the sample to low-resolution galaxy spectra exhibiting unambiguous emission lines. We appeal again to the library assembled in Section 2.3 to generate a sample of 10 000 pseudo-observations of low-resolution galaxy spectra, for  $\bar{S}/\bar{N} = 20$ , with the requirement that the net H $\alpha$ + [N II] emission equivalent width be greater than 5 Å (throughout this paper, we adopt the convention of positive equivalent widths for emission lines). Fig. 13 (bottom row) shows that, in this case, the constraints on the interstellar parameters  $12 + \log(\text{O}/\text{H})$  and  $\hat{\tau}_v$  improve significantly. The corresponding median uncertainties both drop by about 20 per cent. The constraints on  $\psi_s$  improve mainly at the low end of the explored range (the uncertainty dropping by about 15 per cent at  $\psi_s \lesssim 0.1 \text{ Gyr}^{-1}$ ), where the omission of galaxies with weak emission lines is most significant. The improvement is only marginal for  $M_*/L_r$ ,  $f_{\text{SFH}}$  and  $\mu$ , which are constrained primarily by stellar continuum and absorption-line features rather than by nebular emission lines.

The ability with our approach to exploit strong absorption and emission features in low-resolution galaxy spectra depends critically on the quality of the observations. To illustrate this, we show in Fig. 14 (top row) the analogue of Fig. 13 (top row) when adopting a median S/N per pixel  $\bar{S}/\bar{N} = 5$  instead of 20. The reduced ability to interpret spectral features at  $\bar{S}/\bar{N} = 5$  makes the median



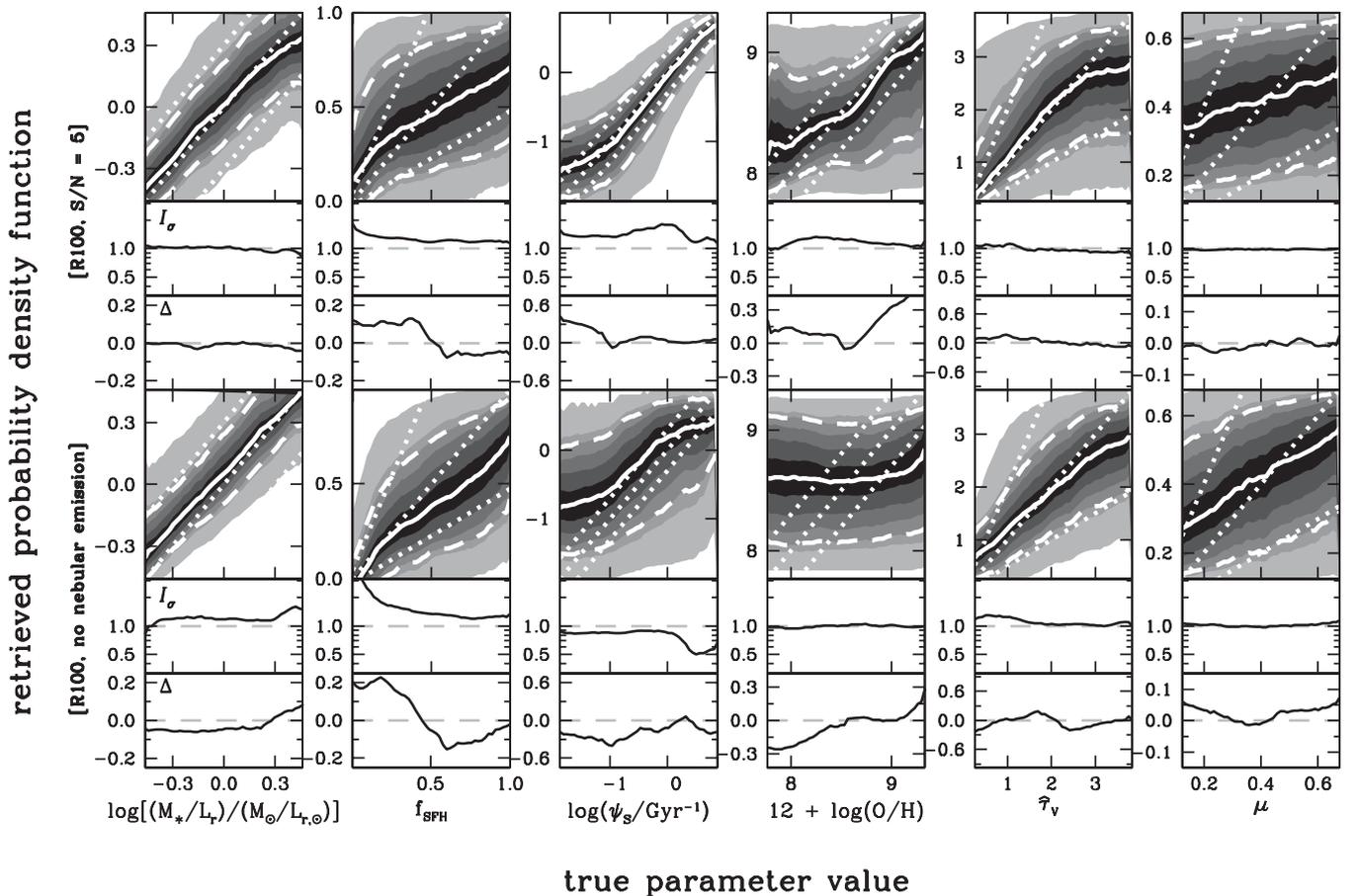
**Figure 13.** Average probability density functions of the same six physical parameters as in Fig. 4 retrieved, using the Bayesian approach described in Section 2.4, from low-resolution optical spectra of a sample of 10 000 pseudo-galaxies. The spectra cover the wavelength range from  $\lambda = 3600$  to  $7400 \text{ \AA}$  at the resolution of  $50 \text{ \AA}$  FWHM ( $R = 100$  at  $\lambda = 5000 \text{ \AA}$ ) with median S/N per pixel  $\overline{S/N} = 20$ . Top row: for a sample of 10 000 pseudo-galaxies extracted randomly from the spectral library assembled in Section 2.3. The improvement factor  $I_\sigma$  and the gain in accuracy  $\Delta$  (equations 3.13–3.14) are shown as a function of true parameter value at the bottom of each panel to quantify differences in the retrieved likelihood distributions relative to the standard case of Fig. 6. Bottom row: for a sample of 10 000 star-forming pseudo-galaxies extracted from the same spectral library, with the requirement that the net  $H\alpha + [N II]$  emission equivalent width be greater than  $5 \text{ \AA}$ . In all panels, the lines and shading have the same meaning as in Fig. 6.

uncertainties in the different retrieved parameters increase by 30 per cent (for  $M_*/L_r, f_{\text{SFH}}, \mu$ ) up to 90 per cent (for  $\psi_S$ ) relative to the case  $\overline{S/N} = 20$ . It also introduces substantial biases in the retrieved likelihood distributions. A comparison of Fig. 14 (top row) with Fig. 6 reveals that, in fact, low-resolution spectroscopy with  $\overline{S/N} = 5$  over the wavelength range from  $\lambda = 3600$  to  $7400 \text{ \AA}$  provides hardly better constraints on the different parameters than high-quality *ugriz* photometry (as confirmed by the fact that  $I_\sigma$  is close to unity and  $\Delta$  close to zero in Fig. 13, top row). The constraints on  $\psi_S$  and  $12 + \log(O/H)$  are somewhat tighter because of the influence of actively star-forming galaxies with strong emission lines on the average retrieved probability density functions. The star formation history parameter is only slightly better constrained, as even strong Balmer absorption lines are difficult to interpret at low  $\overline{S/N}$ . The constraints on  $\tau_v$  are as poor as inferred from multi-band photometry, reflecting the inability to extract useful information from the  $H\alpha/H\beta$  ratio in low-quality spectra. In Fig. 14 (top row), the constraints on  $M_*/L_r$  are actually weaker and more severely biased at the extremities of the explored range than those obtained from rest-frame *ugriz* photometry in Fig. 6. This suggests that high-quality photometry including the near-infrared *iz* bands can better constrain galaxy stellar masses than low-resolution optical spectroscopy with  $\overline{S/N} = 5$ .

