A&A 509, A53 (2010)
DOI: 10.1051/0004-6361/200913168

Astronomy & Astrophysics

Spectral classification of emission-line galaxies from the Sloan Digital Sky Survey

I. An improved classification for high-redshift galaxies

F. Lamareille

Laboratoire d'Astrophysique de Toulouse-Tarbes, Université de Toulouse, CNRS, 14 avenue Edouard Belin, 31400 Toulouse, France
e-mail: flamare@ast.obs-mip.fr

Received 24 August 2009 / Accepted 22 October 2009

ABSTRACT

Aims. We study the spectral classification of emission-line galaxies as starforming galaxies or active galactic nuclei (AGNs). With the high-quality data from the Sloan Digital Sky Survey (SDSS) we define an improved classification to be used for high-redshift galaxies.
Methods. We classify emission-line galaxies of the SDSS according to the latest standard recipe using [OIII]λ5007, [NII]λ6584, [SII]λ6717+6731, Hα, and Hβ emission lines. We obtain four classes: starforming galaxies, Seyfert 2, LINERs, and composites. We then examine where these galaxies fall in the blue diagram used at high redshift (i.e. log([OIII]λ5007/Hβ) vs. log([OII]$\lambda\lambda$3726+3729/Hβ).
Results. We define new improved boundaries in the blue diagram for starforming galaxies, Seyfert 2, LINERs, SF/Sy2, and SF-LIN/comp classes. We maximize the success rate to 99.7% for the detection of starforming galaxies to 86% for the Seyfert 2 (including the SF/Sy2 region) and to 91% for the LINERs. We also minimize the contamination to 16% in the region of starforming galaxies. We cannot reliably separate composites from starforming galaxies and LINERs, but we define an SF-LIN/comp region where most of them fall (64%).

Key words. galaxies: fundamental parameters

1. Introduction

Spectral classification of emission-line galaxies at low redshift is now routinely done with high-quality calibrations. Using a set of five strong emission lines – [OIII]λ5007, [NII]λ6584, [SII]λ6717+6731, Hα, and Hβ – one can reliably distinguish starforming galaxies, Seyfert 2 galaxies, low ionization nuclear emission regions (hereafter LINERs, see Heckman 1980), and composite galaxies with both starforming regions and an active galactic nucleus (hereafter AGN). The equations for such classification have been derived successively by several authors (Baldwin et al. 1981; Veilleux & Osterbrock 1987; Kewley et al. 2001; Kauffmann et al. 2003; Kewley et al. 2006, among others).

At redshifts greater than $z \approx 0.4$, the [NII]λ6584, [SII]λ6717+6731, and Hα emission lines get redshifted out of the wavelength range of all major optical spectroscopic surveys. Therefore, diagnostic diagrams need to be based only on emission lines observed in the blue part of the spectra: [OIII]λ5007, [OII]$\lambda\lambda$3726+3729, and Hβ. Such diagrams, which have been used in the past e.g. by Tresse et al. (1996) or Rola et al. (1997), and which we call the "blue diagram", have been recently studied again by Lamareille et al. (2004). They have derived, from the 2dFGRS data, equations to distinguish starforming galaxies from AGNs. They also show that a region exists in this diagram where both starforming galaxies and AGNs fall (hereafter the "uncertainty region") and thus cannot be distinguished.

With the high-quality Sloan Digital Sky Survey Data Release 7 (hereafter SDSS DR7) data, it is now possible to revisit the blue diagram and to derive new equations that are more compatible with the latest red classification by Kewley et al. (2006) than the ones given in Lamareille et al. (2004). We derive in particular more precise boundaries between the starforming and AGN regions, new boundaries for the regions where AGNs or composites are mixed with starforming galaxies, and new equations to distinguish between Seyfert 2 galaxies and LINERs in the blue diagram.

All spectral classifications, associated numbers, and figures presented in this paper were done with the *JClassif* spectral classification pipeline, which is freely available at the following website: http://www.ast.obs-mip.fr/galaxie/. This paper is organized as follows. We recall the current classification scheme in Sect. 2 and apply it to SDSS DR7 data. Then, we define our new improved classification for high-redshift galaxies in Sect. 3. Finally we show an example of our new classification in Sect. 4.

