

A high-resolution atlas of composite Sloan Digital Sky Survey galaxy spectra

László Dobos,^{1*} István Csabai,^{1,2} Ching-Wa Yip,² Tamás Budavári,² Vivienne Wild³ and Alexander S. Szalay²

¹*Department of Physics of Complex Systems, Eötvös Loránd University, 1117 Budapest, Hungary*

²*Department of Physics and Astronomy, The Johns Hopkins University, Baltimore, MD 21218, USA*

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

Accepted 2011 October 31. Received 2011 October 19; in original form 2011 August 24

ABSTRACT

In this work we present an atlas of composite spectra of galaxies based on the data of the Sloan Digital Sky Survey Data Release 7 (SDSS DR7). Galaxies are classified by colour, nuclear activity and star formation activity to calculate average spectra of high signal-to-noise ratio (S/N) and resolution ($S/N \simeq 132\text{--}4760$ at $\Delta\lambda = 1\text{ \AA}$), using an algorithm that is robust against outliers. Besides composite spectra, we also compute the first five principal components of the distributions in each galaxy class to characterize the nature of variations of individual spectra around the averages. The continua of the composite spectra are fitted with BC03 stellar population synthesis models to extend the wavelength coverage beyond the coverage of the SDSS spectrographs. Common derived parameters of the composites are also calculated: integrated colours in the most popular filter systems, line-strength measurements and continuum absorption indices (including Lick indices). These derived parameters are compared with the distributions of parameters of individual galaxies, and it is shown on many examples that the composites of the atlas cover much of the parameter space spanned by SDSS galaxies. By co-adding thousands of spectra, a total integration time of several months can be reached, which results in extremely low noise composites. The variations in redshift not only allow for extending the spectral coverage bluewards to the original wavelength limit of the SDSS spectrographs, but also make higher spectral resolution achievable. The composite spectrum atlas is available online at <http://www.vo.elte.hu/compositeatlas>.

Key words: methods: statistical – techniques: spectroscopic – atlases – surveys – galaxies: statistics.

1 INTRODUCTION

With the advance of large-scale spectroscopic surveys, spectrum stacking has become a widely used technique in extragalactic astrophysics to derive composite spectra of high signal-to-noise ratios (S/N values) in order to enable the analysis of spectral features that are washed away by noise in single object observations. Recent work have concentrated on averaging and analysing single classes of celestial objects, for example luminous red galaxies (Eisenstein et al. 2003), high-redshift galaxies (Abraham et al. 2004), galaxies in clusters (Dressler et al. 2004), star-forming galaxies with varying inclination (Yip et al. 2010) or quasars (Vanden Berk et al. 2001).

As the spectroscopic survey of the Sloan Digital Sky Survey (SDSS) Legacy Surveys has finished, it is time to summarize the collected information. In this work, we focus on galaxies of vari-

ous types, and present a comprehensive atlas of robustly averaged composite spectra and spectral principal components of galaxies classified by many criteria. We use the original observed spectra as they were reduced and published in SDSS Data Release (DR7; Abazajian et al. 2009).

1.1 Overview of the classification

Galaxies are automatically classified following a revised classification scheme based on the work of Lee et al. (2008). In their work, galaxies were classified by colour, nuclear activity, star formation activity and morphology.

We separate galaxies into three colour classes: red, green and blue, green referring to green-valley galaxies as defined by Strateva et al. (2001). Furthermore, galaxies are classified by star formation activity and nuclear activity into six classes: passive, measured H α emission, star forming, LINER and Seyfert galaxies; the sixth category is active galaxies with star formation signatures hereafter

*E-mail: dobos@complex.elte.hu

Table 1. Basic parameters of the composites by activity and colour class.

Class	λ_{\min} (Å)	λ_{\max} (Å)	Count	S/N
Red P	3350	8900	10431	1884
Red H α	3350	8950	61237	3996
Red SF	3350	8950	481	225
Red A+H II	3350	8950	726	329
Red L	3250	8850	901	483
Red S	3350	8950	577	347
Red all	3350	8950	86690	4760
Green P	3350	8900	2095	791
Green H α	3350	8950	19246	1858
Green SF	3550	8950	1533	482
Green A+H II	3350	8900	1067	431
Green L	3250	8800	327	233
Green S	3350	8950	942	442
Green all	3350	8950	26613	2148
Blue P	3350	8650	312	242
Blue H α	3350	8900	26550	1672
Blue SF	3550	8950	14304	1088
Blue A+H II	3350	8900	2725	541
Blue L	3250	8800	215	132
Blue S	3350	8900	1157	386
Blue all	3350	8900	40504	1856
All P	3350	8900	12838	1670
All H α	3350	8950	107033	3911
All SF	3550	8950	16318	1035
All A+H II	3350	8950	4518	610
All L	3250	8850	1443	470
All S	3350	8950	2676	569
All all	3350	8950	153807	4464
RED 1	3350	8950	17913	1777
RED 2	3350	8950	22934	2458
RED 3	3350	8950	46470	3939
RED 4	3350	8950	17546	2303
RED 5	3350	8950	2469	629
SF 1	3550	9000	301	192
SF 2	3550	9000	2862	650
SF 3	3550	9050	7073	1000
SF 4	3550	9000	4436	798
SF 5	3550	9000	1948	513

denoted as AGN + H II. Star-forming galaxies, LINERs and Seyferts are classified using the Baldwin–Phillips–Terlevich (BPT) diagnostic diagram (Baldwin, Phillips & Terlevich 1981) and the segregation curves determined by Kewley et al. (2001), Kauffmann et al. (2003a) and Kewley et al. (2006). We omit morphological classification of the galaxies for reasons detailed in Section 3, where we also describe the classification scheme in detail.

1.2 Robust average and principal components

To compute the average spectra, we use the robust statistical method of Budavári et al. (2009). This method not only gives the robust average and variance of a large, streaming¹ set of data vectors (the

¹ We apply the term ‘streaming’ to data that are too large to fit into the memory, so iterative loading from the background storage is necessary.

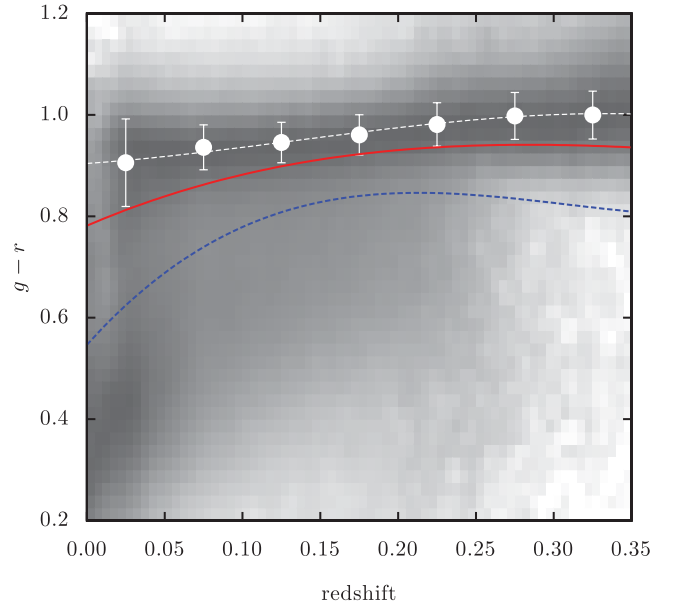


Figure 1. Definitions of the colour cuts. Density plot: logarithm of the distribution of the galaxy sample, normalized in each redshift bin. Red galaxies: above the solid red line; blue galaxies: below the dashed blue line; green-valley galaxies: between the two coloured lines. The data points and the dashed white curve represent the peak of the red sequence and our best fit to them as a function of redshift. See online edition for colour version of this plot.

spectra, in our case), but gives the first few principal components as well. Although we are primarily interested in the average spectra, the eigenspectra can help characterize the variations of objects within the subsets of the sample.

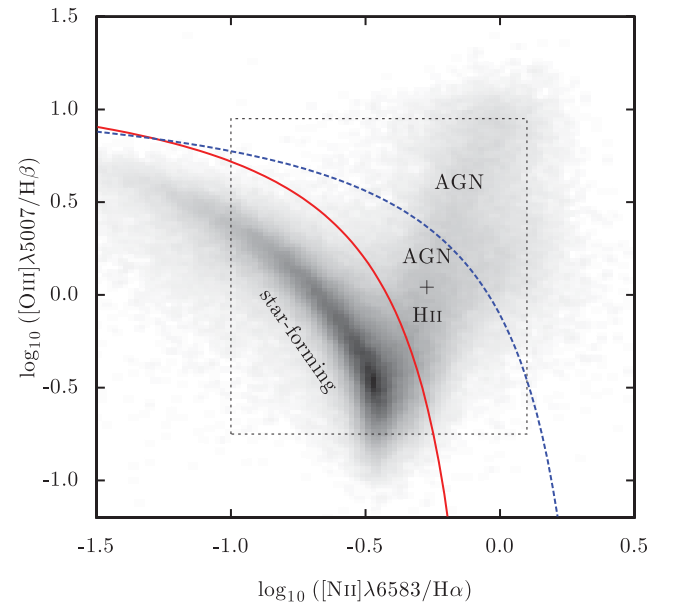


Figure 2. BPT diagram for our sample of SDSS galaxies. The density plot is the square root of the number density of SDSS galaxies with high enough line S/N measurements (see text). The solid red curve indicates the segregation line between star-forming galaxies and AGNs. The blue dashed curve is the extreme starburst line. The dashed rectangle marks the same area as displayed on Fig. 19. See text for references.

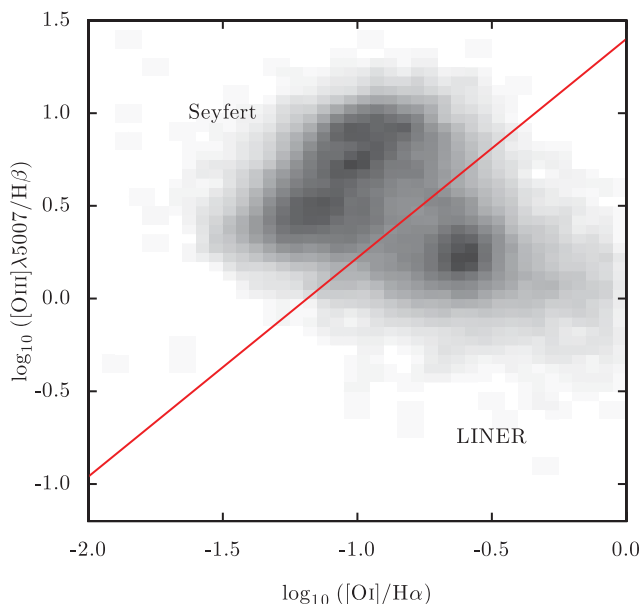


Figure 3. Seyfert/LINER separation. The density plot is the square root of the number density of our sample of SDSS galaxies classified as AGN (without star formation signature) using Fig. 2. The line separates the two sides of the bimodal distribution. See text for references.

Principal component analysis (PCA or Karhunen–Loève transformation) has been widely used in astronomical spectrum analysis in the last two decades. The basic idea of PCA is to derive an orthogonal basis from multidimensional data where the basis vectors

Table 2. Redshift and absolute magnitude ranges for each galaxy activity class.

	z_{\min}	z_{\max}	$M_{r,\min}$	$M_{r,\max}$
Passive	0.03	0.14	−20.5	−21.5
H α measured	0.03	0.14	−20.5	−21.5
Star forming	0.03	0.08	−19.2	−21.0
AGN + H II	0.03	0.14	−20.5	−21.5
LINER	0.04	0.17	−21.0	−22.0
Seyfert	0.03	0.14	−20.5	−21.5

point to the directions of maximum variance of the original data vectors. It was demonstrated that PCA can be successfully applied to high dimensional optical and ultraviolet galaxy spectra for classification purposes (Connolly et al. 1995). Real-life scenarios, when dealing with noisy and gappy data where certain spectral bins are masked out due to bad observations or other reasons, require more advanced techniques (Connolly & Szalay 1999). An application of this robust technique to SDSS spectra was performed for galaxies and quasars by Yip et al. (2004a,b). A novel robust algorithm was developed by Budavári et al. (2009) which combines the robustness and gap-handling capability of the former methods with the ability to handle extraordinarily large data sets by performing the calculations iteratively. This new method significantly reduces the computational and memory requirements of the calculations and enables handling higher dimensional data (thus processing higher resolution spectra).

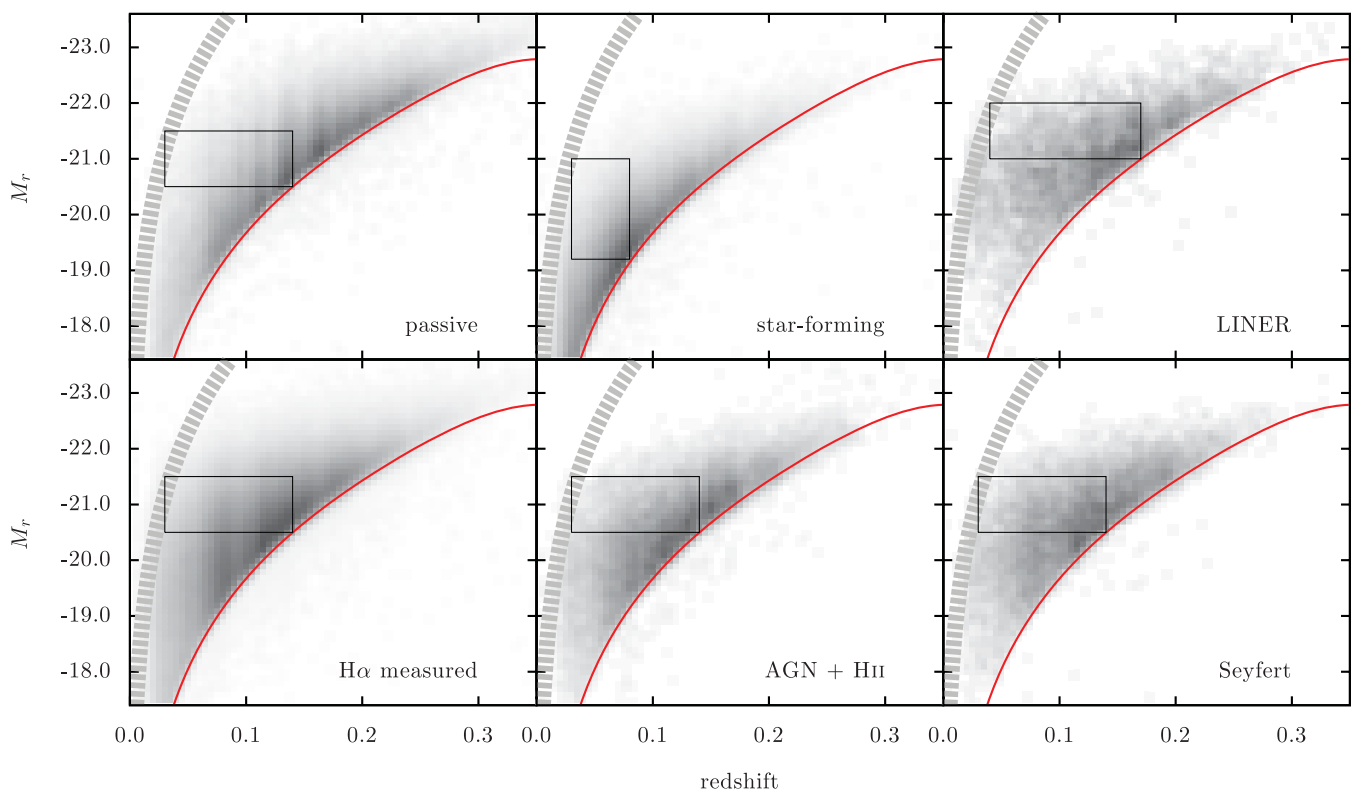


Figure 4. Definitions of redshift and r -band absolute magnitude M_r cuts for every galaxy activity class. Density plot: square root of the galaxy density within the activity class; solid red line: faint-end magnitude limit corresponding to 17.77 mag; grey line: estimated bright-end magnitude limit corresponding to about 13.5 mag. The cuts were manually set to contain as many galaxies as possible considering the magnitude limits of the SDSS spectro sample.

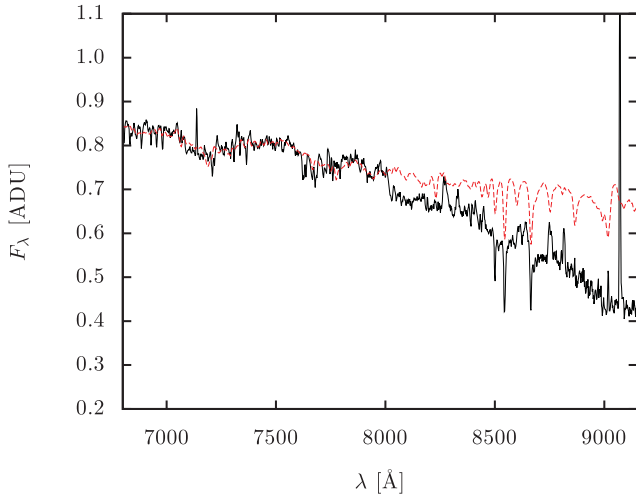


Figure 5. The effect of incomplete sampling. Using low-redshift spectra would allow extending the composites redwards, but bright-end incompleteness causes the averaged spectrum to turn down at longer wavelength, making the composite non-physical. Solid black line: composite spectrum calculated from an incomplete sample; dashed red line: best-fitting model. (Continuum fitting was done in the 3500–9000 Å range.)