The inclusion of nebular emission also has a crucial influence on the ability to constrain galaxy physical parameters from low-resolution optical spectroscopy. Fig. 14 (bottom row) shows the results obtained when attempting to interpret the same pseudo-

observations as in Fig. 13 (top row; which include nebular emission) with the library of models not including nebular emission already used in Fig. 8 (bottom row).<sup>6</sup> In this case, the average retrieved likelihood distributions of all parameters are severely degraded relative to Fig. 13 (top row). In fact, for most parameters, the constraints are worse than those derived in Fig. 6 from high-quality *ugriz* photometry when accounting for the effect of nebular emission on photometric fluxes. Fig. 14 (bottom row) shows that, for example, using models which neglect the contamination of the *r*-band flux by  $H\alpha + [N II]$  and  $[S II]$  line emission biases  $M_*/L_r$  estimates upwards for star-forming galaxies. Moreover, ignoring the contamination of stellar Balmer absorption features by nebular emission in low-resolution spectra causes systematic underestimates of  $f_{\text{SFH}}$  in star-forming galaxies. For actively star-forming galaxies, the constraints on  $\psi_S$  in Fig. 14 (bottom row) are weaker than the rough constraints derived in Fig. 6 from the influence of nebular emission on *ugriz* photometric observations. Not surprisingly, models without nebular emission do not allow the extraction of useful

<sup>6</sup> To avoid artefacts in the resulting probability density functions, we remove prominent emission lines from the pseudo-galaxy spectra before computing the likelihood of models without nebular emission (equation 2.10). To achieve this, we first measure the continuum level across the whole range from  $\lambda = 3600$  to  $7400 \text{ \AA}$ , as described in footnote 5. Then, we reset to the continuum level the flux of every pixel exceeding the continuum flux by more than five times the noise level.



**Figure 14.** Same as Fig. 13, but: (top row) adopting a median S/N per pixel  $\overline{S/N} = 5$  instead of 20; (bottom row) not including nebular emission in the model library used to analyse the sample of 10 000 pseudo-galaxies. In both cases, the improvement factor  $I_\sigma$  and the gain in accuracy  $\Delta$  (equations 3.13–3.14) are shown as a function of true parameter value at the bottom of each panel to quantify differences in the retrieved likelihood distributions relative to the standard case of Fig. 6.

constraints on the interstellar parameters  $12 + \log(O/H)$ ,  $\hat{\tau}_V$  and  $\mu$  from low-resolution galaxy spectra.

Hence, we find that important constraints can be obtained on the stellar and interstellar parameters of galaxies from low-resolution spectra at optical wavelengths. This, however, requires high-quality observations as well as the inclusion of nebular emission in the spectral analysis.

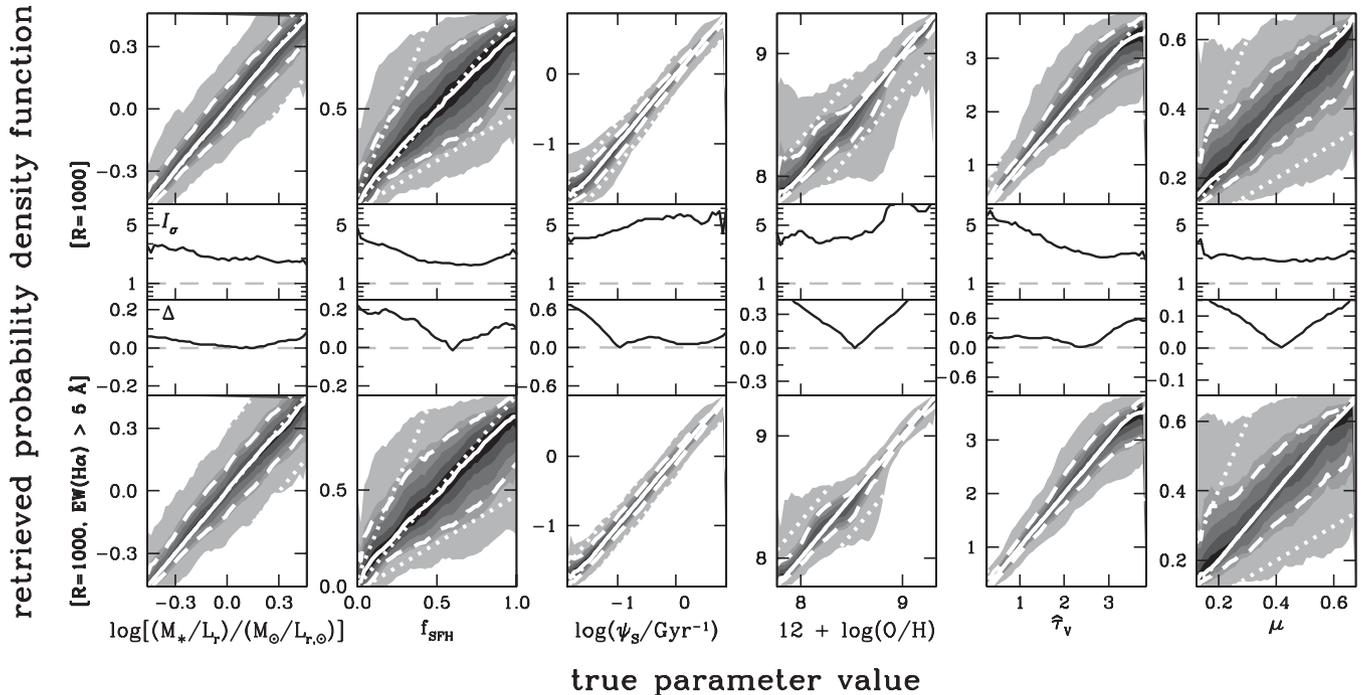
### 3.2.3 Medium-resolution spectroscopy

Medium-resolution spectroscopy should provide even more insight into the physical properties of galaxies than low-resolution spectroscopy, because composite features such as the [O III] doublet, the  $H\alpha + [N II]$  blend and the [S II] doublet can be resolved into individual emission lines. To investigate this, we extract 10 000 spectral energy distributions from the library assembled in Section 2.3 and compute pseudo-observations of galaxies at the resolution of  $5 \text{ \AA}$  FWHM ( $R = 1000$  at  $\lambda = 5000 \text{ \AA}$ ) over the wavelength range from  $\lambda = 3600$  to  $7400 \text{ \AA}$ . We adopt a median S/N per pixel  $\overline{S/N} = 20$ . As before, we use the rest of the 5 million models in the library to retrieve likelihood distributions of physical parameters for each of these 10 000 pseudo-galaxies.

In Fig. 15 (top row), we show the average probability density functions of  $M_*/L_r$ ,  $f_{\text{SFH}}$ ,  $\psi_s$ ,  $12 + \log(O/H)$ ,  $\hat{\tau}_V$  and  $\mu$  retrieved

from the medium-resolution spectra of the 10 000 pseudo-galaxies in our sample. The factor  $I_\sigma$  and the gain  $\Delta$  quantifying the improvements in the retrieved uncertainty and accuracy relative to the standard case of Fig. 6 (equations 3.13–3.14) are also shown at the bottom of each panel. A comparison of Fig. 15 (top row) with Fig. 13 (top row) shows that an increase in spectral resolution by a factor of 10 produces significantly tighter constraints on all parameters. The improvement is most significant for the specific star formation rate, the gas-phase oxygen abundance and the total dust attenuation optical depth, for which the median retrieved uncertainty drops by about 45 per cent from low to medium spectral resolution. The gain is more modest, of the order of 25 per cent, for  $M_*/L_r$ ,  $f_{\text{SFH}}$ , and  $\mu$ . The refined spectral information available at medium resolution also helps in reducing some biases in the retrieved best estimates of  $f_{\text{SFH}}$  (for young galaxies),  $\psi_s$  (for quiescent star-forming galaxies) and  $\mu$  (across most of the explored range) in Fig. 15 (top row) relative to Fig. 13 (top row).