2. The current classification scheme

2.1. Data selection and the red classification

We use SDSS DR7 emission-line measurements of 868 492 galaxies in the redshift range $0.0 < z \lesssim 0.2$. These data are available online at the following address: http://www.mpa-garching.mpg.de/SDSS/DR7/. The measurements are available for 927 552 different spectra, of which 109 219 spectra are duplicated (twice or more) observations of the same galaxy. We averaged the measurements of duplicated spectra in order to increase the signal-to-noise ratio. Measurements that do not

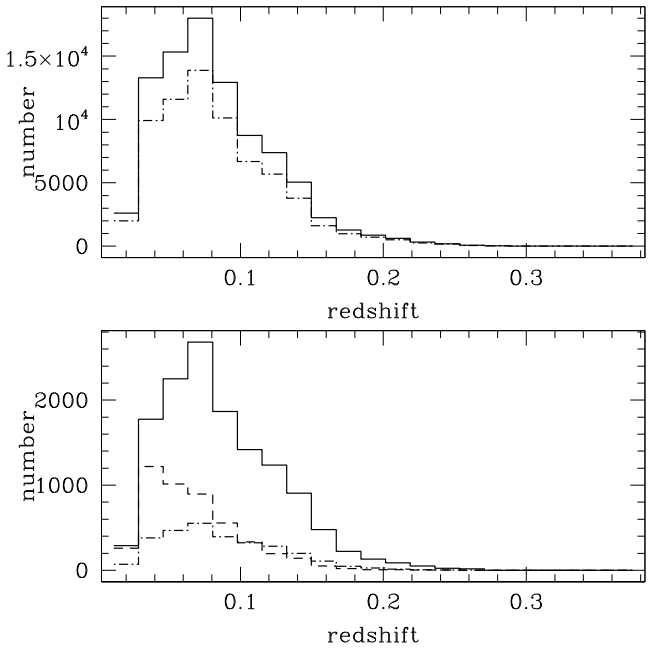

Fig. 1. Redshift histograms of our data sample. *Top*: the solid line is for the whole sample, the dashed-dotted line for starforming galaxies. *Bottom*: the solid line is for composites, the dashed line for LINERs, and the dashed-dotted line for Seyfert 2.

increase the averaged signal-to-noise ratio have been discarded. We select emission-line galaxies with the following criterion: the signal-to-noise ratio in the equivalent width of the emission lines used in our study must be greater than 5. The necessary emission lines for deriving a spectral classification at low redshift are [OIII]λ5007, [NII]λ6584, [SII]$\lambda\lambda$6717+6731, Hα, and Hβ. We also need the [OII]$\lambda\lambda$3726+3729 emission line, which will be used to derive our new high-redshift classification. We ended up with 89 379 emission-line galaxies with the desired minimum signal-to-noise ratio. We sorted these galaxies into four classes according to Kewley et al. (2006) classification scheme. We then end up with the following numbers: 67 778 starforming galaxies, 2 949 Seyfert 2, 4 912 LINERs, and 13 740 composites. Figure 1 shows the redshift histograms of the four classes of emission-line galaxies. We find that the targeted population has the same dependence on redshift for each of the four classes, with a peak around $z \sim 0.07$, except for the LINERs, whose proportion increases at low redshift compared to the other classes. This possible bias has to be noted, even if it does not affect the classification derived in this paper, which is not primarily based on relative proportions between classes.

Figure 2 shows this classification in the standard BPT diagrams. In the left diagram, we note the difference between the old classification of Seyfert 2 and LINERs (Veilleux & Osterbrock 1987) and the new one defined by Kewley et al. (2006). The right diagram cannot be used to distinguish starforming galaxies from composites, which fall in the same region of this diagram. This effect has been clearly explained by Stasińska et al. (2006) with photoionization models: the [OIII]λ5007/Hβ vs. [NII]λ6584/Hα diagnostic diagram is the only one where composites and LINERs clearly separate from starforming galaxies thanks to the so-called "seagull wings". The same applies to LINERs and composites that are not clearly separated in the [OIII]λ5007/Hβ vs. [SII]λ6717+6731/Hα diagnostic diagram. Unlike Kewley et al. (2006), we did not classify as "ambiguous" the composites that fall in the LINERs region, since this diagram does not accurately do it.

2.2. The blue classification

We then derived the blue classification and sort the emission-line galaxies into one of the four classes defined by Lamareille et al. (2004) with the following result : 83 654 secure starforming galaxies, 699 secure AGNs, 3670 candidate starforming galaxies, and 1356 candidate AGNs. Figure 3 shows where the galaxies classified with the red classification fall in the blue diagram. The weaknesses of the blue classification compared to the Kewley et al. (2006) classification scheme applied to the same data are evident in Fig. 3, thanks to the high quality of SDSS's line measurement software and to our signal-to-noise cut. The blue classification is clearly strongly biased against LINERs, which are classified for the majority of them as starforming galaxies or candidate starforming galaxies in the blue diagram. The uncertainty region is actually dominated by AGNs (83%), while a non-negligible number of Seyfert 2 and LINERs (38%) are misclassified as "secure" starforming galaxies. Composites were classified as starforming galaxies by Lamareille et al. (2004) on 2dFRGS data. This has led them to choose an empirical separation that goes more to the right of the blue diagram than necessary.