1.3 Use cases of the spectrum atlas

We would like to emphasize two use cases of this spectrum atlas. One of the keys to the success of template-based photometric redshift estimators is the good set of spectral templates used to fit the broad-band spectral energy distributions of observed galaxies. Spectral templates from models are often used, but they can miss some important features of real spectra. Also, models have to be highly consistent with the observations in terms of integrated colours. As an alternative to models, individual empirical spectra are used, but it is not easy to get a high S/N set of spectra that represents all types of galaxies. Our composites can provide a high S/N, reliable and representative template set for photo- z applications.

The other important use case of our library is generating mock catalogues. We expand the individual galaxy spectra on the eigenbases of the various galaxy classes to determine the so-called eigencoeficients. The best-fitting Gaussian curves to the distribution of these eigencoeficients are determined and the parameters of the Gaussian fits (m , σ) are published, along with the eigenspectra. By using these parameters of the distributions, linear combinations of the average composite spectra and eigenspectra can be easily constructed to further extend the parameter space coverage of the atlas.

1.4 Structure of the paper

In Section 2, we present the data we used for the analysis. In Section 3, we describe our classification scheme in detail. We present the details of the averaging method and the composite spectra in Section 4, where we also discuss the performance and robustness of the averaging algorithm. In Section 5, we fit stellar population synthesis models to the composites and present colours, emission line measurements and Lick indices of them. By comparing the derived physical parameters with the parameter distributions of individual galaxies, we show that the composites of the atlas are representative to the whole SDSS galaxy ensemble. We summarize our work in Section 6.

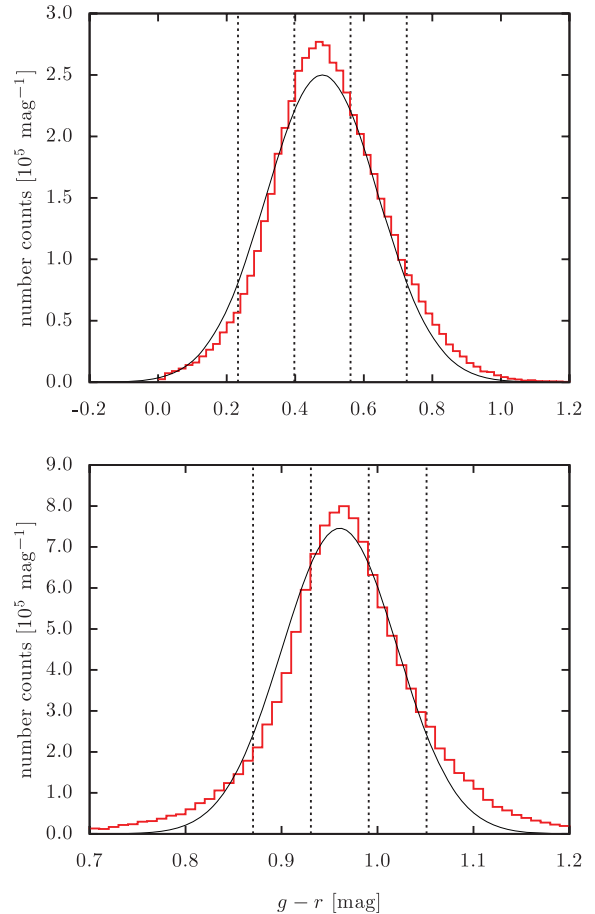


Figure 6. $g - r$ colour distribution of star-forming galaxies (top panel) and passive galaxies (bottom panel). The red histograms show the measured distributions, while the black curves indicate the fitted Gaussians. The vertical dashed lines mark the colour cuts we defined at 0.5σ and 1.5σ distances from the mean.

2 THE DATA

We analyse galaxy spectra from SDSS DR7 (Abazajian et al. 2009) as reduced by the `spectro` pipeline (Stoughton et al. 2002), version 2.6. All spectra (including duplicate observations) are selected for analysis which were identified as galaxies by the `spectro` pipeline and have more than 3800 well-measured spectral bins.

Photometric data are taken from the SDSS DR7 ‘best’ data set as measured by the SDSS photo image reduction pipeline (Stoughton et al. 2002).

K -corrections and absolute magnitudes are taken from the New York University Value-Added Galaxy Catalogue (Blanton et al. 2005). K -corrections are calculated for $z = 0.1$, absolute magnitudes are based on the concordance Λ cold dark matter cosmological model with parameters of $H_0 = 100 \text{ km s}^{-1}$, $\Omega_M = 0.3$ and $\Omega_\Lambda = 0.7$.

For line measurements of observed spectra, we use the MPA/JHU Value-Added Galaxy Catalogue (Kauffmann et al. 2003b; Brinchmann et al. 2004; Tremonti et al. 2004).

Where not otherwise noted, galaxy luminosities are quoted in Petrosian magnitudes. Wavelength values are expressed in vacuum throughout the paper.

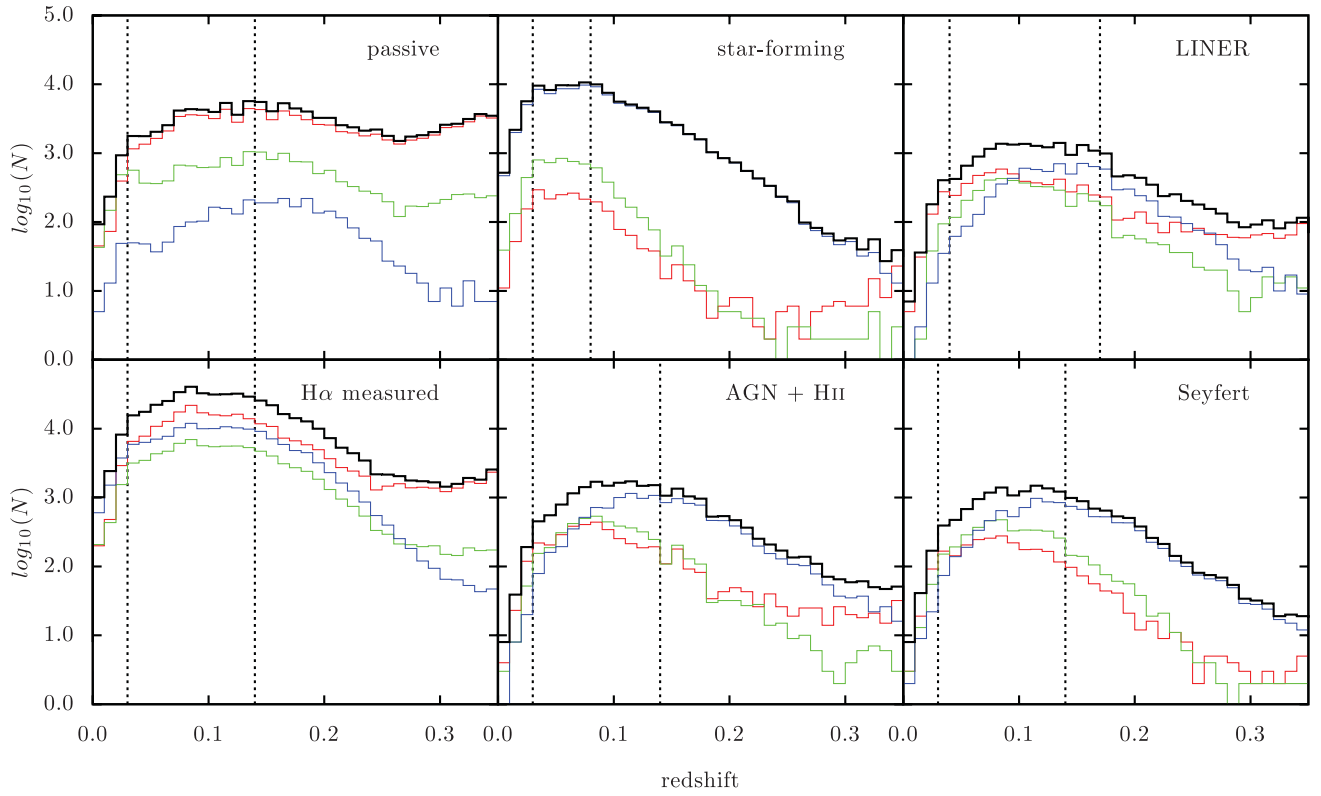


Figure 7. Redshift distribution of the galaxies in each activity class *before* the absolute magnitude cuts. Colours represent the colour classes: red – red galaxies; green – green-valley galaxies; blue – blue galaxies; and black – all galaxies regardless of their colours. The dashed vertical lines represent the redshift cuts, which are different for every activity class.

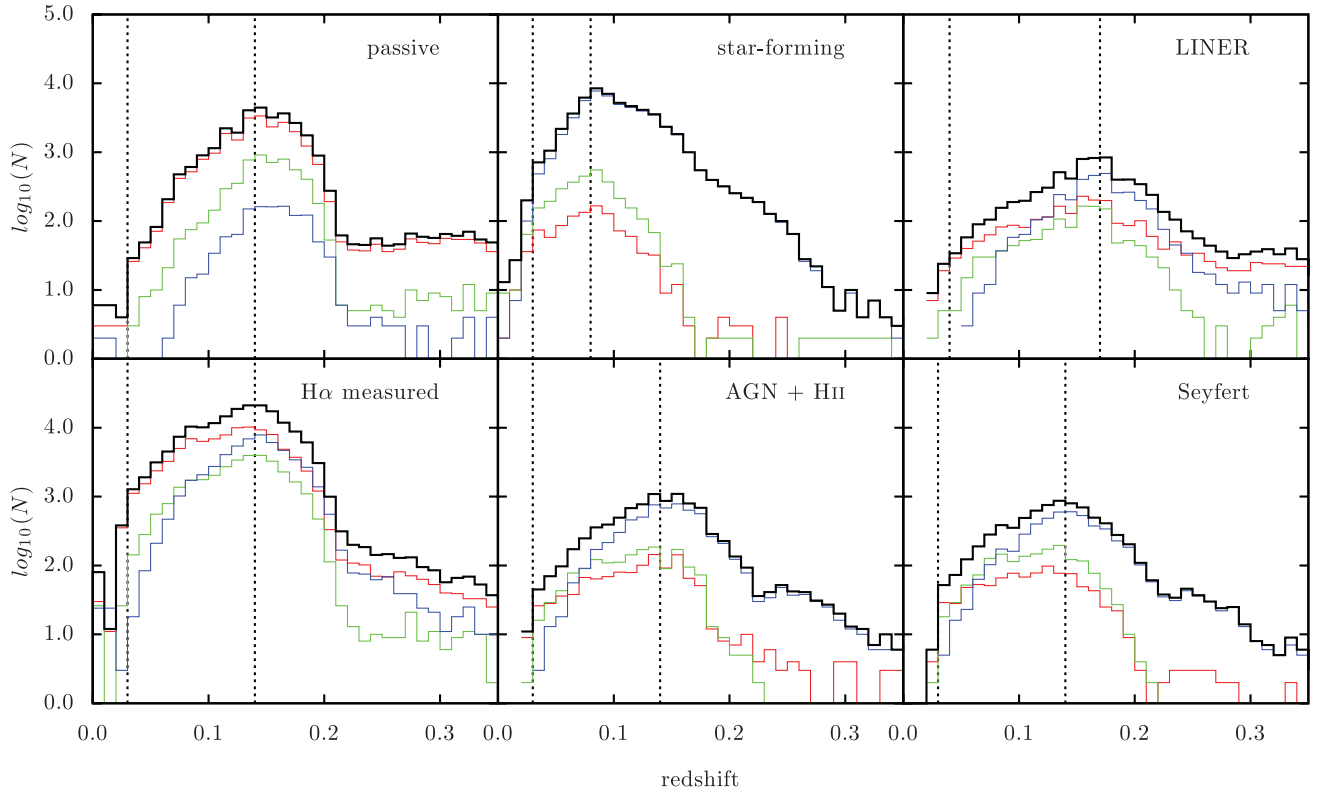


Figure 8. Redshift distribution of the galaxies in each activity class *after* the absolute magnitude cuts. Colours represent the colour classes: red – red galaxies; green – green-valley galaxies; blue – blue galaxies; and black – all galaxies regardless of their colours. The dashed vertical lines represent the redshift cuts, which are different for each activity class.

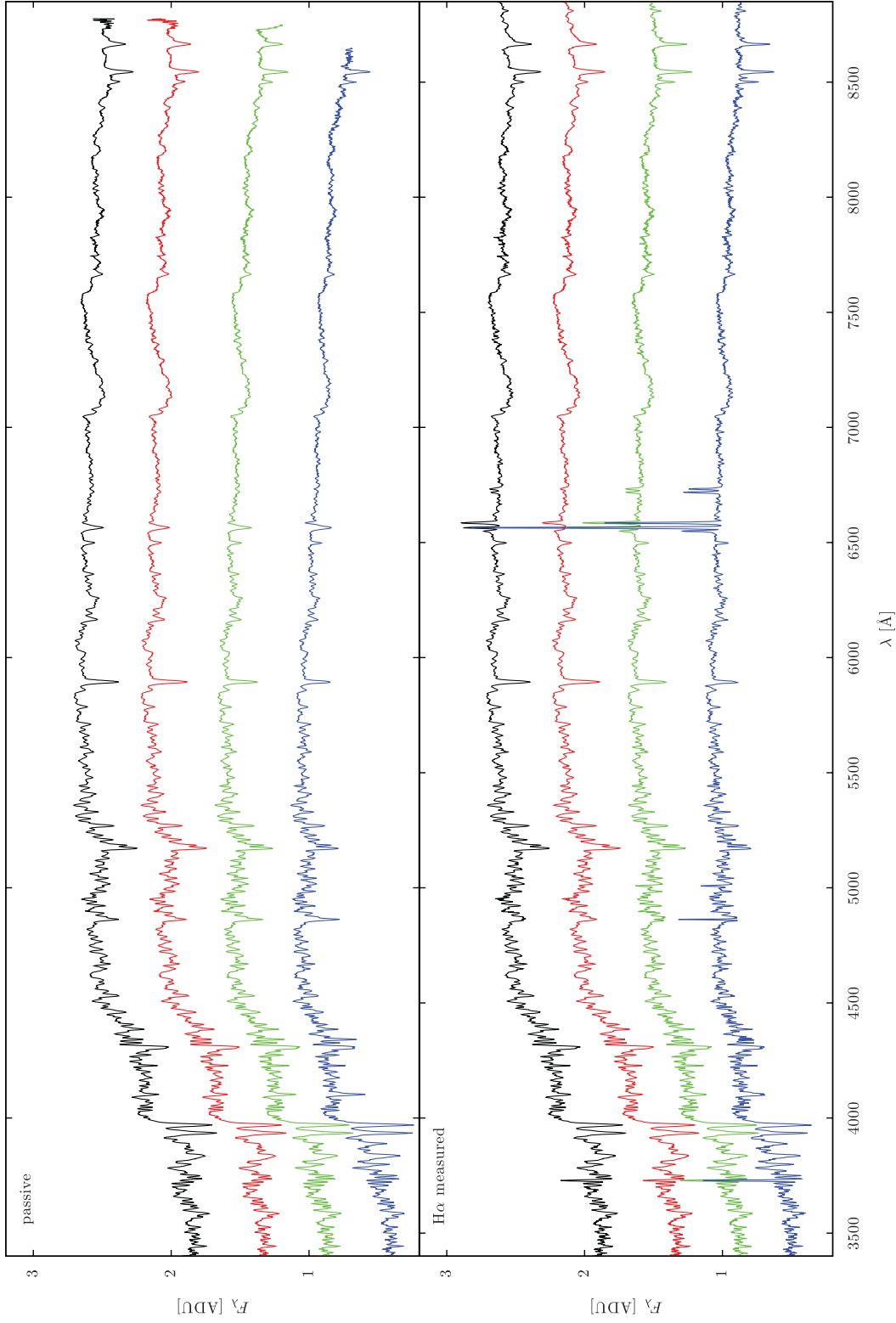


Figure 9. Composite spectra of the different activity and colour classes. Activity class is indicated in the top-left corner of each panel, the spectra of the different colour classes are offset by 0.5 along the y-axis for clarity and indicated by colours from top to bottom (black: all colour, red, green and blue). See the online edition for a colour version of this plot.