By analogy with Fig. 13, we further show in Fig. 15 (bottom row) the average probability density functions obtained when selecting a sample of 10 000 pseudo-observations of medium-resolution galaxy spectra, for  $\overline{S/N} = 20$ , with the requirement that the net  $H\alpha$  emission equivalent width be greater than  $5 \text{ \AA}$ . The retrieved likelihood distributions of  $12 + \log(O/H)$  and  $\hat{\tau}_V$  are significantly narrower than when including also galaxies with weak emission lines in the



**Figure 15.** Average probability density functions of the same six physical parameters as in Fig. 4 retrieved, using the Bayesian approach described in Section 2.4, from medium-resolution optical spectra of a sample of 10 000 pseudo-galaxies. The spectra cover the wavelength range from  $\lambda = 3600$  to  $7400 \text{ \AA}$  at the resolution of  $5 \text{ \AA}$  FWHM ( $R = 1000$  at  $\lambda = 5000 \text{ \AA}$ ) with median S/N per pixel  $\bar{S}/\bar{N} = 20$ . Top row: for a sample of 10 000 pseudo-galaxies extracted randomly from the spectral library assembled in Section 2.3. The improvement factor  $I_\sigma$  and the gain in accuracy  $\Delta$  (equations 3.13–3.14) are shown as a function of true parameter value at the bottom of each panel to quantify differences in the retrieved likelihood distributions relative to the standard case of Fig. 6. Bottom row: for a sample of 10 000 star-forming pseudo-galaxies extracted from the same spectral library, with the requirement that the net H $\alpha$  emission equivalent width be greater than  $5 \text{ \AA}$ . In all panels, the lines and shading have the same meaning as in Fig. 6.

sample, with median retrieved uncertainties of about 0.08 for  $12 + \log(\text{O}/\text{H})$  and 0.30 for  $\hat{\tau}_v$ . As at low spectral resolution, restricting the sample to galaxies with prominent emission lines reduces the uncertainties in  $\psi_s$  mainly at the low end of the explored range. Changes in the retrieved likelihood distributions of  $M_*/L_r$ ,  $f_{\text{SFH}}$  and  $\mu$  are marginal, because these parameters are constrained primarily by stellar continuum and absorption-line features.

We have also investigated how the constraints obtained in Fig. 15 from the analysis of medium-resolution galaxy spectra depend on the assumed S/N of the observations and the inclusion of nebular emission in the modelling (not shown). We find that adopting a median S/N per pixel  $\bar{S}/\bar{N} = 5$  instead of 20 for the 10 000 pseudo-galaxy spectra makes the median retrieved uncertainties in most parameters increase by 30 to 50 per cent (only 15 per cent for  $M_*/L_r$ ) relative to Fig. 15 (top row). The biases in the retrieved best estimates are less severe than found in Fig. 14 (top row) at low spectral resolution. As expected from Fig. 14 (bottom row), accounting for nebular emission also has a critical influence on the retrievability of the specific star formation rate and the interstellar parameters from medium-resolution galaxy spectra. It is worth pointing out that, at medium (and high) spectral resolution, prominent nebular emission lines can be excised from the spectra and analysed separately from the stellar continuum and absorption-line signatures if no tool is available to analyse the stellar and nebular emission simultaneously (e.g. Kauffmann et al. 2003a; Brinchmann et al. 2004; Gallazzi et al. 2005). This is not the case at low spectral resolution, where the analysis of the stellar and nebular emission must be combined to obtain meaningful constraints on galaxy physical parameters.

## 4 APPLICATION TO A REAL SAMPLE

The conclusions drawn in the previous section about the retrievability of galaxy physical parameters from different types of observations at optical wavelengths rely on sophisticated simulations of galaxy spectral energy distributions. As emphasized in Section 2.4, such pseudo-observations are required to test the retrievability of physical parameters from observables, because the physical parameters of real galaxies cannot be known a priori. In this section, we apply our approach to the interpretation of observed medium-resolution spectra of galaxies from the SDSS DR7. We compare the results obtained through the simultaneous interpretation of stellar and nebular emission (at the resolving power  $R \approx 1000$  considered in Section 3.2.3) with results obtained for the same galaxies from previous separate analyses of nebular emission-line and stellar absorption-line features (at the native resolving power  $R \approx 2000$  of the SDSS observations). We also investigate the influence of the adopted prior distributions on the retrieved physical parameters of the galaxies.

### 4.1 Physical parameters of SDSS galaxies

The SDSS DR7 contains about 1 million galaxy spectra covering the wavelength range from  $3800$  to  $9200 \text{ \AA}$  (observer frame) at a spectral resolution of roughly  $3 \text{ \AA}$  FWHM. Of these, we extract a subsample of 12 660 star-forming galaxies with high-quality spectra as follows. We first reject any duplicate spectrum of a same galaxy, all spectra of merger candidates (i.e. corresponding to

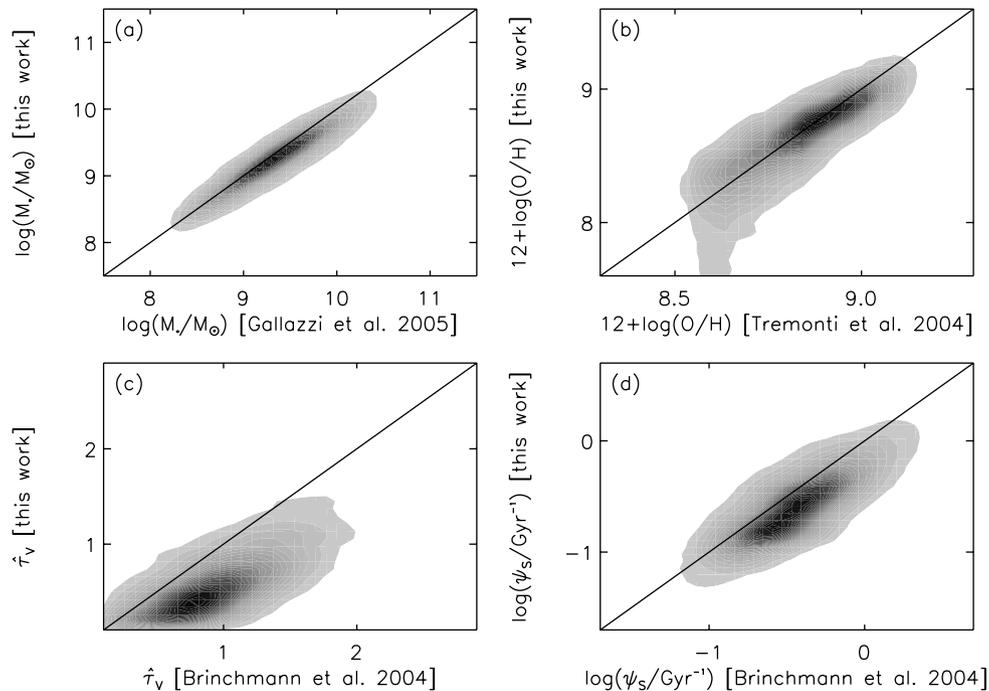
spatial separations less than 1.5 arcsec and redshift separations less than  $200 \text{ km s}^{-1}$ ) and all galaxies at redshift less than  $z = 0.03$  (to ensure that the spectrum includes the  $[\text{O II}]\lambda 3727$  doublet) from the original sample. Among the remaining 829 570 galaxies, we select all emission-line galaxies with S/N greater than 10 in  $[\text{O II}]\lambda 3727$ ,  $\text{H}\beta$ ,  $[\text{O III}]\lambda 5007$ ,  $\text{H}\alpha$ ,  $[\text{N II}]\lambda 6584$  and  $[\text{S II}]\lambda 6716$ . We reject AGNs by applying the conservative criterion of Kauffmann et al. (2003b) in the standard Baldwin et al. (1981) line-diagnostic diagram (see Fig. 2a). The resulting 28 075 star-forming galaxies have a mean redshift of  $0.08 \pm 0.05$ . We exclude all galaxies at redshift  $z > 0.15$  to keep the sample homogeneous and further select galaxies with median S/N per pixel in the range  $5 < \overline{S/N} < 25$ . This leaves 23 213 spectra of galaxies at a mean redshift of  $0.07 \pm 0.03$ , all of which have net  $\text{H}\alpha$  emission equivalent widths greater than  $5 \text{ \AA}$ . We identify 12 660 galaxies in this sample for which good measurements of stellar mass, star formation rate, gas-phase oxygen abundance and dust attenuation optical depth are available from previous studies (see the next paragraph). For the purpose of illustrating the usefulness of our approach to interpret galaxy spectra, we focus on the galaxy physical properties probed by the 3-arcsec-diameter SDSS fibres and do not consider any photometry-based aperture correction.