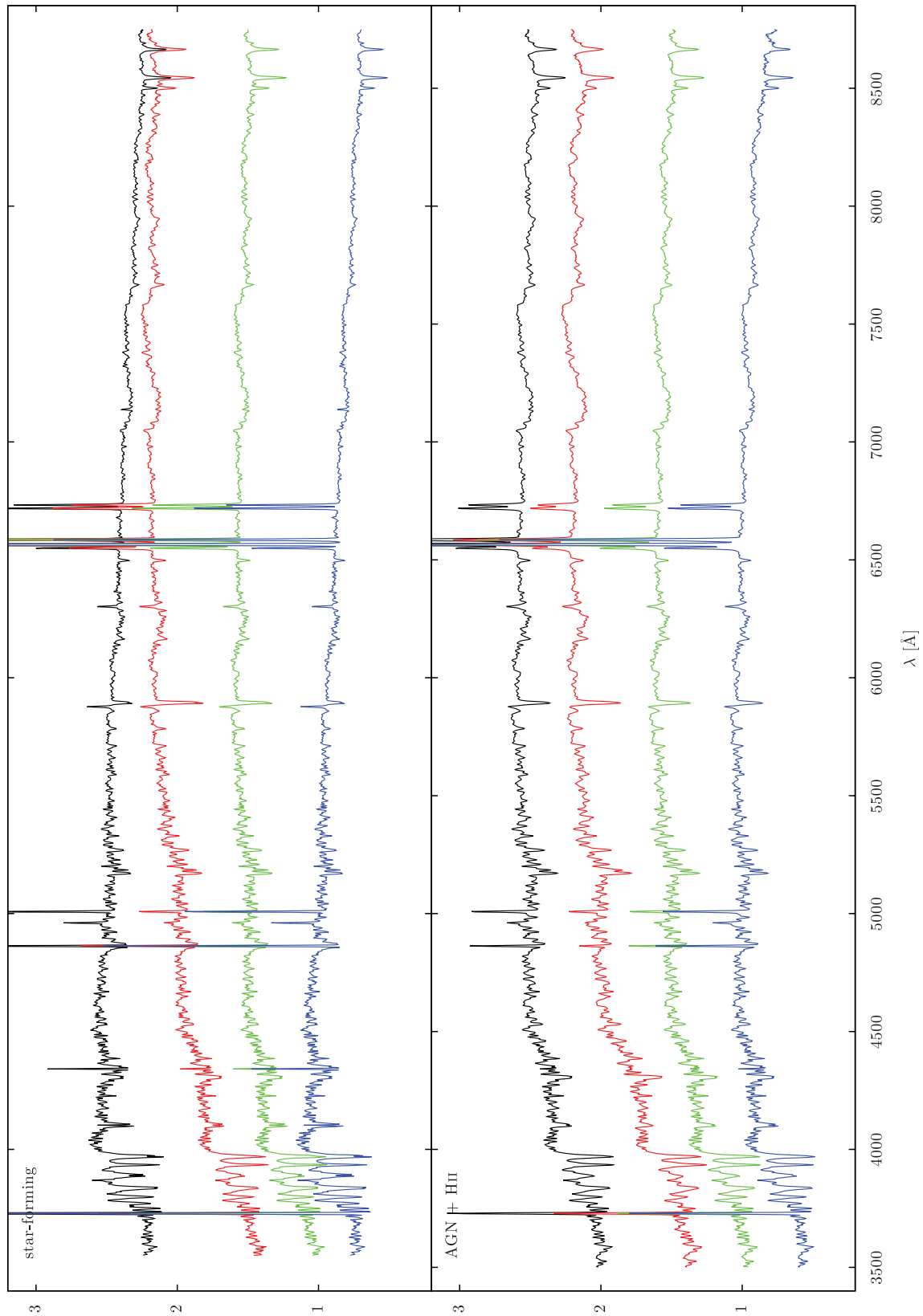


Figure 10. See caption of Fig. 9 for detailed description.

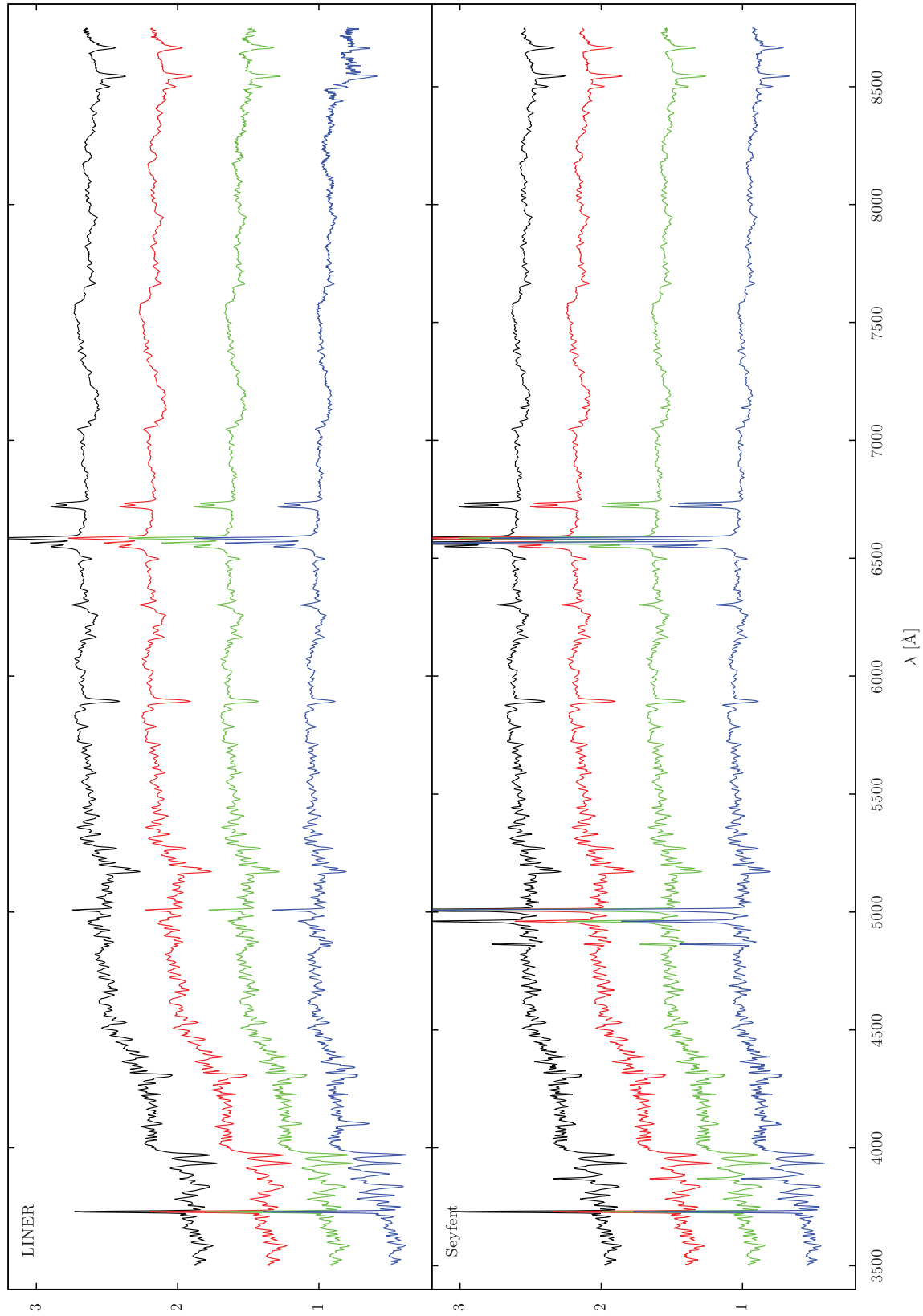


Figure 11. See caption of Fig. 9 for detailed description.

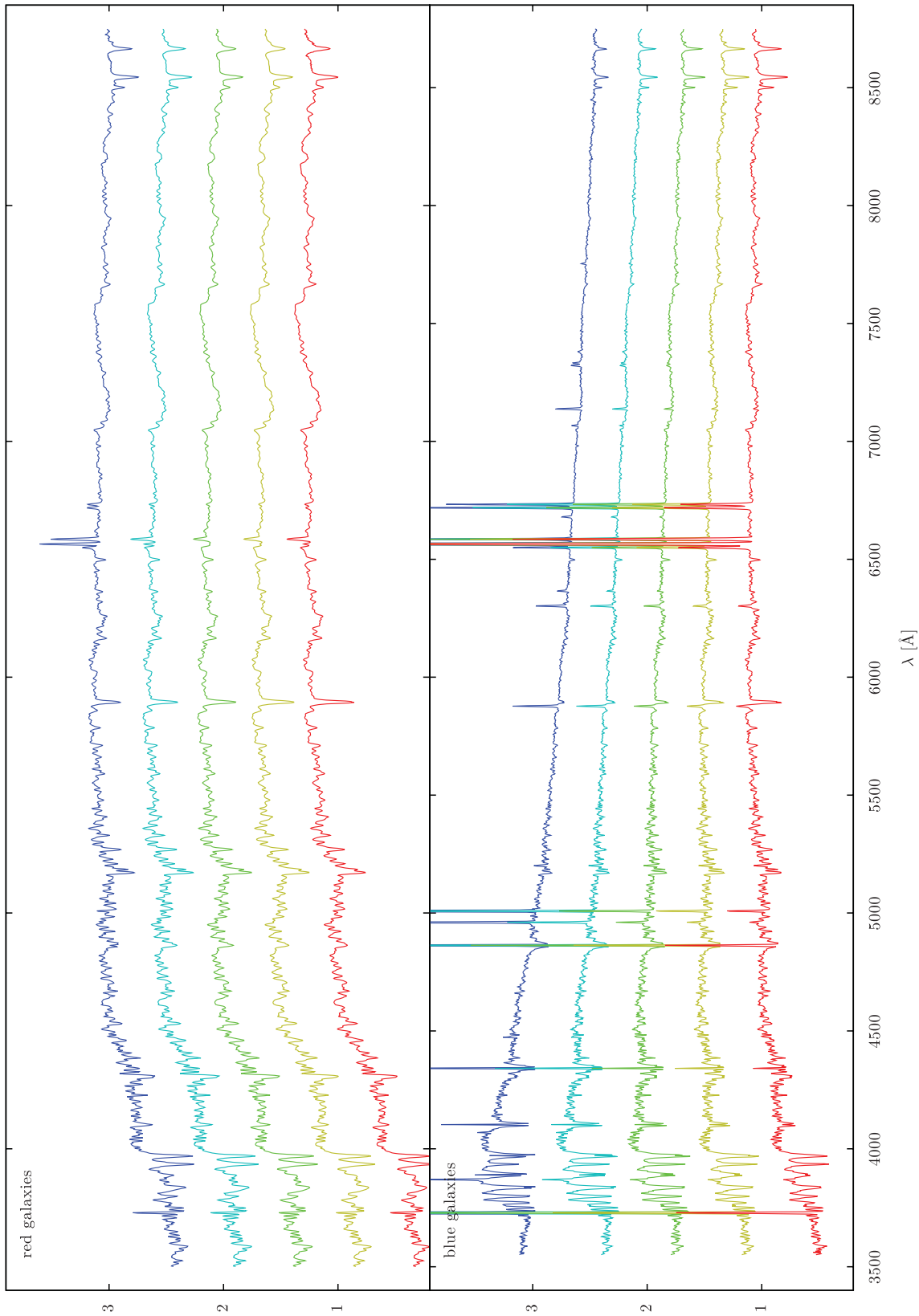


Figure 12. Composite spectra of the refined colour classes as described in Section 3.4. The curves are colour coded from blue (top) to red (bottom) based on the $g - r$ colour of the galaxies. See the online edition for a colour version of this plot.

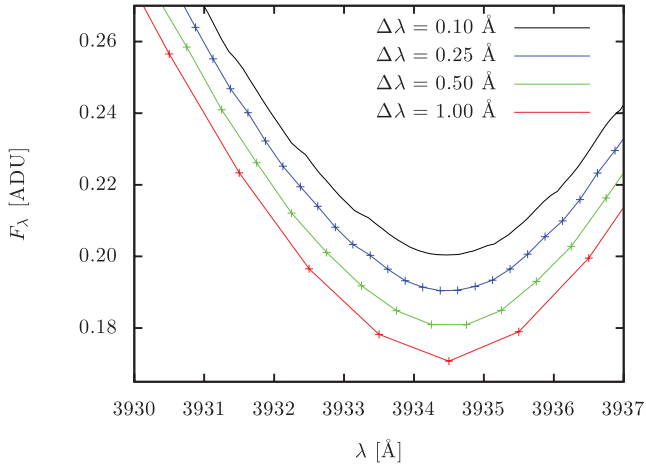


Figure 13. The K line of a galaxy composite spectrum calculated at resolutions $\Delta\lambda = 1, 0.5, 0.25$ and 0.1 Å (from bottom to top). Higher resolution curves are smoother compared to the $\Delta\lambda = 1 \text{ Å}$ curve.

3 CLASSIFICATION SCHEME

We base our galaxy classification scheme on the work of Lee et al. (2008), but define several additional classes as well. In their scheme, galaxies were classified by morphology (early-type and late-type), colour (red and blue) and nuclear activity (passive, star forming, LINER and Seyfert). We extend this scheme by classifying galaxies by the strength of the H α emission (completely passive galaxies and those showing some nebular emission but not enough to classify them using the BPT diagram). We also introduce a transitional class between star-forming galaxies and active galactic nuclei (AGNs) based on the works by Kewley et al. (2001, 2006) and Kauffmann et al. (2003a). Kauffmann et al. (2003a) termed this latter class ‘composite’; we will refer to these galaxies as ‘AGN + H II’ to avoid confusion with *composite spectra*. The colour classification is extended to treat green-valley galaxies separately. Volume-limited cuts are imposed on all subsamples as described in Section 3.3. For the most numerous galaxy classes, we also define further classes based on colour cuts in Section 3.4, and term these classes ‘refined star forming’ or ‘SF’ and ‘refined red’ or ‘RED’. The number of

galaxies used to calculate the composites, the wavelength coverage and the estimated S/N are summarized in Table 1.

In contrast to Lee et al. (2008), we omit morphological classification of the galaxies for the following reasons. Morphological classification relies on the radial colour gradient of galaxies which was not measured perfectly by the SDSS photo pipeline for the whole spectro sample. Thus, galaxy classes defined by morphology would contain about 60 per cent fewer galaxies: only those with good radial colour gradient measurements. Using morphological information would reduce the number of the classified galaxies to a number that would not allow reaching the desired low noise level and resolution. Also, the size- and magnitude-dependent incompleteness of the morphological classification would introduce unwanted bias into the data set.

We further explain the classification criteria in detail in the following subsections.

3.1 Classification by colour

We organize the galaxies into three colour classes based on their redshifts and $g - r$ colours. The three colour classes are the following.

- (i) Red galaxies: having primarily old stellar populations or galaxies with significant intrinsic extinction.
- (ii) Green galaxies: those belonging to the green valley, thought to form the transitional region between active star formation and passivity (Strateva et al. 2001; Baldry et al. 2004; Bell et al. 2004).
- (iii) Blue galaxies: mostly actively star-forming galaxies.

Fig. 1 displays the colour classes we define. Galaxies above the solid red curve are classified as ‘red’, galaxies below the blue curve are classified as ‘blue’ and green galaxies lie in between the two. These curves were determined by fitting Gaussians to the distribution of galaxies by the $g - r$ colour in bins of $\Delta z = 0.05$ in the range $0.0 \leq z \leq 0.35$. The dashed white line is the centre of the red sequence, the solid red line is the 1.2σ envelope of it and the blue line is at 3.5σ from the red sequence. To construct this plot, all galaxies were used independently of their apparent or absolute magnitudes.

In Fig. 1, colour curves are defined by the σ parameter of the fitted Gaussians as explained above. Since the photometric error

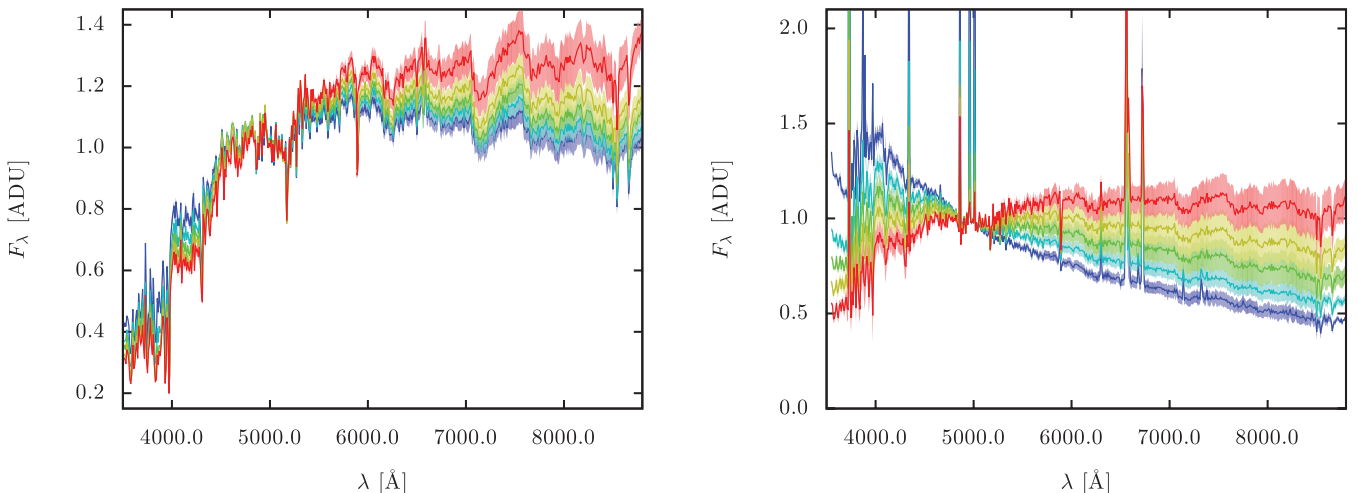


Figure 14. Variance in red (left-hand panel) and blue star-forming (right-hand panel) galaxies. The solid curves show the composites of galaxies classified according to the refined colour classes (redder to bluer). The shaded areas represent the 0.5σ (red galaxies) and 1σ (blue galaxies) variations about the average. See the online edition for a colour version of this plot.

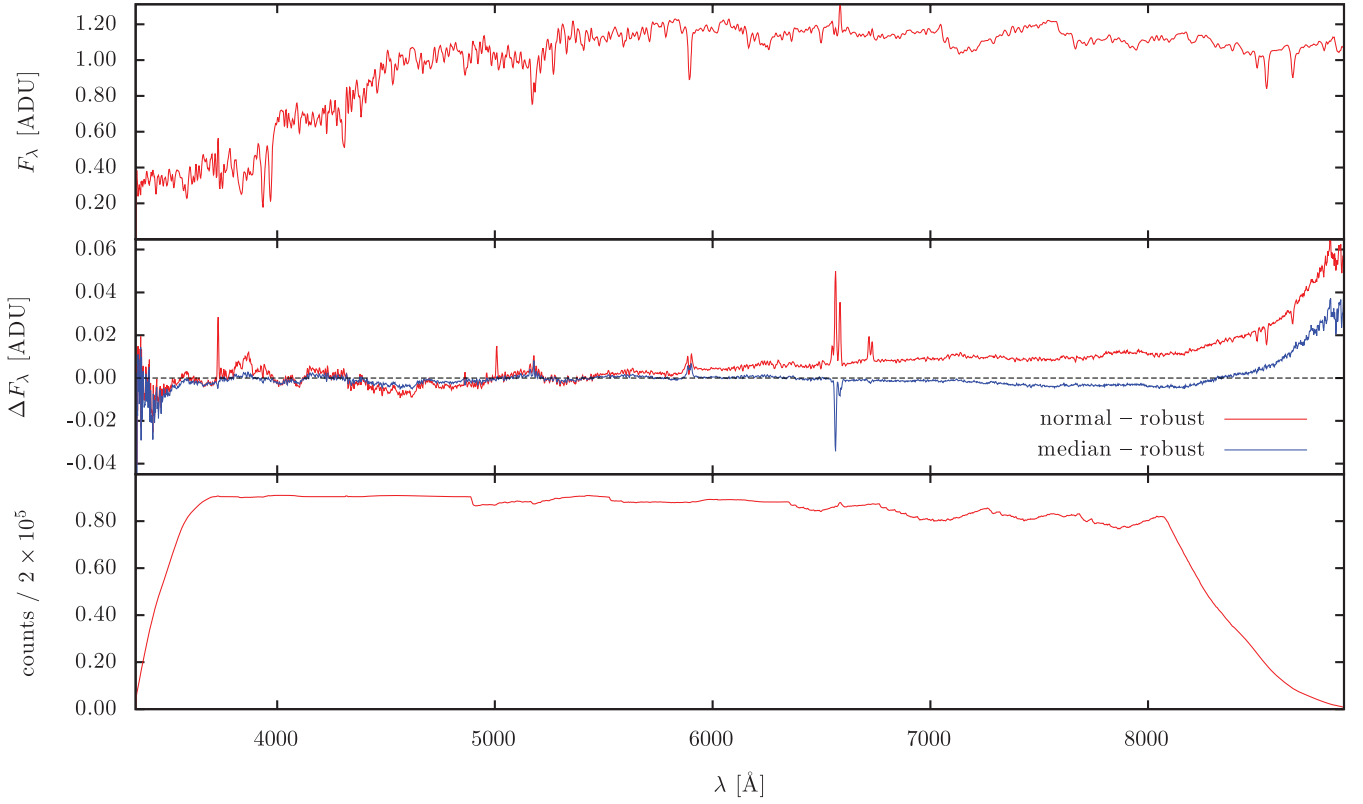


Figure 15. Comparison of the various averaging methods for red galaxies with measured H α . We plot the spectrum in the top panel for reference. The middle panel shows the difference between the robust average and the normal average and the median. The bottom panel shows the number of spectra used to determine the averages of the wavelength bins.

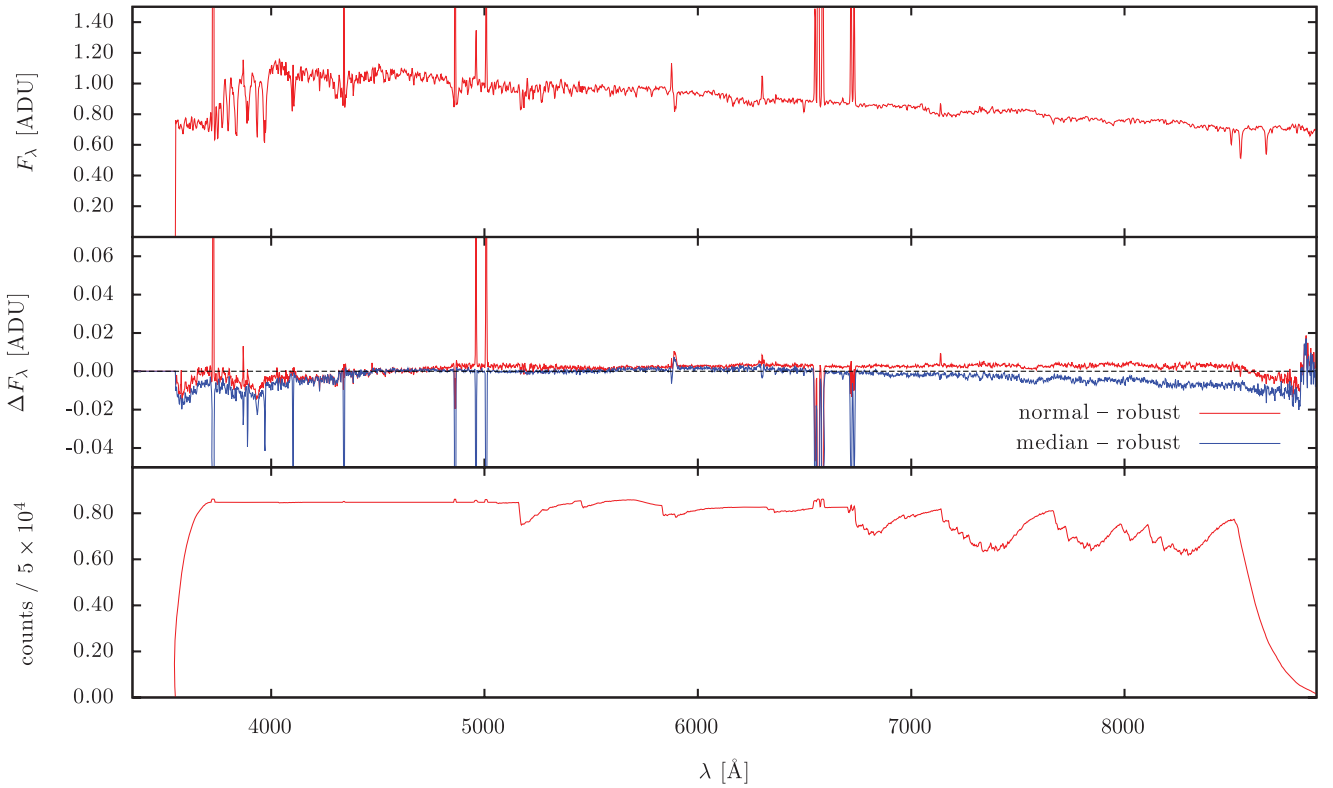


Figure 16. Same as Fig. 15, but for star-forming blue galaxies.

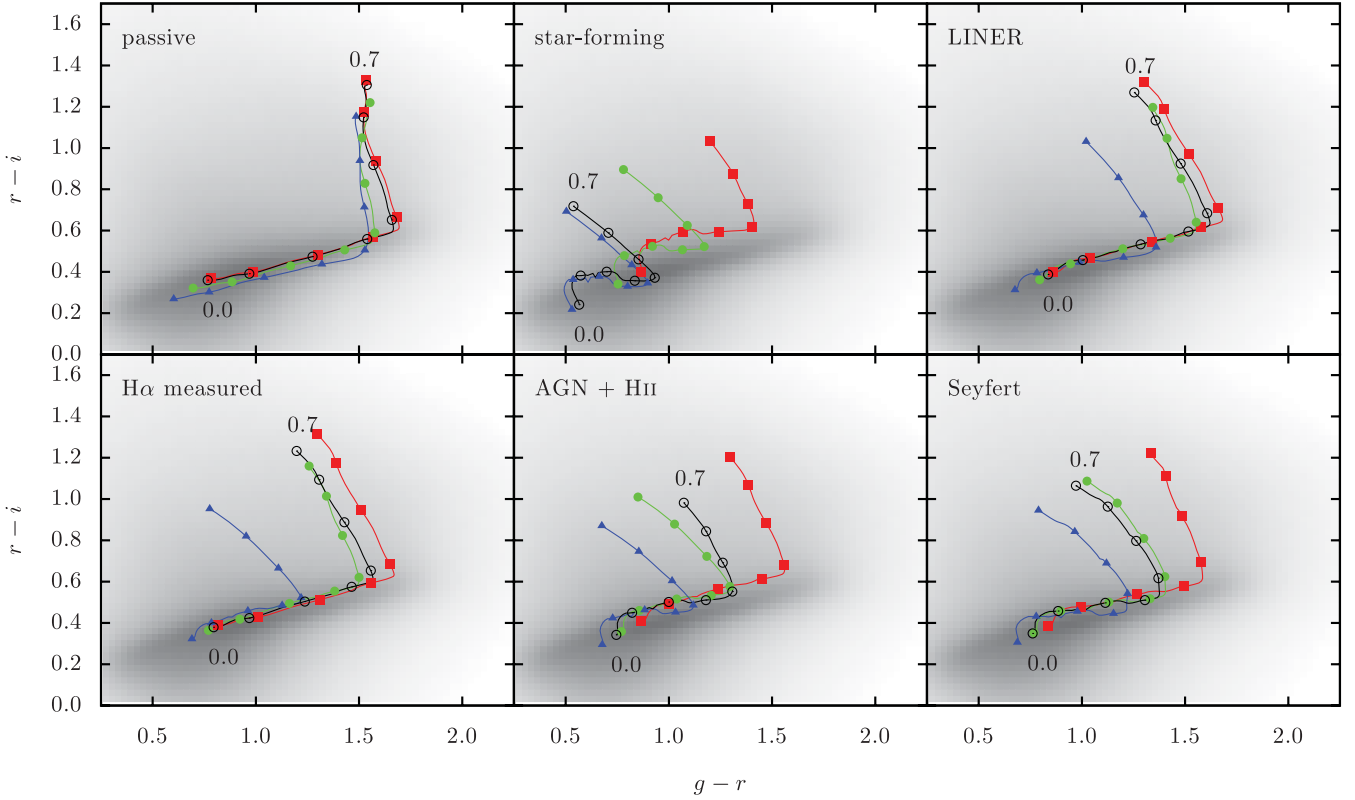


Figure 17. Colour-colour diagram of the composites for the SDSS $g-r$ and $r-i$ colour indices between redshifts in the range $0 \leq z \leq 0.7$. The coloured curves represent the composite spectra of the different colour classes we defined in Section 3.1. The black curve represents the composite spectra calculated from galaxies with any colour. Dots along the curves mark redshifts $z = 0.0, 0.1$, etc. The grey-scale map in the background shows the distribution of the extinction-corrected photometric colours calculated from the fibre magnitudes of SDSS photo galaxies, which is significantly deeper than the spectro set.

increases with increasing redshift, it makes the scatter in the red sequence higher. This causes the colour curves to turn over around $z \simeq 0.2$ and be more restrictive for green-valley and blue galaxies, allowing more variance in the colour of red galaxies. Because of the relatively low number of observed green-valley and blue galaxies at these redshifts, this would have no effect on the average composites, but more high-redshift red galaxies could be used for the analysis. In the present analysis, we set the redshift limits of the subsample volumes below $z \leq 0.17$.

3.2 Classification by activity

Our classification scheme by star formation and nuclear activity is based on the works by Kewley et al. (2001) and Kauffmann et al. (2003a). LINER/Seyfert separation follows Kewley et al. (2006). We define the following six classes based on activity.

- (i) Passive galaxies: no measurable $H\alpha$ emission.
- (ii) $H\alpha$ measured: hydrogen emission is measured, but other lines are not strong enough for classification.
- (iii) Star-forming galaxies: with clear star formation signatures but no active nucleus.
- (iv) AGN + $H\text{II}$ galaxies: with AGN and extreme starburst signatures at the same time.
- (v) LINER galaxies: pure LINER signatures (Heckman 1980).
- (vi) Seyfert galaxies: pure Seyfert signatures.

In Fig. 2, we plot the BPT diagram of the galaxies in our sample. The segregation line (solid red curve) between star-forming galaxies

(below the curve) and AGN (above the curve) is

$$\log_{10} \frac{[\text{O III}]}{H\beta} = 0.61 \left[\log_{10} \frac{[\text{N II}]}{H\alpha} - 0.05 \right]^{-1} + 1.3.$$

The limit of extreme starburst (blue dashed line) is determined as

$$\log_{10} \frac{[\text{O III}]}{H\beta} < 0.61 \left[\log_{10} \frac{[\text{N II}]}{H\alpha} - 0.47 \right]^{-1} + 1.19.$$

The latter line separates pure AGN (above the curve) from AGN showing star formation signatures (between the two curves).

Fig. 3 shows the distribution of galaxies classified as AGN (without star formation signatures). We apply the following cut to distinguish LINERs (below the line) from Seyfert galaxies (above the line):

$$\log_{10} \frac{[\text{O III}]}{H\beta} = 1.18 \log_{10} \frac{[\text{O I}]}{H\alpha} + 1.4.$$

Only galaxies with $H\alpha$ measured at $S/N \geq 3$ are classified as star forming, while only those with all $H\alpha$, $H\beta$, $[\text{O I}]$, $[\text{O III}]$ and $[\text{N II}]$ measured at $S/N \geq 3$ are classified as AGN (either AGN + $H\text{II}$, LINER or Seyfert). For those galaxies that do not have strong enough emission lines to make it on to the BPT diagram, we define two further classes: completely passive with no detectable emission lines and passive with measurable $H\alpha$.

Although their number is negligible, we excluded galaxies with extreme values of line ratios, i.e those that are outside the limits of the plot in Fig. 2.

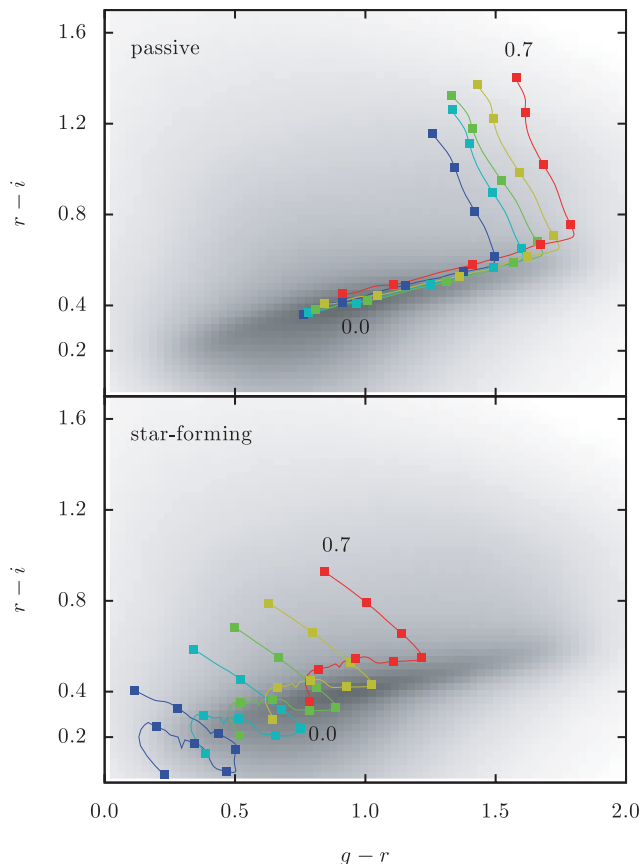


Figure 18. Colour-colour diagrams of the composites for the SDSS $g-r$ and $r-i$ colour indices between redshifts in the range $0 \leq z \leq 0.7$. The coloured curves represent the composite spectrum of the different refined colour classes (top panel: red galaxies; bottom panel: star-forming galaxies), as defined in Section 3.4. Dots along the curves mark redshifts $z = 0.0, 0.1$, etc. The grey-scale map in the background shows the distribution of the photometric colours of SDSS spectro galaxies.