Estimates of star formation rate, gas-phase oxygen abundance and dust attenuation optical depth are available for the 12 600 SDSS galaxies in this sample from the work of Brinchmann et al. (2004) and Tremonti et al. (2004). These estimates were obtained by fitting the fluxes of the  $[\text{O II}]\lambda 3727$ ,  $\text{H}\beta$ ,  $[\text{O III}]\lambda 5007$ ,  $\text{H}\alpha$ ,  $[\text{N II}]\lambda 6584$  and  $[\text{S II}]\lambda 6716$  emission lines (measured to high precision by adjusting a stellar population model to the spectral continuum) with a large

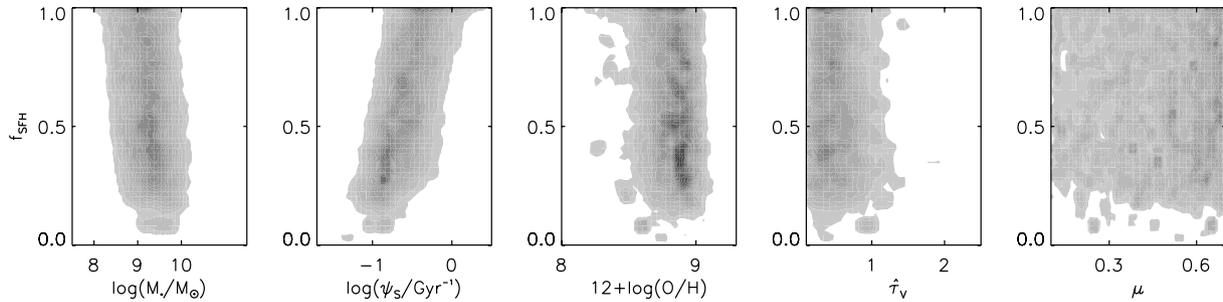
library of Charlot & Longhetti (2001) nebular emission models. In addition, stellar-mass estimates are available for all galaxies from the work of Gallazzi et al. (2005), who fitted the strengths of the  $4000 \text{ \AA}$  break and the  $\text{H}\beta$ ,  $\text{H}\delta_A + \text{H}\gamma_A$ ,  $[\text{MgFe}]'$  and  $[\text{Mg}_2\text{Fe}]$  absorption-line features with a large library of Bruzual & Charlot (2003) stellar population synthesis models. The combination of these stellar-mass estimates with the star formation rate estimates of Brinchmann et al. (2004) provides estimates of the specific star formation rate probed by the SDSS fibre.

The above physical parameters were derived from spectral analyses at the original SDSS resolution of about  $3 \text{ \AA}$  FWHM. To exemplify the power of the approach developed in Section 2, and for the purpose of comparison with the results of Section 3.2.3, we degrade the SDSS spectra to a resolution of  $5 \text{ \AA}$  FWHM, adopting as before a fixed pixel size of  $2.5 \text{ \AA}$ . Also, we focus on the rest-frame wavelength range from  $3700$  to  $7400 \text{ \AA}$ , which does not extend as much in the ultraviolet as that considered in Section 3.2.3 (Fig. 3a), but which does include the  $[\text{O II}]\lambda 3727$  emission doublet. By analogy with the analysis of pseudo-observations in Section 3.2.3, we use the 5 million models in the spectral library assembled in Section 2.3 to retrieve likelihood distributions of physical parameters from the simultaneous analysis of the stellar and nebular emission (equations 2.10–2.12) for each SDSS galaxy in the sample. We note that, when computing likelihood distributions for a given galaxy, we exclude from the library all models older than the age of the Universe at the redshift of that galaxy.

In Fig. 16, we show how the probability density functions of stellar mass, gas-phase oxygen abundance, total V-band attenuation optical depth of the dust and specific star formation rate retrieved



**Figure 16.** Estimates of physical parameters retrieved from the spectra of 12 660 SDSS star-forming galaxies (degraded to a resolution of  $5 \text{ \AA}$  FWHM) using the Bayesian approach described in Section 2.4 plotted against estimates of the same parameters from different sources, as indicated. (a) Stellar mass,  $M_*$ . (b) Gas-phase oxygen abundance,  $12 + \log(\text{O}/\text{H})$ . (c) Total V-band attenuation optical depth of the dust,  $\hat{\tau}_V$ . (d) Specific star formation rate,  $\psi_s$ . In each panel, the likelihood distributions from previous studies were reconstructed from the published 16th, 50th and 84th percentiles and combined with the likelihood distributions obtained in this work to generate 2D probability density functions. The contours depict the normalized co-added 2D probability density function including all 12 660 galaxies in the sample (on a linear scale 25 levels, the outer edge of the lightest grey level corresponding to the 96th percentile of the 2D distribution). The solid line is the identity relation.



**Figure 17.** 2D probability density functions of the fraction of the current stellar mass formed during the last 2.5 Gyr,  $f_{\text{SFH}}$ , versus stellar mass,  $M_*$ , specific star formation rate,  $\psi_S$ , gas-phase oxygen abundance,  $12 + \log(\text{O}/\text{H})$ , total  $V$ -band attenuation optical depth of the dust,  $\hat{\tau}_V$ , and fraction of  $\hat{\tau}_V$  arising from dust in the ambient ISM,  $\mu$ . All parameters were retrieved from the spectra of the same 12 660 SDSS star-forming galaxies as in Fig. 16 (degraded to a resolution of  $5 \text{ \AA}$  FWHM) using the Bayesian approach described in Section 2.4. In each panel, the contours depict the normalized co-added 2D probability density function including all 12 660 galaxies in the sample (on a linear scale of 25 levels, the outer edge of the lightest grey level corresponding to the 96th percentile of the 2D distribution). All physical quantities pertain to the regions of galaxies probed by the 3-arcsec-diameter SDSS fibres.

from our analysis compare with those obtained in previous separate studies of the stellar and nebular emission at higher spectral resolution. Fig. 16(a) shows that, for example, our estimates of  $M_*$  (obtained by multiplying  $M_*/L_r$  by the absolute galaxy luminosity in the observed  $r$  band) agree well with those derived by Gallazzi et al. (2005) from the analysis of selected absorption-line indices. The slightly lower values derived here ( $\sim 0.1$  dex) are attributable to the updated stellar population synthesis prescription (Section 2.2.1). We also find good agreement in Fig. 16(b) between our estimates of  $12 + \log(\text{O}/\text{H})$  and those obtained by Tremonti et al. (2004) from the analysis of selected emission-line luminosities. The discrepancies at low metallicity at  $12 + \log(\text{O}/\text{H}) < 8.7$  are consistent with the uncertainty expected at low metallicity from interpretations of medium-resolution galaxy spectra (Fig. 15, bottom row). In Figs 16(c) and (d), the likelihood distributions retrieved from our analysis favour systematically lower  $\hat{\tau}_V$  and  $\psi_S$  than derived by Brinchmann et al. (2004) from the analysis of emission-line luminosities. This is because Brinchmann et al. (2004) adopt a shallower dust attenuation curve ( $\hat{\tau}_\lambda \propto \lambda^{-0.7}$  both in stellar birth clouds and in the ambient ISM) than that resulting typically from equations (2.7)–(2.9). This makes the inferred dust optical depth and the correction for obscured star formation smaller at fixed observed reddening.

Fig. 16 therefore shows that the likelihood distributions of  $M_*$ ,  $12 + \log(\text{O}/\text{H})$ ,  $\hat{\tau}_V$  and  $\psi_S$  retrieved using our approach agree well (modulo some understood offsets) with those obtained in previous separate analyses of the stellar and nebular emission of SDSS galaxies at twice higher spectral resolution. Such an agreement is expected in part from the fact that these previous analyses rely on earlier versions of the same stellar population synthesis code and nebular emission models as adopted here (Charlot & Longhetti 2001; Bruzual & Charlot 2003; see Sections 2.2.1 and 2.2.2). What is most remarkable is the demonstrated ability with our approach to interpret simultaneously the stellar and nebular emission from galaxies without having to excise nebular emission lines from noisy spectra. As shown in Section 3, this ability extends to low-resolution galaxy spectra and even to multi-band photometry.

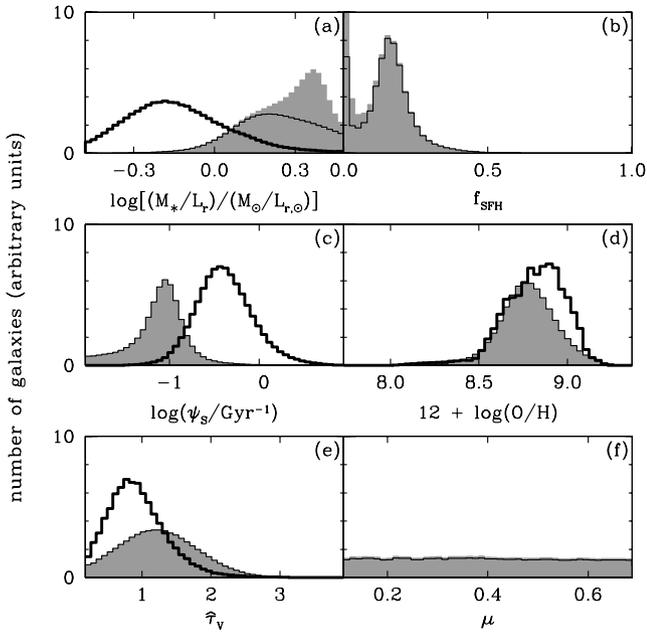
It is of interest to see how the fraction of the current stellar mass formed during the last 2.5 Gyr, which has not been investigated previously, relates to the other physical parameters of SDSS galaxies retrieved using our approach. In Fig. 17, we show the 2D probability density functions of  $f_{\text{SFH}}$  versus  $\log M_*$ ,  $\log \psi_S$ ,  $12 + \log(\text{O}/\text{H})$ ,  $\hat{\tau}_V$  and  $\mu$  for the 12 660 SDSS star-forming galaxies in our sample. The Spearman rank correlation coefficients (evaluated using the median estimates of the different parameters) are  $r_s = -0.27$  for the re-

lation between  $f_{\text{SFH}}$  and  $\log M_*$  and  $+0.59$  for that between  $f_{\text{SFH}}$  and  $\log \psi_S$ . For this sample size, both correlations are significant at better than the  $5\sigma$  level. In contrast, none of the interstellar parameters  $12 + \log(\text{O}/\text{H})$ ,  $\hat{\tau}_V$  and  $\mu$  correlates significantly with  $f_{\text{SFH}}$  (we note in passing that the required high-S/N detection of six emission lines tends to select galaxies with low  $\hat{\tau}_V$  and high  $\mu$ ). Taken at face value, the results of Fig. 17 would suggest that the least massive galaxies, which are the most actively star-forming ones today, have also formed the largest proportion of stars over the last 2.5 Gyr. Such a trend would be consistent with the general expectation that, as galaxies grow through star formation, the exhaustion of the gas supply makes the rate of star formation drop gradually (see also e.g. fig. 22 of Brinchmann et al. 2004). However, this conclusion must be taken with caution, as the present sample is not complete and the SDSS fibres probe only the inner parts of galaxies.

#### 4.2 Influence of the prior distributions of physical parameters

The constraints derived above on the stellar and interstellar properties of SDSS galaxies, along with the analyses of pseudo-observations carried out in Section 3, rely on the adoption of the prior distributions of galaxy physical parameters in Fig. 4. As outlined in Section 2.4, these distributions are designed to be as flat as possible to make the spectral library appropriate for the retrieval of physical parameters from observed spectral energy distributions of galaxies in wide ranges of physical properties. To illustrate the importance of this choice, we examine here the results obtained when adopting different prior distributions reflecting more specific properties of the galaxy population. As an example, we consider the prior distributions of galaxy physical parameters derived from the original semi-analytic recipes of De Lucia & Blaizot (2007).

In Section 2.1, we used the semi-analytic recipes of De Lucia & Blaizot (2007) to generate the star formation and chemical enrichment histories of 500 000 galaxies extracted from the Millennium cosmological simulation of Springel et al. (2005). The parameters  $M_*/L_r$ ,  $f_{\text{SFH}}$ ,  $\psi_S$  and the interstellar metallicity  $Z$  are computed by default in this procedure. De Lucia & Blaizot (2007) do not provide any recipe to compute the gas-phase oxygen abundance nor the ionization parameter of the gas heated by young stars. Also, while attenuation by dust is included using the two-component model of Charlot & Fall (2000, adopting a fixed attenuation curve  $\hat{\tau}_\lambda \propto \lambda^{-0.7}$  both in stellar birth clouds and in the ambient ISM), we find that the recipe used by De Lucia & Blaizot (2007) to relate the  $V$ -band dust optical depth to the galaxy gas mass, metallicity and orientation



**Figure 18.** Prior distributions of the same physical parameters as in Fig. 4 for the 5 million galaxies in the spectral library generated in Section 4.2. These distributions are those arising, at the mean redshift  $z = 0.07$  of the SDSS sample of Section 4.1, from the original semi-analytic post-treatment of the Millennium cosmological simulation of Springel et al. (2005) by De Lucia & Blaizot (2007, see text for detail). In each panel, the shaded and thin solid histograms have the same meaning as in Fig. 4. In panels (a), (c), (d) and (e), the thick solid histograms show the distributions of the best estimates of  $M_*/L_r$ ,  $\psi_S$ ,  $12 + \log(O/H)$  and  $\hat{\tau}_V$  from Gallazzi et al. (2005), Brinchmann et al. (2004) and Tremonti et al. (2004) for the same sample of 12 660 SDSS star-forming galaxies as in Fig. 16.

produces typically much lower  $\hat{\tau}_V$  than observed in SDSS galaxies. To compute prior distributions of interstellar parameters associated with the original semi-analytic recipes of De Lucia & Blaizot (2007), we therefore proceed as follows. For each of the 500 000 galaxies in the library, we draw 10 different realizations of  $\log U_0$ ,  $\xi_d$  and  $\mu$  from the distributions in Table 1. We also draw 10 different realizations of  $\hat{\tau}_V$  from a Gaussian distribution centred on 1.2 with a standard deviation of 0.6. This roughly represents the global properties of SDSS galaxies for a fixed attenuation curve  $\hat{\tau}_\lambda \propto \lambda^{-0.7}$  in stellar birth clouds and in the ambient ISM (Brinchmann et al. 2004). The new spectral library contains 5 million models, which we can use to interpret observations of galaxy spectral energy distributions in the same way as the library assembled in Section 2.3.

Fig. 18 shows the prior distributions of  $M_*/L_r$ ,  $f_{\text{SFH}}$ ,  $\psi_S$ ,  $12 + \log(O/H)$ ,  $\hat{\tau}_V$  and  $\mu$  in this library at the mean redshift  $z = 0.07$  of the SDSS sample of Section 4.1. In each panel, the shaded histogram shows the distribution including all galaxies, while the thin solid histogram shows the contribution by star-forming galaxies alone. As noted in Section 2.1, 35 per cent of the galaxies in the original library of De Lucia & Blaizot (2007) do not form stars at low redshift, while the distributions of  $\psi_S$  and  $f_{\text{SFH}}$  for the remaining galaxies are narrow and characteristic of quiescent star formation (Figs 18b and c). As a result, most galaxies have relatively large  $M_*/L_r$  (Fig. 18a). In Fig. 18(d), the narrow distribution of  $12 + \log(O/H)$  arises from the correlation of the interstellar metallicity  $Z$  with  $\psi_S$  for star-forming galaxies in this library. Also shown as thick solid histograms in Fig. 18 are the distributions of the best estimates of  $M_*/L_r$ ,  $\psi_S$ ,  $12 + \log(O/H)$  and  $\hat{\tau}_V$  from Gallazzi et al.