3.3 Volume-limited samples

In order to avoid the Malmquist bias, we define redshift and absolute magnitude ranges for every galaxy class, trying to maximize the number counts of the volume-limited samples.

Although setting restrictive constraints on the redshift and absolute magnitude reduces the wavelength coverage of the composites, it is necessary to get physically meaningful results. The spectroscopic targeting algorithms of SDSS (Eisenstein et al. 2001; Stoughton et al. 2002; Strauss et al. 2002) set the sharp faint-end r -band magnitude limit at 17.7 mag for the main galaxy sample. On the other hand, bright-end incompleteness is much less well controlled. One of the main reasons for this incompleteness is that nearby galaxies are too big in apparent diameter compared to the spectroscopic fibre diameter. We estimate the bright-end apparent r -band magnitude limit to be around 13.5 mag when determining the volumes.

We set the volumes as plotted in Fig. 4 and summarized in Table 2.

3.3.1 Effect of incomplete sampling

We note that the continua of average composite spectra are physical in the sense that they represent a non-negative linear combination of

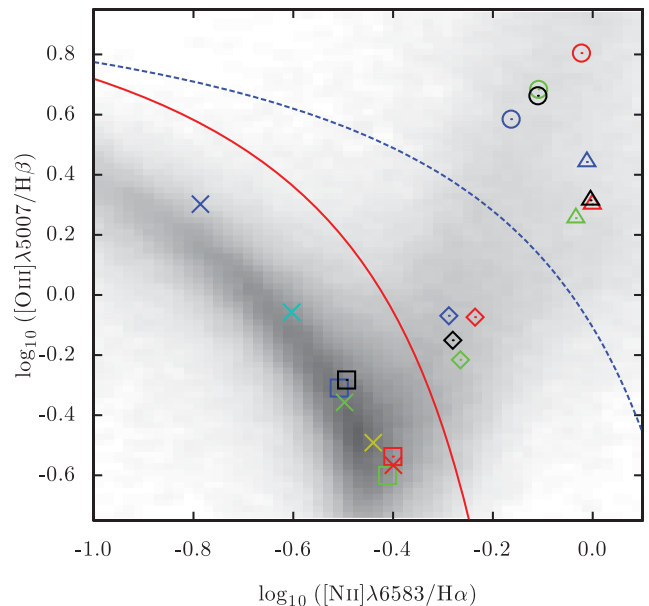


Figure 19. BPT diagram of the composite spectra. The density plot is the logarithm of the number density of all SDSS galaxies. The solid red curve indicates the segregation line between star-forming galaxies and AGNs. The blue dashed curve is the extreme starburst line. The loci of the composites are indicated with points: rectangles – star-forming galaxies; diamonds – AGN+H II galaxies; triangles – LINERs and circles – Seyferts. The different colours denote the different colour classes; black points mark the loci of composites of all galaxies (regardless of their colours) in a given activity class. Colour crosses mark the loci of the star-forming galaxies, according to the refined classification. See the online edition for a colour version of this plot.

real stellar populations. This makes it meaningful to fit composites with theoretical population synthesis models. However, incomplete sampling of galaxies for composite calculation has the effect of making the composites non-physical.

While we consider spectral evolution negligible in the small redshift intervals we choose, care must be taken about the dependence of spectra on the absolute magnitude of the objects. For instance, choosing a lower redshift cut might enable extending the wavelength coverage of the composites redwards, but bright-end incompleteness would lead to the dominance of lower luminosity galaxies which (in case of blue galaxies, for example) tend to be bluer. This would result in a downturn of the composite spectrum as it can be seen in Fig. 5.

3.4 Refined classification of blue and red galaxies

Template-based photometric redshift (photo- z) estimation algorithms (for a recent review, see Hildebrandt et al. (2010), and references therein) rely on good-quality empirical galaxy spectra or theoretical models to fit the broad-band observations with, though certain techniques exist to improve the reliability of the models (Budavári et al. 2000; Blanton et al. 2003). Most often, the problem with using theoretical models for photo- z is the mismatch between the colours calculated from the models and the colours of the observed galaxies. Also, photo- z techniques are usually based on a small number of templates in order to reduce computational complexity and to avoid spurious results due to fitting too many free parameters of the not-so-reliable models.

Table 3. Values of the emission-line equivalent widths (in Å) for every galaxy class, continued in Table 4.

	λ (Å)	Red P	Red H α	Red SF	Red A+H II	Red L	Red S	Red all	Green P	Green H α	Green SF	Green A+H II	Green L	Green S	Green all
O II	3727.09	–	4.163	14.122	7.523	7.337	8.480	5.202	–	2.369	13.465	7.680	9.742	9.010	2.847
O II	3729.88	–	–	4.857	8.665	7.729	9.871	–	–	1.467	17.105	8.734	10.975	10.246	1.723
H θ	3798.98	–	–	–	–	–	–	–	–	–	–	–	–	–	–
H η	3836.47	–	–	–	–	–	–	–	–	–	–	–	–	–	–
He I	3889.00	–	–	1.390	1.111	–	1.261	–	–	–	1.382	–	–	1.197	–
S II	4072.30	–	–	–	–	–	–	–	–	–	–	–	–	–	–
H δ	4102.89	–	–	2.068	1.235	–	1.131	–	–	–	2.029	1.252	–	–	–
H γ	4341.68	–	1.433	3.889	2.464	–	2.235	1.868	–	–	3.766	1.888	–	1.730	–
O III	4364.44	–	–	–	–	–	–	–	–	–	–	–	–	–	–
H β	4862.68	–	3.091	9.453	5.607	1.471	4.784	4.087	–	–	9.060	4.745	1.159	3.401	–
O III	4932.60	–	–	–	–	–	–	–	–	–	–	–	–	–	–
O III	4960.30	–	–	1.448	1.121	–	5.680	–	–	–	1.308	–	–	5.063	–
O III	5008.24	–	1.086	5.125	4.198	2.949	18.697	1.998	–	–	4.692	3.252	2.622	16.649	–
O I	6302.05	–	–	1.272	1.228	1.580	1.611	–	–	–	1.247	1.238	1.490	1.496	–
O I	6365.54	–	–	–	–	–	–	–	–	–	–	–	–	–	–
N I	6529.03	–	–	–	–	–	–	–	–	–	–	–	–	–	–
N II	6549.86	–	–	4.204	5.429	5.731	5.875	3.902	–	–	4.150	6.879	3.603	7.439	1.658
H α	6564.61	–	14.908	41.231	26.957	10.188	22.476	19.225	–	3.751	39.508	21.372	6.435	16.527	4.380
N II	6585.27	–	5.677	12.908	13.600	9.450	15.528	7.629	–	2.137	12.745	11.214	6.453	12.834	2.436
S II	6718.29	–	2.053	7.254	4.735	3.932	4.790	2.853	–	–	6.906	4.380	3.617	4.245	1.100
S II	6732.67	–	1.478	5.205	3.748	3.351	4.242	2.088	–	–	4.973	3.682	3.316	3.932	–

Table 4. Values of the emission-line equivalent widths (in Å) for every galaxy class, continued from Table 3.

	λ (Å)	Blue P	Blue H α	Blue SF	Blue A+H II	Blue L	Blue S	Blue all	All P	All H α	All SF	All A+H II	All L	All S	All all
O II	3727.09	–	4.163	14.122	7.523	7.337	8.480	5.202	–	2.369	13.465	7.680	9.742	9.010	2.847
O II	3729.88	–	–	4.857	8.665	7.729	9.871	–	–	1.467	17.105	8.734	10.975	10.246	1.723
H θ	3798.98	–	–	–	–	–	–	–	–	–	–	–	–	–	–
H η	3836.47	–	–	–	–	–	–	–	–	–	–	–	–	–	–
He I	3889.00	–	–	1.390	1.111	–	1.261	–	–	–	1.382	–	–	1.197	–
S II	4072.30	–	–	–	–	–	–	–	–	–	–	–	–	–	–
H δ	4102.89	–	–	2.068	1.235	–	1.131	–	–	–	2.029	1.252	–	–	–
H γ	4341.68	–	1.433	3.889	2.464	–	2.235	1.868	–	–	3.766	1.888	–	1.730	–
O III	4364.44	–	–	–	–	–	–	–	–	–	–	–	–	–	–
H β	4862.68	–	3.091	9.453	5.607	1.471	4.784	4.087	–	–	9.060	4.745	1.159	3.401	–
O III	4932.60	–	–	–	–	–	–	–	–	–	–	–	–	–	–
O III	4960.30	–	–	1.448	1.121	–	5.680	–	–	–	1.308	–	–	5.063	–
O III	5008.24	–	1.086	5.125	4.198	2.949	18.697	1.998	–	–	4.692	3.252	2.622	16.649	–
O I	6302.05	–	–	1.272	1.228	1.580	1.611	–	–	–	1.247	1.238	1.490	1.496	–
O I	6365.54	–	–	–	–	–	–	–	–	–	–	–	–	–	–
N I	6529.03	–	–	–	–	–	–	–	–	–	–	–	–	–	–
N II	6549.86	–	–	4.204	5.429	5.731	5.875	3.902	–	–	4.150	6.879	3.603	7.439	1.658
H α	6564.61	–	14.908	41.231	26.957	10.188	22.476	19.225	–	3.751	39.508	21.372	6.435	16.527	4.380
N II	6585.27	–	5.677	12.908	13.600	9.450	15.528	7.629	–	2.137	12.745	11.214	6.453	12.834	2.436
S II	6718.29	–	2.053	7.254	4.735	3.932	4.790	2.853	–	–	6.906	4.380	3.617	4.245	1.100
S II	6732.67	–	1.478	5.205	3.748	3.351	4.242	2.088	–	–	4.973	3.682	3.316	3.932	–

The most numerous galaxies (passive red ellipticals and star-forming blue spirals) show significant variations in their colours. Using only a few templates might cause the photo- z estimates to be coarse because the templates are unable to characterize this intrinsic colour variance of the objects. More reliable empirical templates, however, could be used to refine photometric redshift estimation. In order to address this issue, we further refine the most numerous galaxy classes by colour and calculate their composite spectra.

3.4.1 Blue galaxies

In the case of star-forming galaxies (mostly spirals and irregulars), the intrinsic colour variance is caused by many effects. Star formation rate is primarily responsible for the rest-frame colours of galaxies, while dust content and inclination play an important role too (Yip et al. 2010). Also, since fibre spectrographs with limited aperture only observe the central, redder region of the galaxies, the distance of the galaxies and their bulge-to-disc ratio is an important

Table 5. Values of the emission-line equivalent widths (in Å) for the refined galaxy classes.

	λ (Å)	RED 1	RED 2	RED 3	RED 4	RED 5	SF 1	SF 2	SF 3	SF 4	SF 5
O II	3727.09	2.924	1.803	1.621	1.992	2.323	28.630	21.238	13.033	9.077	7.940
O II	3729.88	1.577	1.254	1.200	1.445	1.531	–	26.363	16.573	11.115	9.387
H θ	3798.98	–	–	–	–	–	–	–	–	–	–
H η	3836.47	–	–	–	–	–	–	–	–	–	–
He I	3889.00	–	–	–	–	–	1.618	1.292	1.378	1.343	1.392
S II	4072.30	–	–	–	–	–	–	–	–	–	–
H δ	4102.89	–	–	–	–	–	2.706	2.103	2.069	1.920	1.842
H γ	4341.68	–	–	–	–	–	6.111	4.408	3.916	3.457	3.166
O III	4364.44	–	–	–	–	–	–	–	–	–	–
H β	4862.68	1.273	–	–	–	–	17.348	11.736	9.573	8.072	7.133
O III	4932.60	–	–	–	–	–	–	–	–	–	–
O III	4960.30	–	–	–	–	–	10.865	3.198	1.175	–	–
O III	5008.24	–	–	–	–	–	35.024	10.803	4.252	2.568	1.997
O I	6302.05	–	–	–	–	–	2.400	1.600	1.239	1.107	1.072
O I	6365.54	–	–	–	–	–	–	–	–	–	–
N I	6529.03	–	–	–	–	–	–	–	–	–	–
N II	6549.86	2.276	–	–	–	1.038	4.771	4.464	4.372	4.355	5.050
H α	6564.61	6.305	2.206	1.123	1.409	1.958	84.884	53.510	41.972	36.237	33.311
N II	6585.27	3.229	1.598	–	1.281	1.657	13.877	13.506	13.443	13.223	13.223
S II	6718.29	1.276	–	–	–	–	12.821	9.729	7.407	5.949	5.322
S II	6732.67	1.033	–	–	–	–	9.335	6.959	5.312	4.353	4.001

factor causing colour variations in small aperture magnitudes. We will denote these classes as ‘SF’ followed by a number.

3.4.2 Red galaxies

In the case of red galaxies (mostly ellipticals), the origin of the colour variation is mainly an evolutionary effect caused by their passively evolving old stellar populations (for a comprehensive review, see Renzini (2006), and references therein); the effect of dust content on their optical colours is thought to be minimal.

For the purposes of refined colour classification, we will use all galaxies of the former ‘passive’ and ‘H α measured’ classes and denote these class as ‘RED’ followed by a number.

3.4.3 The refined colour classes

We define the refined colour classes based on the $g - r$ colour distributions of star-forming and passive galaxies as plotted in Fig. 6. The histograms of the distributions are fitted with Gaussians and the colour limits are defined at 0.5σ and 1.5σ distances from the mean. Altogether, five additional colour classes are defined, as listed in Table 6.

Parameters m and σ denote the mean and standard deviation of the fitted distributions, respectively. The fitted parameters are summarized in Table 7.

Table 6. Definitions of the refined colour cuts.

#	Limits
1	$g - r < m - 1.5\sigma$
2	$m - 1.5\sigma \leq g - r < m - 0.5\sigma$
3	$m - 0.5\sigma \leq g - r \leq m + 0.5\sigma$
4	$m + 0.5\sigma < g - r \leq m + 1.5\sigma$
5	$m + 1.5\sigma < g - r$

Table 7. Parameters of the refined colour cuts.

	m	σ
SF	0.48	0.16
RED	0.95	0.06

3.5 Redshift distribution

The redshift distribution of the galaxies is plotted in Fig. 7 for every galaxy activity class in redshift bins of $\Delta z = 0.01$. It is clear from the curves that all samples except red galaxies are flux limited according to the spectroscopic targeting algorithm of SDSS in the investigated redshift range. The bright-end red galaxy sample is volume limited to $z = 0.35$ (Eisenstein et al. 2001). Fig. 8 shows the same redshift distributions *after* applying the absolute magnitude cuts to get the volume-limited samples. Note that the requirement to select volume-limited samples does not allow maximizing galaxy counts by freely choosing the used redshift intervals; absolute magnitude limits also have to be taken into account.

4 COMPOSITE SPECTRA

4.1 Pre-processing

We selected spectra based on the criteria defined in Section 3 and pre-processed them before calculating the composites. First, all spectra were corrected for foreground galactic extinction using the dust map of Schlegel, Finkbeiner & Davis (1998) and the extinction curve from O’Donnell (1994). Secondly, spectra were normalized as described in Section 4.2. Thirdly, spectra were converted to rest frame and rebinned to a grid with the intended resolution ($\Delta\lambda = 1.0, 0.5$ or 0.25 Å).

4.2 Normalization

To derive meaningful composites, one has to normalize the galaxy spectra in a robust way. Normalizing spectra at a given wavelength

Table 8. Values of the continuum absorption indices for every galaxy class, continued in Table 9.