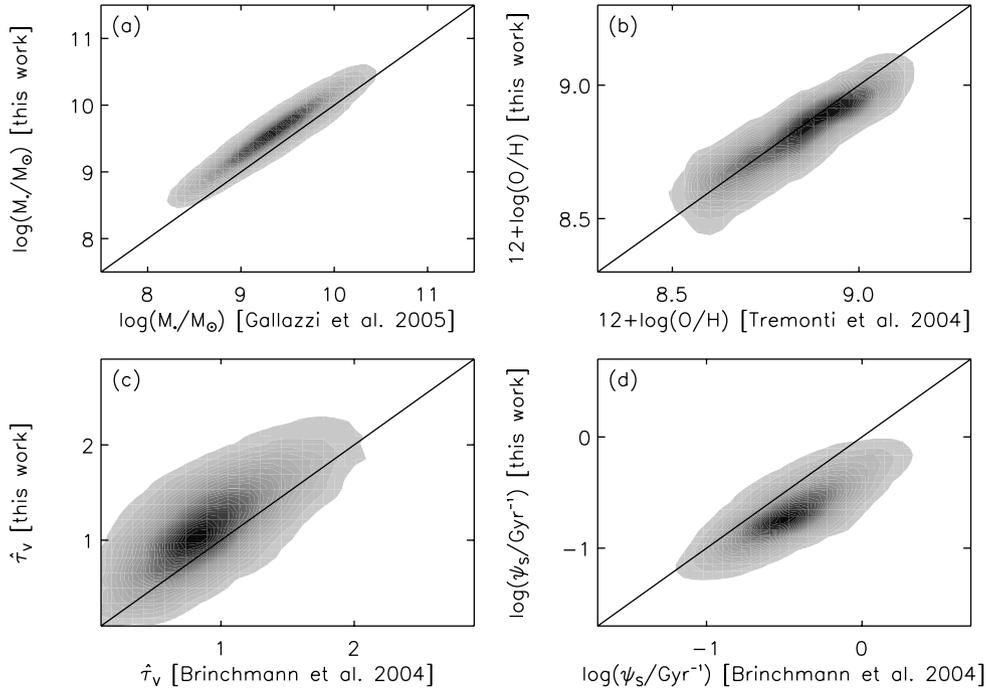
(2005), Brinchmann et al. (2004) and Tremonti et al. (2004) for the 12 660 SDSS star-forming galaxies of Section 4.1. The required detection of six emission lines with  $S/N > 10$  in these galaxies implies that they are typically more actively star-forming (and hence have lower  $M_*/L_r$ ) and less dusty than the bulk of SDSS galaxies the Millennium cosmological simulation is expected to represent. This is the origin of the differences between the shaded and thick solid histograms in Fig. 18.

In Fig. 19, we show the analogue of Fig. 16 obtained when using the prior distributions in Fig. 18, instead of those in Fig. 4, to retrieve the probability density functions of  $M_*$ ,  $12 + \log(O/H)$ ,  $\hat{\tau}_V$  and  $\psi_S$  for the 12 660 SDSS star-forming galaxies in our sample. The upward bias in the prior distribution of  $M_*/L_r$  relative to the observed sample (Fig. 18a) causes the estimates of  $M_*$  in Fig. 19(a) to be systematically larger than those derived by Gallazzi et al. (2005). In Fig. 19(d), the retrieved  $\psi_S$  is about 60 per cent smaller than that estimated by Brinchmann et al. (2004), as in Fig. 16(d). The offset in this case results from the downward bias in the prior distribution of  $\psi_S$  (Fig. 18c) rather than from differences in the dust attenuation curve, which is assumed to be the same by De Lucia & Blaizot (2007) and Brinchmann et al. (2004). Meanwhile, the upward bias in the prior distribution of  $\hat{\tau}_V$  (Fig. 18e) causes estimates of this parameter in Fig. 19(c) to be systematically larger than those derived by Brinchmann et al. (2004). In Fig. 19(b), the estimates of  $12 + \log(O/H)$  retrieved using the prior distribution of Fig. 18(d) agree well with those derived by Tremonti et al. (2004).

The results of Figs 16 and 19 illustrate the importance of the choice of appropriate prior distributions when applying the Bayesian approach described in Section 2.4 to retrieve physical parameters from observed spectral energy distributions of galaxies. While the prior distributions of Fig. 18 are expected to be representative of the SDSS sample as a whole, adopting these distributions to analyse subsamples of galaxies with properties different from the bulk can induce unwanted offsets in the retrieved physical parameters. In cases where some specific physical property of the observed galaxies can be known in advance (e.g. actively star-forming, early-type, metal-poor), the use of appropriate restrictive prior distributions can help reduce the uncertainties in the retrieved parameters. When no advance information is available, it is preferable to appeal to flat (maximum-ignorance) prior distributions encompassing wide ranges of galaxy physical properties.

## 5 SUMMARY AND CONCLUSION

We have developed a new approach to constrain galaxy physical parameters from the combined interpretation of stellar and nebular emission in wide ranges of photometric and spectroscopic observations. Our approach relies on the simulation of a comprehensive library of 5 million galaxy spectral energy distributions using a set of state-of-the-art models of star formation and chemical enrichment histories, stellar population synthesis, nebular emission and attenuation by dust. This library can be used to retrieve probability density functions of physical parameters from the Bayesian analysis of any type of multi-wavelength galaxy observation. We focus in this paper on the retrievability of the observer-frame absolute  $r$ -band stellar mass-to-light ratio ( $M_*/L_r$ ), the fraction of a current galaxy stellar mass formed during the last 2.5 Gyr ( $f_{\text{SFH}}$ ), the specific star formation rate ( $\psi_S$ ), the gas-phase oxygen abundance [ $12 + \log(O/H)$ ], the total effective  $V$ -band absorption optical depth of the dust ( $\hat{\tau}_V$ ) and the fraction of this arising from dust in the ambient ISM ( $\mu$ ) from different types of observations at optical wavelengths: five-band *ugriz* photometry; equivalent-width measurements of



**Figure 19.** Same as Fig. 16, but using the prior distributions in Fig. 18 instead of those in Fig. 4 to retrieve the probability density functions of  $M_*$ ,  $12 + \log(\text{O}/\text{H})$ ,  $\hat{\tau}_V$  and  $\psi_S$  for the 12 660 SDSS star-forming galaxies in the sample.

strong emission lines; low-resolution spectroscopy (50 Å FWHM over the wavelength range  $\lambda = 3600\text{--}7400$  Å) and medium-resolution spectroscopy (5 Å FWHM over the same wavelength range).

Since we cannot know the true properties of any sample of observed galaxies, assessing the retrievability of galaxy physical parameters from these different types of observations requires the simulation of pseudo-observations. We therefore simulate pseudo-observations by convolving the spectral energy distributions of models with known parameters with appropriate instrument responses and then applying artificial noise to mimic true observations. For each type of photometric or spectroscopic observation, we simulate 10 000 pseudo-observations in this way and use the other  $\sim 5$  million models in the library to retrieve likelihood distributions of  $M_*/L_r$ ,  $f_{\text{SFH}}$ ,  $\psi_S$ ,  $12 + \log(\text{O}/\text{H})$ ,  $\hat{\tau}_V$  and  $\mu$  for each pseudo-galaxy (using equations 2.10–2.11). Table 2 summarizes the main conclusions of our study. For each parameter and each type of observation, we list the median retrieved uncertainty (defined as half the 16th–84th percentile range of the retrieved likelihood distribution) and accuracy (defined as the absolute difference between the retrieved best estimate and true parameter value) determined from the analysis of the 10 000 pseudo-observations in Section 3.