		Red P	Red H α	Red SF	Red A+H II	Red L	Red S	Red all	Green P	Green H α	Green SF	Green A+H II	Green L	Green S	Green all
CN ₁	mag	-0.033	-0.038	-0.062	-0.044	-0.055	-0.031	-0.041	0.050	0.032	-0.060	-0.029	0.045	-0.004	0.033
CN ₂	mag	0.023	0.002	-0.037	-0.009	-0.005	0.005	-0.004	0.098	0.077	-0.035	0.006	0.086	0.032	0.077
Ca4227	Å	0.875	0.785	0.447	0.615	0.728	0.663	0.720	1.260	1.172	0.468	0.744	1.160	0.861	1.147
G4300	Å	3.810	2.905	1.185	2.205	2.718	2.841	2.613	5.492	4.917	1.296	2.811	4.942	3.732	4.834
Fe4383	Å	3.467	3.124	1.640	2.698	3.120	3.595	2.928	4.945	4.638	1.737	3.173	4.814	4.351	4.576
Ca4455	Å	1.255	1.122	0.580	0.900	1.019	1.017	1.035	1.599	1.509	0.614	1.035	1.415	1.219	1.477
Fe4531	Å	3.156	2.720	1.913	2.384	2.614	2.644	2.591	3.511	3.336	1.965	2.612	3.309	2.983	3.302
C4668	Å	4.718	3.792	1.839	3.065	3.908	3.153	3.491	6.368	5.928	1.951	3.757	6.116	4.145	5.897
H β	Å	3.187	0.212	-4.591	-1.915	1.492	-1.474	-0.550	1.913	1.364	-4.348	-1.222	0.673	-0.739	1.240
Fe5015	Å	4.695	3.583	-0.179	0.914	1.673	-7.702	2.748	5.008	4.525	0.155	1.825	2.987	-5.896	4.369
Mg ₁	mag	0.055	0.051	0.025	0.048	0.059	0.067	0.049	0.101	0.093	0.027	0.058	0.104	0.079	0.094
Mg ₂	mag	0.159	0.145	0.086	0.127	0.149	0.150	0.139	0.238	0.223	0.090	0.150	0.234	0.178	0.224
Mg <i>b</i>	Å	2.671	2.646	1.812	2.557	2.980	2.923	2.585	3.839	3.686	1.884	2.887	4.113	3.347	3.684
Fe5270	Å	2.314	2.221	1.517	1.893	2.014	2.082	2.104	2.680	2.624	1.563	2.071	2.532	2.358	2.588
Fe5335	Å	2.159	2.144	1.543	1.907	2.019	2.057	2.050	2.382	2.370	1.581	2.017	2.310	2.260	2.327
Fe5406	Å	1.369	1.341	0.945	1.176	1.253	1.272	1.278	1.573	1.552	0.972	1.262	1.518	1.449	1.524
Fe5709	Å	0.823	0.784	0.594	0.701	0.760	0.803	0.754	0.906	0.890	0.611	0.758	0.871	0.874	0.877
Fe5782	Å	0.641	0.707	0.538	0.687	0.710	0.716	0.693	0.756	0.772	0.564	0.728	0.784	0.769	0.763
Na D	Å	2.566	2.209	1.282	2.363	2.447	2.121	2.208	3.366	3.266	1.495	2.864	3.471	2.678	3.302
TiO ₁	mag	0.022	0.022	0.016	0.020	0.024	0.025	0.022	0.032	0.031	0.016	0.023	0.032	0.027	0.031
TiO ₂	mag	0.053	0.053	0.041	0.051	0.056	0.056	0.052	0.073	0.070	0.042	0.056	0.074	0.062	0.070
H δ_A	Å	2.656	2.405	3.501	2.829	3.544	2.447	2.608	-1.280	-0.514	3.409	2.176	-0.668	1.180	-0.445
H γ_A	Å	-0.344	-0.822	0.128	-0.668	0.334	-1.528	-0.705	-4.969	-4.202	0.026	-1.476	-4.852	-2.984	-4.172
H δ_F	Å	2.676	2.265	2.411	2.261	3.096	2.084	2.287	0.625	0.934	2.373	1.973	0.823	1.501	0.944
H γ_F	Å	1.562	0.643	0.225	0.347	1.653	0.101	0.559	-1.092	-0.865	0.211	0.035	-1.279	-0.531	-0.867
Ca II λ 8498	Å	0.794	0.854	0.757	0.820	0.806	0.842	0.842	1.150	1.155	0.775	1.001	1.122	0.880	1.147
Ca II λ 8542	Å	2.723	2.741	2.488	2.673	2.615	2.710	2.720	3.334	3.342	2.514	2.939	3.259	2.757	3.323
Ca II λ 8662	Å	2.719	2.802	2.683	2.777	2.696	2.779	2.798	3.061	3.082	2.691	2.940	2.982	2.759	3.062
Mg I λ 8807	Å	0.448	0.473	0.343	0.415	0.382	0.449	0.460	0.372	0.382	0.359	0.343	0.339	0.465	0.371
CNB	mag	0.115	0.080	0.028	0.060	0.087	0.051	0.071	0.202	0.166	0.030	0.073	0.161	0.080	0.163
H + K	mag	0.333	0.248	0.183	0.216	0.258	0.229	0.235	0.373	0.336	0.187	0.239	0.329	0.263	0.331
Ca I	mag	0.007	0.009	0.006	0.007	0.006	0.005	0.008	0.015	0.014	0.007	0.009	0.013	0.007	0.014
G	mag	0.173	0.141	0.073	0.110	0.134	0.134	0.128	0.232	0.211	0.077	0.132	0.207	0.168	0.206
Hb	mag	0.104	0.046	-0.036	0.007	0.073	0.013	0.032	0.073	0.064	-0.032	0.020	0.051	0.026	0.062
MgG	mag	0.067	0.064	0.040	0.060	0.071	0.067	0.062	0.102	0.096	0.042	0.069	0.107	0.079	0.096
MH	mag	0.034	0.036	0.018	0.024	0.023	-0.028	0.032	0.060	0.055	0.020	0.031	0.051	-0.014	0.055
FC	mag	0.062	0.056	0.037	0.049	0.051	0.052	0.053	0.073	0.070	0.038	0.054	0.069	0.060	0.070
NaD	mag	0.057	0.044	0.014	0.044	0.053	0.039	0.043	0.080	0.075	0.019	0.057	0.080	0.053	0.076
D4000	-	1.785	1.572	1.348	1.481	1.617	1.543	1.533	2.070	1.938	1.364	1.576	1.971	1.677	1.924
D4000 _n	-	1.578	1.401	1.223	1.320	1.427	1.365	1.366	1.868	1.737	1.234	1.397	1.758	1.482	1.724

is not appropriate for noisy measurements; thus we use the following technique. We calculate the median flux in the 4200–4300, 4600–4800, 5400–5500 and 5600–5800 Å wavelength intervals and use this value to scale spectra together. The wavelength intervals are chosen to be devoid of any strong emission lines. Spectra are rescaled to have the median value of the flux in these intervals equal to 1.

We do not weight or otherwise scale the flux of the spectra based on the apparent or intrinsic luminosity of the objects or S/N of the observations. Although this could affect the S/N of the composites negatively, the robust algorithm can successfully overcome this problem by automatically assigning lower weights to spectra with lower S/N values.

4.3 Robust algorithm

In order to take the significant noise and the varying S/N of the original data set into account, we use the robust averaging method of Budavári et al. (2009) to determine the average and variance

spectra of the different galaxy classes. This algorithm not only computes the average and the variance iteratively by taking care of gaps in the measurements, but also yields the first few principal components of the spectrum distribution (see Section 4.5). We will detail the efficiency of the robust method with respect to ordinary average and median composite spectra in Section 4.6.

The composites calculated by the robust algorithm are plotted in Figs 9–12 for each activity and colour class.

4.4 High resolution

Because composites are combined from many spectra at varying redshifts, it is possible to get a couple of factors higher resolution than the resolution of the original measurements. The average error in redshift measurement of the high S/N ($S/N \geq 30$) galaxy spectra of SDSS is $z_{\text{err}} = 1.6 \times 10^{-4}$, which allows room for oversampling the composites. On the other hand, oversampling can introduce noise compared to the almost noiseless composites at 1 Å binning.

Table 9. Values of the continuum indices for every galaxy class, continued from Table 8.

		Blue P	Blue H α	Blue SF	Blue A+H II	Blue L	Blue S	Blue all	All P	All H α	All SF	All A+H II	All L	All S	All all
CN ₁	mag	-0.033	-0.038	-0.062	-0.044	-0.055	-0.031	-0.041	0.050	0.032	-0.060	-0.029	0.045	-0.004	0.033
CN ₂	mag	0.023	0.002	-0.037	-0.009	-0.005	0.005	-0.004	0.098	0.077	-0.035	0.006	0.086	0.032	0.077
Ca4227	Å	0.875	0.785	0.447	0.615	0.728	0.663	0.720	1.260	1.172	0.468	0.744	1.160	0.861	1.147
G4300	Å	3.810	2.905	1.185	2.205	2.718	2.841	2.613	5.492	4.917	1.296	2.811	4.942	3.732	4.834
Fe4383	Å	3.467	3.124	1.640	2.698	3.120	3.595	2.928	4.945	4.638	1.737	3.173	4.814	4.351	4.576
Ca4455	Å	1.255	1.122	0.580	0.900	1.019	1.017	1.035	1.599	1.509	0.614	1.035	1.415	1.219	1.477
Fe4531	Å	3.156	2.720	1.913	2.384	2.614	2.644	2.591	3.511	3.336	1.965	2.612	3.309	2.983	3.302
C4668	Å	4.718	3.792	1.839	3.065	3.908	3.153	3.491	6.368	5.928	1.951	3.757	6.116	4.145	5.897
H β	Å	3.187	0.212	-4.591	-1.915	1.492	-1.474	-0.550	1.913	1.364	-4.348	-1.222	0.673	-0.739	1.240
Fe5015	Å	4.695	3.583	-0.179	0.914	1.673	-7.702	2.748	5.008	4.525	0.155	1.825	2.987	-5.896	4.369
Mg ₁	mag	0.055	0.051	0.025	0.048	0.059	0.067	0.049	0.101	0.093	0.027	0.058	0.104	0.079	0.094
Mg ₂	mag	0.159	0.145	0.086	0.127	0.149	0.150	0.139	0.238	0.223	0.090	0.150	0.234	0.178	0.224
Mg <i>b</i>	Å	2.671	2.646	1.812	2.557	2.980	2.923	2.585	3.839	3.686	1.884	2.887	4.113	3.347	3.684
Fe5270	Å	2.314	2.221	1.517	1.893	2.014	2.082	2.104	2.680	2.624	1.563	2.071	2.532	2.358	2.588
Fe5335	Å	2.159	2.144	1.543	1.907	2.019	2.057	2.050	2.382	2.370	1.581	2.017	2.310	2.260	2.327
Fe5406	Å	1.369	1.341	0.945	1.176	1.253	1.272	1.278	1.573	1.552	0.972	1.262	1.518	1.449	1.524
Fe5709	Å	0.823	0.784	0.594	0.701	0.760	0.803	0.754	0.906	0.890	0.611	0.758	0.871	0.874	0.877
Fe5782	Å	0.641	0.707	0.538	0.687	0.710	0.716	0.693	0.756	0.772	0.564	0.728	0.784	0.769	0.763
Na D	Å	2.566	2.209	1.282	2.363	2.447	2.121	2.208	3.366	3.266	1.495	2.864	3.471	2.678	3.302
TiO ₁	mag	0.022	0.022	0.016	0.020	0.024	0.025	0.022	0.032	0.031	0.016	0.023	0.032	0.027	0.031
TiO ₂	mag	0.053	0.053	0.041	0.051	0.056	0.056	0.052	0.073	0.070	0.042	0.056	0.074	0.062	0.070
H δ_A	Å	2.656	2.405	3.501	2.829	3.544	2.447	2.608	-1.280	-0.514	3.409	2.176	-0.668	1.180	-0.445
H γ_A	Å	-0.344	-0.822	0.128	-0.668	0.334	-1.528	-0.705	-4.969	-4.202	0.026	-1.476	-4.852	-2.984	-4.172
H δ_F	Å	2.676	2.265	2.411	2.261	3.096	2.084	2.287	0.625	0.934	2.373	1.973	0.823	1.501	0.944
H γ_F	Å	1.562	0.643	0.225	0.347	1.653	0.101	0.559	-1.092	-0.865	0.211	0.035	-1.279	-0.531	-0.867
Ca II λ 8498	Å	0.794	0.854	0.757	0.820	0.806	0.842	0.842	1.150	1.155	0.775	1.001	1.122	0.880	1.147
Ca II λ 8542	Å	2.723	2.741	2.488	2.673	2.615	2.710	2.720	3.334	3.342	2.514	2.939	3.259	2.757	3.323
Ca II λ 8662	Å	2.719	2.802	2.683	2.777	2.696	2.779	2.798	3.061	3.082	2.691	2.940	2.982	2.759	3.062
Mg I λ 8807	Å	0.448	0.473	0.343	0.415	0.382	0.449	0.460	0.372	0.382	0.359	0.343	0.339	0.465	0.371
CNB	mag	0.115	0.080	0.028	0.060	0.087	0.051	0.071	0.202	0.166	0.030	0.073	0.161	0.080	0.163
H + K	mag	0.333	0.248	0.183	0.216	0.258	0.229	0.235	0.373	0.336	0.187	0.239	0.329	0.263	0.331
Ca I	mag	0.007	0.009	0.006	0.007	0.006	0.005	0.008	0.015	0.014	0.007	0.009	0.013	0.007	0.014
G	mag	0.173	0.141	0.073	0.110	0.134	0.134	0.128	0.232	0.211	0.077	0.132	0.207	0.168	0.206
Hb	mag	0.104	0.046	-0.036	0.007	0.073	0.013	0.032	0.073	0.064	-0.032	0.020	0.051	0.026	0.062
MgG	mag	0.067	0.064	0.040	0.060	0.071	0.067	0.062	0.102	0.096	0.042	0.069	0.107	0.079	0.096
MH	mag	0.034	0.036	0.018	0.024	0.023	-0.028	0.032	0.060	0.055	0.020	0.031	0.051	-0.014	0.055
FC	mag	0.062	0.056	0.037	0.049	0.051	0.052	0.053	0.073	0.070	0.038	0.054	0.069	0.060	0.070
NaD	mag	0.057	0.044	0.014	0.044	0.053	0.039	0.043	0.080	0.075	0.019	0.057	0.080	0.053	0.076
D4000	-	1.785	1.572	1.348	1.481	1.617	1.543	1.533	2.070	1.938	1.364	1.576	1.971	1.677	1.924
D4000 _n	-	1.578	1.401	1.223	1.320	1.427	1.365	1.366	1.868	1.737	1.234	1.397	1.758	1.482	1.724

This noise can become significant for sparsely populated galaxy classes such as the green-valley galaxies.

Another important factor that counteracts the efforts to get higher resolution is the intrinsic broadening of the spectral features of the individual spectra due to the high-velocity kinetics in the galaxies. To overcome this problem, deconvolution of the Doppler broadening kernel from the original measurements would be necessary. We do not address this issue in the current work.

Fig. 13 shows the comparison of some spectral features of a galaxy composite with different oversampling (at the resolution of 1, 0.5, 0.25 and 0.1 Å). Note how the line features become smoother with higher resolution. It is important, though, that oversampled spectra might not contain more information because the wide velocity dispersion washes out fine details.

Evidently, linear interpolation could not be used to get the higher resolution composites from the lower resolution ones. Higher order curves could be used for oversampling low-resolution composites to get smooth line profiles, but they are likely to introduce artefacts. Our method of oversampling overcomes this problem.

4.5 Generating empirical mock catalogues using eigenspectra

Besides the robust average spectrum of every galaxy class, we also calculate the variance spectra. These can be used to determine the wavelength ranges with outstanding variance (usually the emission lines) and quantify the noise in each wavelength bin. The variance spectra are available online. However, if one is interested in the overall variance of the spectra (in contrast to the bin-by-bin variance) of a galaxy class, it is not enough to use the variance spectrum where bins are treated independently. In order to quantify the variance among the spectra of a certain galaxy class over the entire wavelength range, and take the covariance of the bins into account, the first few principal components (eigenspectra) have to be calculated. Fortunately, our robust averaging algorithm yields the eigenspectra as a ‘by-product’ of the averaging.

We calculate the first five eigenspectra of every galaxy class and expand each measured spectrum on this basis. In most cases, the distribution of the expansion coefficients is very close to a Gaussian distribution centred on zero. We determine the standard deviation

Table 10. Value of the continuum absorption indices for the refined colour classes.