Several interesting conclusions can be drawn from Table 2. We find that, for example, rest-frame *ugriz* photometry can provide roughly 50 per cent uncertainty on  $M_*/L_r$  measurements, i.e. almost as tight as the 25 to 35 per cent uncertainty provided by medium- and low-resolution optical spectroscopy (photometric constraints worsen somewhat if the galaxy redshift is not known; see Fig. 11). Meanwhile, optical spectroscopy is required to constrain  $f_{\text{SFH}}$  through the strong absorption-line signatures of intermediate-age stars. Table 2 also shows that observations of the net emission equivalent widths of prominent optical lines are sufficient to constrain  $\psi_S$  to within 50–70 per cent (depending on the spectral resolution) and  $12 + \log(\text{O}/\text{H})$  to within 0.05–0.23 in star-forming

galaxies. Optical spectroscopy can help reduce these uncertainties and the potential small offsets in the retrieved best estimates of  $\psi_S$  and  $12 + \log(\text{O}/\text{H})$ . In addition, spectroscopy is required to constrain both parameters in quiescent star-forming galaxies with weak emission lines ( $\psi_S \lesssim 0.07 \text{ Gyr}^{-1}$ ; see Fig. 12). Medium-resolution spectroscopy provides the best constraints on dust parameters in Table 2, with uncertainties of typically 0.35 in  $\hat{\tau}_V$  and 0.10 in  $\mu$  (these include uncertainties in the optical properties and spatial distribution of the dust as well as in the orientation of the galaxy; see Section 2.2.3). At lower spectral resolution, the less distinct signatures of emission lines make the uncertainty in  $\hat{\tau}_V$  rise by about 80 per cent, while the equivalent widths of optical emission lines at any spectral resolution constrain only the part of the attenuation arising from stellar birth clouds.

We stress that the constraints in Table 2 rely critically on the ability with our approach to interpret simultaneously the stellar and nebular emission from galaxies. In the case of low-resolution spectroscopy, no reliable constraint can be obtained on any of the physical parameters in Table 2 (except for the mass-to-light ratio of early-type galaxies) when nebular emission is not included in the analysis (Fig. 14). The same is true at medium spectral resolution, although if the resolution is high enough, prominent nebular emission lines can be excised from the spectra and analysed separately from the stellar continuum and absorption-line signatures. We find that the likelihood distributions of  $M_*$ ,  $12 + \log(\text{O}/\text{H})$ ,  $\hat{\tau}_V$  and  $\psi_S$  retrieved from the combined analysis of stellar and nebular emission in the spectra of 12 660 SDSS star-forming galaxies (degraded to a resolution of 5 Å FWHM) using our approach agree well (modulo some recent model improvements) with those obtained in previous separate analyses of the stellar and nebular emission of these galaxies at twice higher spectral resolution (Brinchmann et al. 2004; Tremonti et al. 2004; Gallazzi et al. 2005, see Fig. 16). Accounting for nebular emission is less crucial when deriving constraints on  $M_*/L_r$  from *ugriz* photometry (Fig. 8, bottom row).

**Table 2.** Summary of the constraints on galaxy physical parameters retrieved from different types of optical observations using the method developed in Section 2. For each parameter and each type of observation, the quoted uncertainty (defined as half the 16th–84th percentile range of the retrieved likelihood distribution) and accuracy (defined as the absolute difference between the retrieved best estimate and true parameter value) are median quantities determined from the analysis of 10 000 pseudo-observations in Section 3.

Parameter	Constraint	Photometry <sup>a</sup>	Net emission-line EWs <sup>b</sup>		Spectroscopy <sup>b</sup>			
		rest-frame <i>ugriz</i>	$R = 100$	$R = 1000$	$R = 100$		$R = 1000$	
		All <sup>c</sup>	SF <sup>d</sup>	SF <sup>e</sup>	All <sup>c</sup>	SF <sup>f</sup>	All <sup>c</sup>	SF <sup>g</sup>
$\log[(M_*/L_r)/(M_\odot/L_{r,\odot})]$	Uncertainty	0.19	–	–	0.13	0.14	0.10	0.11
	Accuracy	0.04	–	–	0.01	0.01	0.00	0.00
$f_{\text{SFH}}$	Uncertainty	0.36	–	–	0.23	0.23	0.17	0.19
	Accuracy	0.17	–	–	0.05	0.05	0.02	0.03
$\log(\psi_S/\text{Gyr}^{-1})$	Uncertainty	0.67	0.23	0.18	0.24	0.22	0.12	0.12
	Accuracy	0.15	0.12	0.05	0.01	0.00	0.00	0.00
$12 + \log(\text{O}/\text{H})$	Uncertainty	0.50	0.23	0.05	0.28	0.21	0.13	0.08
	Accuracy	0.33	0.11	0.03	0.04	0.03	0.00	0.00
$\hat{\tau}_V$	Uncertainty	0.91	0.48 <sup>h</sup>	0.41 <sup>h</sup>	0.64	0.52	0.35	0.30
	Accuracy	0.21	0.57 <sup>h</sup>	0.14 <sup>h</sup>	0.05	0.00	0.00	0.00
$\mu$	Uncertainty	0.20	–	–	0.16	0.15	0.10	0.10
	Accuracy	0.10	–	–	0.04	0.02	0.00	0.00

<sup>a</sup> For fixed S/N = 30 in all bands.

<sup>b</sup> For median S/N per pixel  $\overline{S/N} = 20$  across the wavelength range  $\lambda = 3600\text{--}7400 \text{ \AA}$ .

<sup>c</sup> Averaged over all galaxy types.

<sup>d</sup> For greater-than- $3\sigma$  detections of [O II] $\lambda$ 3727; H $\beta$ ; [O III] $\lambda$  $\lambda$ 4959, 5007; [N II] $\lambda$ 6548 + H $\alpha$  + [N II] $\lambda$ 6584 and [S II] $\lambda$  $\lambda$ 6716, 6731.

<sup>e</sup> For greater-than- $3\sigma$  detections of [O II] $\lambda$ 3727; H $\beta$ ; [O III] $\lambda$ 4959; [O III] $\lambda$ 5007; [N II] $\lambda$ 6548; H $\alpha$ ; [N II] $\lambda$ 6584; [S II] $\lambda$ 6716 and [S II] $\lambda$ 6731.

<sup>f</sup> For galaxies with net H $\alpha$ + [N II] emission equivalent width greater than  $5 \text{ \AA}$ .

<sup>g</sup> For galaxies with net H $\alpha$  emission equivalent width greater than  $5 \text{ \AA}$ .

<sup>h</sup> Constraint on the birth-cloud component of the attenuation only (i.e.  $\hat{\tau}_V^{\text{BC}}$  and not  $\hat{\tau}_V$ ).

Another important factor we can investigate in a straightforward way using our approach is the influence of S/N on the retrieval of galaxy physical parameters from different types of observations. For multi-band photometry, we find that lowering the S/N from 30 to 10 makes the uncertainty in the retrieved  $M_*/L_r$  in Table 2 larger by about 15 per cent (Fig. 8). For spectroscopy, adopting a median S/N per pixel  $\overline{S/N} = 5$  instead of 20 makes the median uncertainties in the different retrieved parameters increase by about 30 to 90 per cent and introduces substantial biases in the retrieved likelihood distributions (Fig. 14). The ability to make such assessments is valuable as it can help decide between different strategies when designing an observational campaign. For example, as noted in Section 3.2.2, low-resolution spectroscopy with  $\overline{S/N} = 5$  over the wavelength range from  $\lambda = 3600$  to  $7400 \text{ \AA}$  provides hardly better constraints on the different parameters in Table 2 than high-quality *ugriz* photometry (Figs 6 and 14). In contrast, the constraints from low-resolution spectroscopy with  $\overline{S/N} = 20$  over the same wavelength range are generally far superior to those obtained from multi-band photometry (Table 2).