		RED 1	RED 2	RED 3	RED 4	RED 5	SF 1	SF 2	SF 3	SF 4	SF 5
CN ₁	mag	−0.002	0.032	0.063	0.079	0.066	−0.047	−0.069	−0.064	−0.054	−0.047
CN ₂	mag	0.043	0.078	0.110	0.124	0.110	−0.045	−0.049	−0.039	−0.027	−0.019
Ca4227	Å	1.038	1.209	1.304	1.301	1.286	0.205	0.341	0.429	0.539	0.622
G4300	Å	4.289	5.175	5.563	5.580	5.465	0.092	0.597	1.130	1.659	2.095
Fe4383	Å	4.177	4.809	5.087	5.108	5.046	1.123	1.194	1.576	2.064	2.461
Ca4455	Å	1.418	1.575	1.604	1.574	1.536	0.096	0.361	0.576	0.735	0.854
Fe4531	Å	3.190	3.426	3.499	3.483	3.430	1.183	1.625	1.890	2.122	2.318
C4668	Å	5.160	6.034	6.615	6.927	6.669	0.683	1.198	1.751	2.337	2.773
Hβ	Å	1.407	1.734	1.597	1.434	1.409	−12.371	−6.644	−4.584	−3.640	−3.117
Fe5015	Å	4.305	4.728	4.825	4.741	4.567	−16.711	−3.991	0.305	1.743	2.286
Mg ₁	mag	0.072	0.089	0.109	0.120	0.113	0.019	0.019	0.024	0.032	0.039
Mg ₂	mag	0.186	0.218	0.251	0.268	0.256	0.054	0.068	0.083	0.101	0.115
Mg <i>b</i>	Å	3.210	3.626	4.039	4.232	4.085	1.301	1.520	1.763	2.080	2.333
Fe5270	Å	2.544	2.690	2.706	2.690	2.669	0.800	1.229	1.495	1.710	1.842
Fe5335	Å	2.370	2.447	2.380	2.316	2.320	0.963	1.295	1.525	1.708	1.802
Fe5406	Å	1.525	1.594	1.582	1.543	1.525	0.594	0.800	0.933	1.054	1.130
Fe5709	Å	0.889	0.923	0.901	0.878	0.897	0.376	0.505	0.578	0.654	0.705
Fe5782	Å	0.770	0.780	0.774	0.776	0.785	0.303	0.412	0.523	0.631	0.716
Na D	Å	2.732	3.077	3.575	4.066	4.481	0.015	0.633	1.162	1.898	2.826
TiO ₁	mag	0.026	0.030	0.033	0.035	0.034	0.012	0.014	0.016	0.017	0.018
TiO ₂	mag	0.062	0.069	0.076	0.080	0.078	0.032	0.034	0.040	0.046	0.050
Hδ _A	Å	0.762	−0.698	−1.724	−1.993	−1.728	3.012	3.859	3.591	3.057	2.602
Hγ _A	Å	−2.713	−4.343	−5.508	−5.856	−5.574	−1.284	0.332	0.260	−0.311	−0.868
Hδ _F	Å	1.568	0.893	0.379	0.222	0.341	1.653	2.489	2.463	2.217	1.986
Hγ _F	Å	−0.081	−0.834	−1.457	−1.655	−1.514	−1.718	−0.004	0.301	0.169	−0.037
Ca II λ8498	Å	1.078	1.167	1.144	1.127	1.187	0.470	0.561	0.752	0.807	0.814
Ca II λ8542	Å	3.231	3.370	3.318	3.272	3.445	1.967	2.166	2.476	2.581	2.556
Ca II λ8662	Å	3.074	3.114	3.010	2.961	3.182	2.329	2.476	2.682	2.728	2.676
Mg I λ8807	Å	0.410	0.392	0.365	0.342	0.478	0.101	0.200	0.342	0.408	0.421
CNB	mag	0.125	0.175	0.213	0.229	0.214	−0.006	0.021	0.027	0.037	0.046
H + K	mag	0.310	0.352	0.372	0.375	0.368	0.113	0.163	0.182	0.198	0.218
Ca I	mag	0.012	0.013	0.016	0.016	0.016	0.002	0.005	0.006	0.007	0.008
G	mag	0.193	0.224	0.232	0.227	0.224	0.030	0.052	0.071	0.090	0.106
Hb	mag	0.067	0.071	0.067	0.065	0.065	−0.166	−0.070	−0.036	−0.020	−0.010
MgG	mag	0.080	0.093	0.108	0.116	0.110	0.026	0.033	0.039	0.047	0.054
MH	mag	0.044	0.054	0.063	0.068	0.065	−0.064	0.000	0.020	0.028	0.033
FC	mag	0.066	0.071	0.074	0.075	0.074	0.017	0.029	0.037	0.043	0.046
NaD	mag	0.060	0.071	0.085	0.098	0.105	−0.031	−0.006	0.011	0.029	0.050
D4000	–	1.814	1.993	2.105	2.146	2.145	1.161	1.273	1.342	1.415	1.514
D4000 _n	–	1.611	1.787	1.903	1.941	1.919	1.081	1.166	1.217	1.271	1.334

σ_{eig} of the distribution of the eigencoefficients in each galaxy class. Then one can calculate the following spectra:

$$F_{\lambda}^{\pm} = F_{\lambda}^{\text{avg}} \pm k \sigma_{\text{eig}} F_{\lambda}^{\text{eig}}, \quad (1)$$

where F_{λ}^{avg} and F_{λ}^{eig} are the average and one of the eigenspectra, respectively. The coefficient k is a constant factor that can be varied to get slightly different spectra.

As an illustration, we plot F_{λ}^{\pm} for the refined classes of red and blue galaxies in Fig. 14. In case of the red galaxies, the first eigenspectrum clearly determines the colour of the galaxy; almost the entire variation in the overall slope of spectrum is picked up by the first principal component. In case of the star-forming galaxies, most of the variance is in the emission lines, so one has to use the second eigenvector to generate spectra with colour variations. As Fig. 14 shows, this technique can be successfully used to generate composites of any colour between the averages of the pre-defined classes. One drawback of the method is that adding higher order eigenspectra to the averages introduces some noise.

The importance of this technique is that it allows generating mock spectral libraries. By drawing the value of k of equation (1)

from a normal distribution, the F_{λ}^{\pm} spectra will be representative to the whole ensemble of SDSS galaxies. Eigenspectra of the various galaxy classes are published online along with the parameters (mean and σ) of the distributions of the expansion coefficients of individual observations on the eigenbases. By using these parameters, linear combinations of the average composite spectra and eigenspectra can be easily calculated to further extend the parameter space coverage of the presented atlas.

4.6 Performance of the robust method

To quantify the performance of the robust method, we compare the robustly averaged composites with normally averaged and median composites. Averaging is thought to result in composites that are physical in the sense that they are constructed as non-negative linear combinations of real stellar continua. Median, while it is robust against noise, differs from averaging because it treats the spectral bins independently and might yield non-physical continua. One disadvantage of averaging spectra is that emission-line ratios are

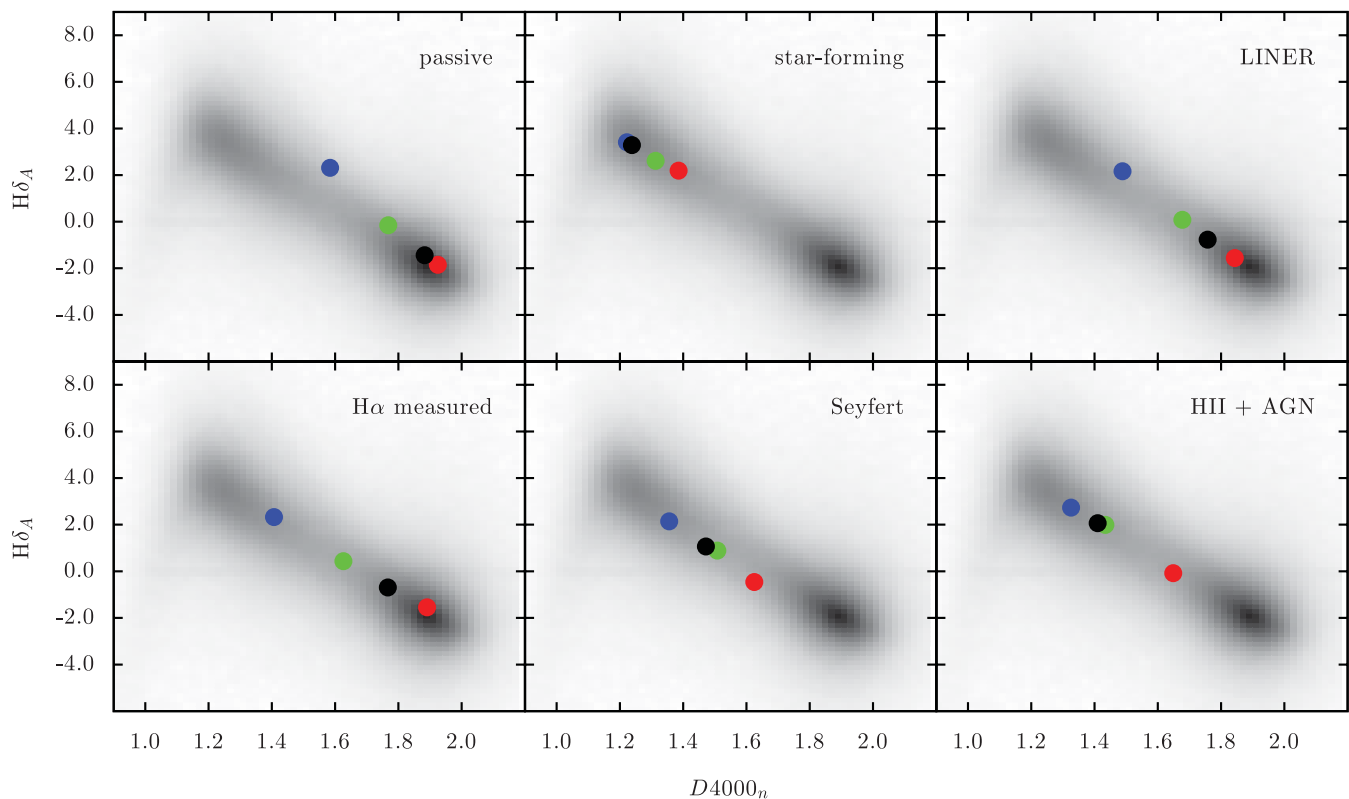


Figure 20. Recent star formation signatures in the composites. The density plot is the square root of the number density of all SDSS galaxies. The loci of the composites are indicated with circles. The different colours denote the different colour classes; black points mark the loci of composites of all galaxies (regardless of their colours) in a given activity class. See the online edition for a colour version of this plot.

not guaranteed to be preserved, whereas Vanden Berk et al. (2001) reported that taking the median spectra of quasars did not alter the emission-line ratios. We will revisit the question of emission lines of average composites in Section 5.3. The main advantage of robust averaging over median calculation is that the former can be done iteratively, and there is no need to store and sort all the data in memory.

We plot the difference between the robust average and the non-robust composites in Figs 15 and 16 (middle panels). The robust composite and the number counts used to calculate the composites in each wavelength bin are also plotted for reference in the top and bottom panels, respectively.

Interestingly, for red galaxies, the median runs much closer to the robust average than the normal average, while in the case of blue galaxies, the differences are not as significant as in the case of the red galaxies. Significant differences between the robust and non-robust composites are visible at very short and very long wavelengths, and also at emission lines. The differences between the continua are possibly due to the low number counts in those wavelength ranges. For low number counts, non-robust methods tend to become severely biased. The robust algorithm assigns lower weights to galaxies with extremely strong emission lines, which is considered to be the reason behind the large difference between the robust and non-robust methods in the emission-line regions.

5 COMPOSITES ON THE MAP

In this section, we use standard techniques to estimate the stellar continua, emission-line strengths, absorption-line indices and physical parameters of the composites. We follow the methods de-

scribed by Kauffmann et al. (2003a,b), Brinchmann et al. (2004) and Tremonti et al. (2004). These techniques were developed and used by the cited authors to compose the MPA/JHU Value-Added Galaxy Catalogue² of SDSS DR7.

In this paper, we do not intend to make any conclusions about physical parameters of the galaxy samples based on the parameters derived from the composites. Our goal is simply to show that the composites calculated using our method cover much of the parameter space of the individual galaxies. To achieve this, we plot several standard diagrams widely used in the literature. These diagrams show the distributions of the parameters of individual galaxies of the SDSS or MPA/JHU value-added catalogues overplotted with the parameters of the composites.

5.1 Continuum model fits

The wavelength coverage of our empirically derived high S/N composite spectra spans from about 3350 to 9050 Å. The actual coverage depends on the redshift cuts of the volume-limited samples of the galaxy classes; higher redshift samples allow extending the coverage bluewards. In order to extend this limited coverage, we fit the composites with theoretical model spectra generated using the single stellar population synthesis models of Bruzual & Charlot (2003). Three sets of models were generated with different metallicity values of $Z = 0.008, 0.02$ and 0.05 . Each set contains 10 single burst models with ages of $t = 0.005, 0.025, 0.100, 0.290, 0.640$,

² <http://www.mpa-garching.mpg.de/SDSS/>

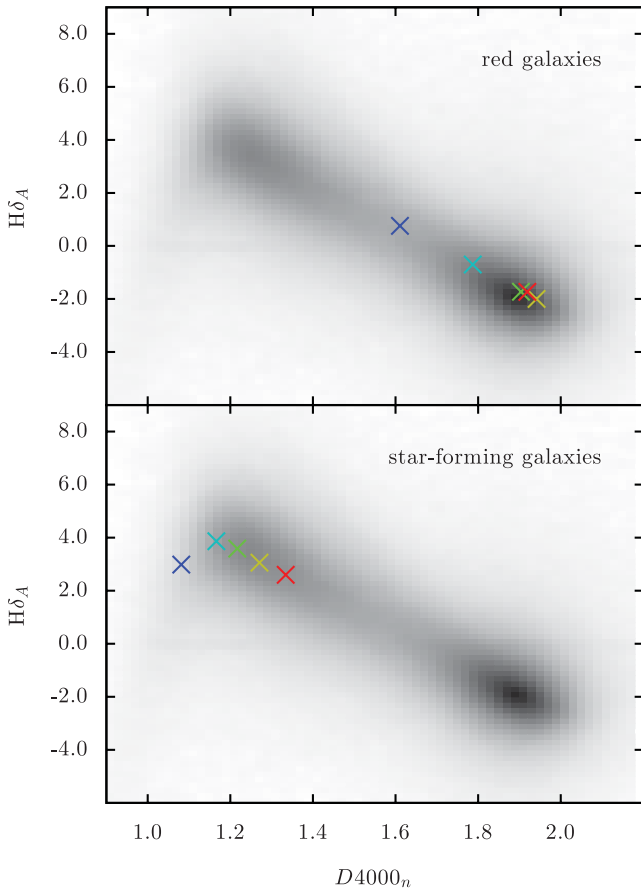


Figure 21. Recent star formation signatures in the refined composites. The density plot is the square root of the number density of all SDSS galaxies. The crosses mark the loci of the composites based on the refined colour classification; see Section 3.4.

0.900, 1.4, 2.5, 5.0 and 11.0 Gyr based on the initial mass function of Chabrier (2003) and the evolutionary tracks ‘Padova1994’ (Alongi et al. 1993; Bressan et al. 1993; Fagotto et al. 1994a,b). We use the non-negative least-squares (NNLS) method to determine the best linear combination of models of different ages (but same metallicity) to fit the composite spectra. To estimate the effect of intrinsic extinction by dust, we use the simple dust model of Bruzual & Charlot (2003), which is described by the following formula:

$$\tau_{\lambda}(t) = \begin{cases} \tau_V(\lambda/5500 \text{ \AA})^{-0.7} & \text{for } t \leq 10^7 \text{ yr,} \\ \mu \tau_V(\lambda/5500 \text{ \AA})^{-0.7} & \text{for } t > 10^7 \text{ yr.} \end{cases}$$

We let the value of the optical depth τ_V vary while keeping the value of $\mu = 0.3$ fixed. In this particular dust model, μ controls the ratio of the effect of extinction on stars older/younger than 10 Myr. We also keep the value of the velocity dispersion of the stars σ_v a free parameter. For each composite, we repeat the fitting for each model set, and the metallicity of the best-fitting model set is used.

To compute the synthetic magnitude of the composites in broad-band filters where the wavelength coverage of the composites is not wide enough, we extend the empirical spectra with the best-fitting models. We also use the same continuum fits to measure the emission-line equivalent widths of the composites.

5.2 Colour–colour trajectories

As we intend to use the composite atlas as a basis for template-based photometric redshift estimation, it is important that the composites cover much of the colour space spanned by the SDSS galaxies. Figs 17 and 18 display the SDSS $g - r$ and $r - i$ colour–colour trajectories of composites for every galaxy class we defined. The trajectories cover the redshift range $0.0 \leq z \leq 0.7$, and they nicely overlap with the measured galaxy distributions. This is a good confirmation that the robust averaging algorithm does not introduce any bias into galaxy colours.

It is observable in Fig. 17 that the bluest galaxies ($g - r < 0.5$) are not covered by the composites of the main galaxy activity classes. The refined classification of star-forming galaxies, however, yields composites of extremely blue colours, as it can be seen in Fig. 18.

5.3 Emission-line measurements

As averaging spectra may have the effect of altering line ratios, we have to verify whether the emission lines of the composites are physical or not. Fortunately, in our case the robust averaging algorithm does not alter the line ratios severely, as we will show below.

First, emission-line equivalent widths are determined by using the best-fitting continuum model as described in Section 5.1. Then, the BPT diagram is plotted for the composites of the star-forming, AGN + H II, LINER and Seyfert galaxies in Fig. 19. The loci of the composites overlap with the line ratios calculated from the individual galaxy measurements. It is important to mention, however, that the variance in line ratios of galaxies is much larger than what a few composite spectra can characterize.

One interesting feature of Fig. 19 is that Seyferts are located along a well-defined line parametrized by the $g - r$ colour: redder Seyferts tend to have higher $[\text{O III}]/\text{H}\beta$ ratios. This is consistent with the findings of Kewley et al. (2006): the stellar populations of AGNs are the older the further away they are located from the star-forming line on the BPT diagram. For AGN + H II galaxies, no similar behaviour is readily visible.

In the case of the star-forming galaxies, we also plot the results from the refined colour classification. A clear trend is observable here too: bluer galaxies have higher $[\text{O III}]/\text{H}\beta$ ratios and lower $[\text{N II}]/\text{H}\alpha$ ratios. Bluer star-forming composites (crosses), however, tend to be slightly above the star-forming line of the MPA/JHU results (density plot). The most trivial explanation to this effect is that averaging alters the line ratios, and this bias is more significant in the case of stronger emission lines.

Emission-line equivalent widths of all composites are listed in Tables 3–5 and are available online.

5.4 Continuum indices

We calculate numerous absorption-line indices of the composites: Lick indices (Worthey et al. 1994; Worthey & Ottaviani 1997), BH indices (Huchra et al. 1996) and DTT indices (Diaz, Terlevich & Terlevich 1989). Indices characterizing the size of the 4000-Å break are from Bruzual (1983) and Balogh et al. (1999). Absorption indices for all composites are summarized in Tables 8–10 and are available online.

One widely used diagram to separate star-forming, starburst and post-starburst galaxies from passive ones based on absorption-line features is the plot of the $D4000_n$ – $\text{H}\delta_A$ plane (Kauffmann et al.

Table 11. Modes and medians of the probability distributions of the physical parameters of the composites resulting from the Bayesian continuum fitting technique. t_{form} : age of the continuous model component; γ : time-scale parameter of the star formation rate; t_M : average age of the stellar population, weighted by the stellar mass; t_r : average age of the stellar population, weighted by the r -band magnitude; Z : metallicity; τ_V : optical depth of the dust component; SSFR: specific star formation rate; M/L : mass-to-light ratio.

	t_{form} (Gyr)		γ (Gyr)		t_M (Gyr)		t_r (Gyr)		Z (Z_{\odot})		τ_V		log SSFR		log M/L	
	Median	Mode	Median	Mode	Median	Mode	Median	Mode	Median	Mode	Median	Mode	Median	Mode	Median	Mode
P	10.33	11.42	0.68	0.89	6.42	6.42	5.31	4.57	1.09	0.89	0.80	0.06	-12.11	-11.40	0.49	0.49
H α	10.33	11.69	0.66	0.97	6.42	6.67	5.55	4.57	1.21	1.09	1.16	0.67	-11.93	-11.40	0.57	0.57
SF	8.69	10.88	0.50	0.81	5.17	4.17	4.08	3.10	1.09	1.71	1.90	1.53	-11.05	-10.70	0.57	0.57
A+H II	9.78	12.78	0.60	0.85	5.92	5.42	5.06	4.57	1.24	1.71	1.53	1.28	-11.58	-11.05	0.57	0.57
L	10.33	12.78	0.66	0.85	6.67	6.67	5.55	4.57	1.28	1.55	1.28	1.28	-11.93	-11.23	0.57	0.57
S	9.78	12.78	0.62	0.77	6.17	5.42	5.06	4.57	1.17	1.71	1.41	1.28	-11.58	-11.05	0.57	0.57
all	10.33	11.97	0.68	0.97	6.42	6.67	5.55	4.57	1.21	1.09	1.04	0.67	-12.11	-11.40	0.57	0.57
P	8.97	10.88	0.62	0.83	5.17	4.67	3.83	3.34	0.89	0.62	0.80	0.06	-11.40	-11.05	0.31	0.31
H α	9.24	11.97	0.58	0.77	5.42	4.67	4.32	3.83	1.01	0.55	1.28	1.16	-11.23	-10.88	0.40	0.40
SF	7.33	4.87	0.44	0.13	4.17	3.92	3.10	2.36	1.01	0.55	1.65	1.28	-10.53	-10.35	0.40	0.31
A+H II	8.15	6.24	0.50	0.30	4.92	4.67	3.83	3.34	0.97	0.55	1.53	1.16	-10.88	-10.53	0.40	0.40
L	9.51	11.97	0.58	0.77	5.67	4.67	4.57	3.83	1.01	0.74	1.28	1.16	-11.40	-11.05	0.49	0.49
S	8.69	6.24	0.54	0.77	5.17	4.67	4.08	3.83	0.97	0.55	1.41	1.16	-11.05	-10.70	0.40	0.40
all	9.24	11.97	0.58	0.77	5.42	4.67	4.32	3.83	1.01	0.55	1.16	1.16	-11.40	-10.88	0.40	0.40
P	7.87	6.78	0.54	0.91	4.17	3.42	2.85	2.12	0.62	0.20	0.92	0.06	-11.05	-10.53	0.14	0.14
H α	7.33	4.87	0.44	0.32	4.17	3.92	3.10	2.36	0.86	0.16	1.28	1.16	-10.53	-10.35	0.31	0.31
SF	4.33	3.24	0.34	0.01	2.42	1.92	1.63	1.38	0.93	0.12	1.28	1.16	-9.82	-9.82	-0.03	-0.03
A+H II	6.51	4.87	0.38	0.03	3.92	2.92	2.85	2.12	0.93	0.16	1.41	1.28	-10.35	-10.35	0.23	0.23
L	7.33	5.42	0.46	0.32	4.17	3.42	3.10	2.85	0.82	0.16	1.16	1.16	-10.70	-10.53	0.23	0.23
S	7.06	4.87	0.42	0.13	4.17	3.42	3.10	2.36	0.89	0.16	1.28	1.16	-10.53	-10.35	0.31	0.31
all	6.78	4.87	0.40	0.32	4.17	3.42	3.10	2.36	0.89	0.16	1.28	1.28	-10.53	-10.35	0.23	0.23
P	10.06	11.15	0.66	0.89	6.17	6.42	5.06	4.57	1.09	0.89	0.80	0.06	-11.93	-11.23	0.49	0.49
H α	9.78	11.15	0.62	0.77	5.92	4.67	4.82	3.83	1.13	1.09	1.16	1.16	-11.58	-11.05	0.49	0.49
SF	4.87	3.24	0.34	0.01	2.67	2.17	1.87	1.63	0.93	0.12	1.28	1.16	-10.00	-9.82	0.06	0.06
A+H II	7.87	5.69	0.48	0.30	4.67	3.92	3.59	2.85	0.97	0.55	1.41	1.16	-10.70	-10.53	0.40	0.40
L	9.78	12.78	0.62	0.85	6.17	6.67	5.06	4.57	1.24	1.52	1.41	1.28	-11.76	-11.05	0.57	0.57
S	8.69	6.24	0.54	0.30	5.17	4.67	4.08	3.83	1.01	0.55	1.41	1.16	-11.05	-10.70	0.40	0.40
all	9.78	11.15	0.62	0.77	5.92	4.67	4.82	3.83	1.13	1.09	1.16	1.16	-11.58	-11.05	0.49	0.49
RED 1	9.24	10.88	0.58	0.77	5.42	4.67	4.32	3.83	0.97	0.55	1.16	1.16	-11.23	-10.88	0.40	0.40
RED 2	10.06	11.15	0.64	0.77	5.92	4.67	4.82	3.83	1.09	0.89	1.04	0.67	-11.76	-11.23	0.49	0.49
RED 3	10.33	11.97	0.68	0.97	6.42	6.67	5.55	4.57	1.17	1.09	1.04	0.67	-12.11	-11.40	0.49	0.49
RED 4	10.60	12.78	0.70	0.83	6.92	6.67	6.04	5.55	1.28	1.17	1.16	1.28	-12.11	-11.76	0.57	0.57
RED 5	10.60	12.78	0.68	0.85	7.17	7.92	6.29	5.55	1.48	1.94	1.53	1.53	-12.11	-11.76	0.74	0.74
SF 1	1.33	1.06	0.44	0.56	0.67	0.42	0.40	0.40	0.86	0.78	1.16	0.92	-8.94	-8.94	-0.63	-0.63
SF 2	2.69	1.87	0.36	0.01	1.42	0.92	0.89	0.89	0.93	0.12	1.16	0.80	-9.47	-9.47	-0.28	-0.37
SF 3	4.06	2.96	0.34	0.01	2.42	1.92	1.63	1.38	0.97	0.12	1.28	1.16	-9.82	-9.82	-0.03	-0.03
SF 4	5.69	3.51	0.36	0.13	3.42	2.67	2.36	1.63	1.01	0.78	1.41	1.16	-10.17	-10.00	0.23	0.23
SF 5	7.60	4.87	0.46	0.30	4.67	3.92	3.34	2.36	1.01	0.55	1.65	1.53	-10.70	-10.35	0.40	0.40

2003b,c). As an illustration to the parameter space coverage of our composite atlas, we plot the $D4000_n$ - $H\delta_A$ diagram in Figs 20 and 21. According to Kauffmann et al. (2003c), in this plane, galaxies are located along a straight line: low-mass star-forming galaxies are close to the top-left corner of the plot, while high-mass passive galaxies are in the bottom-right corner. Intermediate-mass galaxies ($10^{10} \leq M_{\odot} \leq 10^{11.5}$) show some bimodality and populate the two big clumps as well as the line between them. Starburst galaxies are located slightly above the straight line (Kauffmann et al. 2003b). The composites clearly follow the relationship among colour, $H\alpha$ and $H\delta_A$: passive red galaxies with no or weak $H\alpha$ emissions are in the bottom-right corner, blue star-forming galaxies with strong $H\alpha$ emissions are in the top-left corner of the plots. Most AGNs fall into the intermediate-mass region, only the reddest LINERs overlapping with the clump of the completely passive red galaxies. In terms of recent star formation activity, the two interesting galaxy classes are passive blue galaxies and blue LINERs. These galaxy classes are likely to contain many post-starburst galaxies

as they are slightly above the linear distribution of the majority of galaxies.

5.5 Physical parameters from Bayesian analysis

The model fitting method described in Section 5.1 is known not to yield correct values for the physical parameters of the stellar populations building up the galaxies, and more sophisticated techniques are required to estimate the age and the metallicity of the galaxies.

To estimate the physical parameters of the continua of the composites, we entirely adopt the method described in section 4 of Kauffmann et al. (2003b). They used a Bayesian approach to sample the parameter space of the BC03 stellar population synthesis models (Bruzual & Charlot 2003) paying special attention to generate physically plausible star formation histories and constrain other parameters (metallicity, dust content). Based on the stochastic library of star formation histories, they generated a huge library of

model spectra. The star formation histories were characterized by a continuous rate starting at t_{form} and decreasing exponentially with the time-scale γ superimposed by random, short-term bursts.

To estimate physical parameters of a single spectrum, models of the library are individually fitted to the composites, and from the goodness of these fits a probability of the model parameter set is calculated. This method gives not just a single best-fitting model to the spectrum as the NNLS method described in Section 5.1 does, but yields the whole probability distribution of the model parameters and physical properties. As certain physical parameters are very hard to constrain based on optical spectroscopy, it is much better to characterize galaxy spectra with the probability distribution of their parameters instead of a single best-fitting model. The distributions of the physical parameters of our composites are determined using the described method and the modes and medians of the distributions are listed in Table 11. It is clear from the results that our composites cover a very broad range of stellar population ages, specific star formation rates and mass-to-light ratios. On the other hand, metallicities are weakly constrained and close to solar. For the same parameter measurements of individual SDSS galaxies, see Brinchmann et al. (2004).

6 SUMMARY

We have presented a comprehensive atlas of composite spectra of SDSS galaxies. The S/N of the composites can reach as high as $S/N \simeq 132\text{--}4760$ at $\Delta\lambda = 1 \text{ \AA}$ sampling. We have shown that the volume-limited sampling is very important to get physically meaningful composites at very short and very long wavelength. We have argued that our robust averaging method is superior to both traditional averaging and median calculation in terms of computational efficiency and physical interpretation. We have calculated the most important spectral feature indicators of the composites and shown that the parameters of the spectra span almost as wide a range as the spectral parameters of individual galaxies do. This makes the atlas a representative collection of the numerous types of galaxies observed by SDSS.

The composite spectrum atlas is available online at <http://www.vo.elte.hu/compositeatlas>.

ACKNOWLEDGMENTS

This work was supported by the following Hungarian grants: NKTH: Polányi, OTKA-80177 and KCKHA005.

The Project is supported by the European Union and co-financed by the European Social Fund (grant agreement no. TÁMOP 4.2.1/B-09/1/KMR-2010-0003).

REFERENCES

- Abazajian K. N. et al., 2009, *ApJS*, 182, 543
 Abraham R. G. et al., 2004, *AJ*, 127, 2455

- Alongi M., Bertelli G., Bressan A., Chiosi C., Fagotto F., Greggio L., Nasi E., 1993, *A&AS*, 97, 851
 Baldry I. K., Glazebrook K., Brinkmann J., Ivezić Ž., Lupton R. H., Nichol R. C., Szalay A. S., 2004, *ApJ*, 600, 681
 Baldwin J. A., Phillips M. M., Terlevich R., 1981, *PASP*, 93, 5
 Balogh M. L., Morris S. L., Yee H. K. C., Carlberg R. G., Ellingson E., 1999, *ApJ*, 527, 54
 Bell E. F. et al., 2004, *ApJ*, 608, 752
 Blanton M. R. et al., 2003, *AJ*, 125, 2348
 Blanton M. R. et al., 2005, *AJ*, 129, 2562
 Bressan A., Fagotto F., Bertelli G., Chiosi C., 1993, *A&AS*, 100, 647
 Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, *MNRAS*, 351, 1151
 Bruzual A. G., 1983, *ApJ*, 273, 105
 Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
 Budavári T., Szalay A. S., Connolly A. J., Csabai I., Dickinson M., 2000, *AJ*, 120, 1588
 Budavári T., Wild V., Szalay A. S., Dobos L., Yip C.-W., 2009, *MNRAS*, 394, 1496
 Chabrier G., 2003, *PASP*, 115, 763
 Connolly A. J., Szalay A. S., Bershadsky M. A., Kinney A. L., Calzetti D., 1995, *AJ*, 110, 1071
 Connolly A. J., Szalay A. S., 1999, *AJ*, 117, 2052
 Diaz A. I., Terlevich E., Terlevich R., 1989, *MNRAS*, 239, 325
 Dressler A., Oemler A., Jr, Poggianti B. M., Smail I., Trager S., Sackett P. S., Couch W. J., Ellis R. S., 2004, *ApJ*, 617, 867
 Eisenstein D. J. et al., 2001, *AJ*, 122, 2267
 Eisenstein D. J. et al., 2003, *ApJ*, 585, 694
 Fagotto F., Bressan A., Bertelli G., Chiosi C., 1994a, *A&AS*, 104, 365
 Fagotto F., Bressan A., Bertelli G., Chiosi C., 1994b, *A&AS*, 105, 29
 Heckman T. M., 1980, *A&A*, 87, 152
 Hildebrandt H. et al., 2010, *A&A*, 523, A31
 Huchra J. P., Brodie J. P., Caldwell N., Christian C., Schommer R., 1996, *ApJS*, 102, 29
 Kauffmann G. et al., 2003, *MNRAS*, 346, 1055
 Kauffmann G. et al., 2003, *MNRAS*, 341, 33
 Kauffmann G. et al., 2003, *MNRAS*, 341, 54
 Kewley L. J., Dopita M. A., Sutherland R. S., Heisler C. A., Trevena J., 2001, *ApJ*, 556, 121
 Kewley L. J., Groves B., Kauffmann G., Heckman T., 2006, *MNRAS*, 372, 961
 Lee J. H., Lee M. G., Park C., Choi Y.-Y., 2008, *MNRAS*, 389, 1791
 O'Donnell J. E., 1994, *ApJ*, 422, 158
 Renzini A., 2006, *ARA&A*, 44, 141
 Schlegel D. J., Finkbeiner D. P., Davis M., 1998, *ApJ*, 500, 525
 Stoughton C. et al., 2002, *AJ*, 123, 485
 Strateva I. et al., 2001, *AJ*, 122, 1861
 Strauss M. A. et al., 2002, *AJ*, 124, 1810
 Tremonti C. A. et al., 2004, *ApJ*, 613, 898
 Vanden Berk D. E. et al., 2001, *AJ*, 122, 549
 Worthey G., Faber S. M., Gonzalez J. J., Burstein D., 1994, *ApJS*, 94, 687
 Worthey G., Ottaviani D. L., 1997, *ApJS*, 111, 377
 Yip C. W. et al., 2004a, *AJ*, 128, 585
 Yip C. W. et al., 2004b, *AJ*, 128, 2603
 Yip C.-W., Szalay A. S., Wyse R. F. G., Dobos L., Budavári T., Csabai I., 2010, *ApJ*, 709, 780

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