It is important to keep in mind that the results reported in Table 2 about the retrievability of galaxy physical parameters from different types of observations correspond to the optimistic case in which the models we rely on are good approximations of true galaxies (Section 2.4). We have ensured this in the best possible way by appealing to state-of-the-art models including the most recent progress in galaxy spectral modelling to generate the library of spectral energy distributions on which our approach is based. With this in mind, the tool we have developed should be useful not only to interpret existing galaxy data sets, but also to design future observations. In principle, such observations can be of any photometric or spectroscopic type (or a combination thereof) across the wave-

length range covered by spectral evolution models. We have focused in this paper on a few examples of photometric and spectroscopic observations at rest-frame optical wavelengths. In future work, we will present applications of our approach to the interpretation of the ultraviolet and infrared emission from galaxies at various redshifts. The tool we have developed is intended to be made available to the general astronomical community. In the meantime, we encourage colleagues interested in the application of our approach to the interpretation of specific galaxy observations to contact us.

## ACKNOWLEDGMENTS

We thank Maryam Shirazi for her help in the assembly of the SDSS galaxy sample. We also thank Marijn Franx, Brent Groves and Vivienne Wild for helpful discussions. This work was funded in part by the Marie Curie Initial Training Network ELIXIR of the European Commission under contract PITN-GA-2008-214227.

Funding for the SDSS has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the US Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society and the Higher Education Funding Council for England. The SDSS Web site is <http://www.sdss.org>. The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, the Astrophysical Institute Potsdam, the University of Basel, Cambridge University, Case Western Reserve University, the University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and

Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences, Los Alamos National Laboratory, the Max Planck Institute for Astronomy, the Max Planck Institute for Astrophysics, New Mexico State University, Ohio State University, the University of Pittsburgh, the University of Portsmouth, Princeton University, the US Naval Observatory and the University of Washington.

## REFERENCES

- Abazajian K. N. et al., 2009, *ApJS*, 182, 543  
 Baldwin J. A., Phillips M. M., Terlevich R., 1981, *PASP*, 93, 5  
 Bell E. F., de Jong R. S., 2001, *ApJ*, 550, 212  
 Bertelli G., Girardi L., Marigo P., Nasi E., 2008, *A&A*, 484, 815  
 Bertelli G., Nasi E., Girardi L., Marigo P., 2009, *A&A*, 508, 355  
 Blaizot J., Guiderdoni B., Devriendt J. E. G., Bouchet F. R., Hatton S. J., Stoehr F., 2004, *MNRAS*, 352, 571  
 Blitz L., Shu F. H., 1980, *ApJ*, 238, 148  
 Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, *MNRAS*, 351, 1151  
 Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000  
 Chabrier G., 2003, *PASP*, 115, 763  
 Charlot S., Fall S. M., 2000, *ApJ*, 539, 718  
 Charlot S., Longhetti M., 2001, *MNRAS*, 323, 887  
 Cid Fernandes R., Mateus A., Sodré L., Stasińska G., Gomes J. M., 2005, *MNRAS*, 358, 363  
 Coelho P., Bruzual G., Charlot S., Weiss A., Barbuy B., Ferguson J. W., 2007, *MNRAS*, 382, 498  
 Colless M. et al., 2001, *MNRAS*, 328, 1039  
 Conroy C., White M., Gunn J. E., 2010, *ApJ*, 708, 58  
 Croton D. J. et al., 2006, *MNRAS*, 367, 864  
 da Cunha E., Charlot S., Elbaz D., 2008, *MNRAS*, 388, 1595  
 Davis M. et al., 2003, in Guhathakurta P., ed., *Proc. SPIE Conf. Vol. 4834, Discoveries and Research Prospects from 6- to 10-Meter-Class Telescopes II*. SPIE, Bellingham, p. 161  
 De Lucia G., Blaizot J., 2007, *MNRAS*, 375, 2  
 Ferland G. J., 1996, in Ferland G. J., ed., *Hazy, A Brief Introduction to Cloudy 90*. Internal Report, Univ. Kentucky, Lexington  
 Gallazzi A., Charlot S., Brinchmann J., White S. D. M., Tremonti C. A., 2005, *MNRAS*, 362, 41  
 González Delgado R. M., Cerviño M., Martins L. P., Leitherer C., Hauschildt P. H., 2005, *MNRAS*, 357, 945  
 González-Lópezlira R. A., Bruzual-A. G., Charlot S., Ballesteros-Paredes J., Loinard L., 2010, *MNRAS*, 403, 1213  
 Goranova Y. et al., 2009, <http://terapix.iap.fr/cpl/T0006-doc.pdf>  
 Heckman T. M., Kauffmann G., Brinchmann J., Charlot S., Tremonti C., White S. D. M., 2004, *ApJ*, 613, 109  
 Ilbert O. et al., 2006, *A&A*, 457, 841  
 Ilbert O. et al., 2009, *ApJ*, 690, 1236  
 Kauffmann G. et al., 2003a, *MNRAS*, 341, 33  
 Kauffmann G. et al., 2003b, *MNRAS*, 346, 1055  
 Kawamura A. et al., 2009, *ApJS*, 184, 1  
 Kewley L. J., Dopita M. A., 2002, *ApJS*, 142, 35  
 Kong X., Charlot S., Brinchmann J., Fall S. M., 2004, *MNRAS*, 349, 769  
 Lamareille F., Contini T., Le Borgne J.-F., Brinchmann J., Charlot S., Richard J., 2006, *A&A*, 448, 893  
 Le Fèvre O. et al., 2005, *A&A*, 439, 845  
 Maraston C., 2005, *MNRAS*, 362, 799  
 Marigo P., Girardi L., Bressan A., Groenewegen M. A. T., Silva L., Granato G. L., 2008, *A&A*, 482, 883  
 Panter B., Jimenez R., Heavens A. F., Charlot S., 2007, *MNRAS*, 378, 1550  
 Pierini D., Gordon K. D., Witt A. N., Madsen G. J., 2004, *ApJ*, 617, 1022  
 Rocha M., Jonsson P., Primack J. R., Cox T. J., 2008, *MNRAS*, 383, 1281  
 Salim S. et al., 2005, *ApJ*, 619, L39  
 Salim S. et al., 2007, *ApJS*, 173, 267  
 Sánchez-Blázquez P. et al., 2006, *MNRAS*, 371, 703  
 Schaerer D., de Barros S., 2010, *A&A*, 515, A73  
 Scoville N. et al., 2007, *ApJS*, 172, 1  
 Springel V. et al., 2005, *Nat*, 435, 629  
 Thomas D., Maraston C., Johansson J., 2011, *MNRAS*, 412, 2183  
 Tremonti C. A. et al., 2004, *ApJ*, 613, 898  
 Tuffs R. J., Popescu C. C., Völk H. J., Kylafis N. D., Dopita M. A., 2004, *A&A*, 419, 821  
 Vazdekis A., Sánchez-Blázquez P., Falcón-Barroso J., Cenarro A. J., Beasley M. A., Cardiel N., Gorgas J., Peletier R. F., 2010, *MNRAS*, 404, 1639  
 Walcher C. J., Coelho P., Gallazzi A., Charlot S., 2009, *MNRAS*, 398, L44  
 Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., Croton D. J., Moore B., 2006, *MNRAS*, 372, 1161  
 Wild V., Kauffmann G., Heckman T., Charlot S., Lemson G., Brinchmann J., Reichard T., Pasquali A., 2007, *MNRAS*, 381, 543  
 Wild V., Budavári T., Blaizot J., Walcher C. J., Johansson P. H., Lemson G., De Lucia G., Charlot S., 2008, in Bailer-Jones C. A. L., ed., *AIP Conf. Ser. Vol. 1082, Classification and Discovery in Large Astronomical Surveys*. Am Inst. Phys., New York, p. 119  
 Wild V. et al., 2011a, *MNRAS*, 410, 1593  
 Wild V., Charlot S., Brinchmann J., Heckman T., Vince O., Pacifici C., Chevillard J., 2011b, *MNRAS*, 417, 1760  
 Wuyts S., Franx M., Cox T. J., Hernquist L., Hopkins P. F., Robertson B. E., van Dokkum P. G., 2009, *ApJ*, 696, 348  
 York D. G. et al., 2000, *AJ*, 120, 1579

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