Nevertheless, the empirical separation defined by Lamareille et al. (2004) from 2dFGRS data does not follow the actual boundary between starforming galaxies and AGNs, as seen with SDSS DR7 data. This separation may then be improved. We may also define a new uncertainty region and a separation between Seyfert 2 and LINERs. The composites cannot be distinguished from starforming galaxies, since they fall in the same region of the blue diagram. As mentioned above, this trend is also present in the log([OIII]λ5007/Hβ) vs. log([SII]$\lambda\lambda$6717+6731/Hα) diagnostic diagram (see Fig. 2 right). Unfortunately, it cannot be avoided at high redshift without the [NII]λ6584 emission-line measurement.

The red classification is insensitive to reddening since it uses ratios of emission lines that are close in wavelength. Conversely, the blue classification uses a line ratio – [OII]$\lambda\lambda$3726+3729/Hβ – involving two lines that are not close in wavelength. Using equivalent widths instead of fluxes, as in this paper, removes direct dependence on reddening. Still, the reddening does not exactly affect the flux of emission lines and the flux of the underlying stellar continuum in the same way. There is thus an indirect dependence of the [OII]$\lambda\lambda$3726+3729/Hβ line ratio with reddening when calculated with equivalent widths. This dependence is greatly reduced compared to ratios of line fluxes.

3. The improved classification

We now define the new improved blue classification of emission-line galaxies. Figure 4 shows how the objects of different classes, according to the red classification, fall in the new blue diagram.

3.1. The new starforming – AGN boundary

We define a new boundary that follows the starforming galaxy region more precisely, as in Kauffmann et al. (2003) compared to the Kewley et al. (2001) boundary in the log([OIII]λ5007/Hβ) vs. log([NII]λ6584/Hα) diagnostic diagram. According to the old blue classification, 87 324 galaxies were classified as secure or candidate starforming galaxies. But 19% of them are actually not

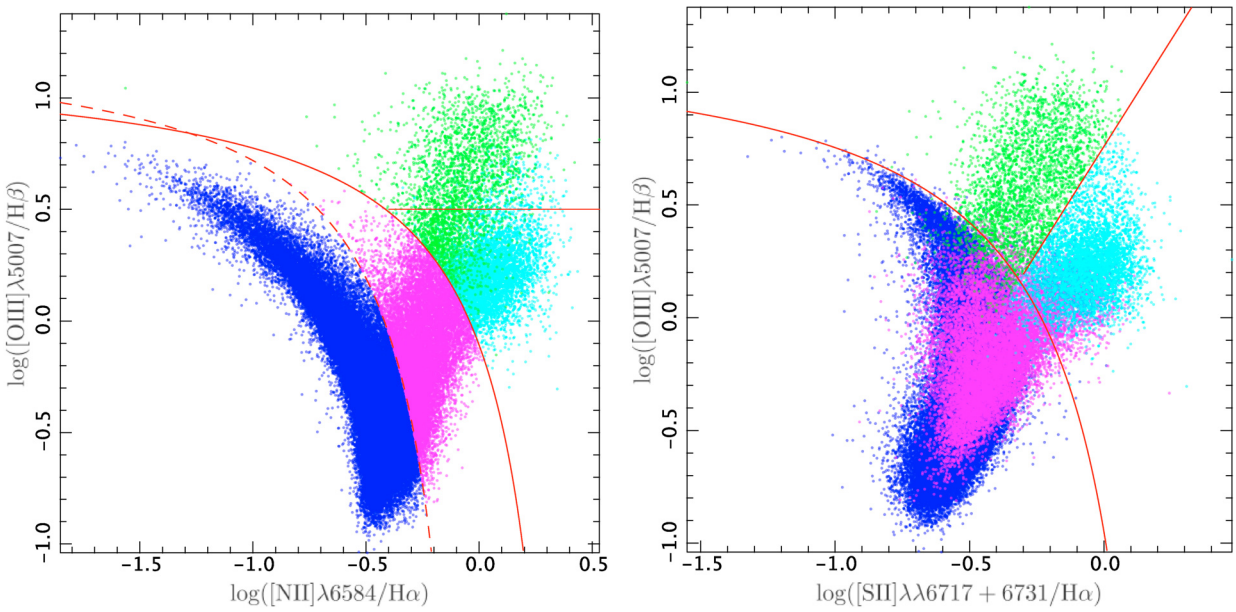

Fig. 2. This is the reference "red" classification of emission-line galaxies at low redshift. The two diagnostic diagrams show the relation between two line ratios: log([OIII]λ5007/Hβ) vs. log([NII]λ6584/Hα) (*left*) and log([OIII]λ5007/Hβ) vs. log([SII]$\lambda\lambda$6717+6731/Hα) (*right*). Starforming galaxies are shown in blue, composites in magenta, Seyfert 2 in green, and LINERs in cyan. The red curves show the empirical or theoretical separations: the solid curve (*left and right*) is Kewley et al. (2001), the dotted curve (*left*) is Kauffmann et al. (2003), the horizontal line (*left*) is Veilleux & Osterbrock (1987), and the solid line (*right*) is Kewley et al. (2006).

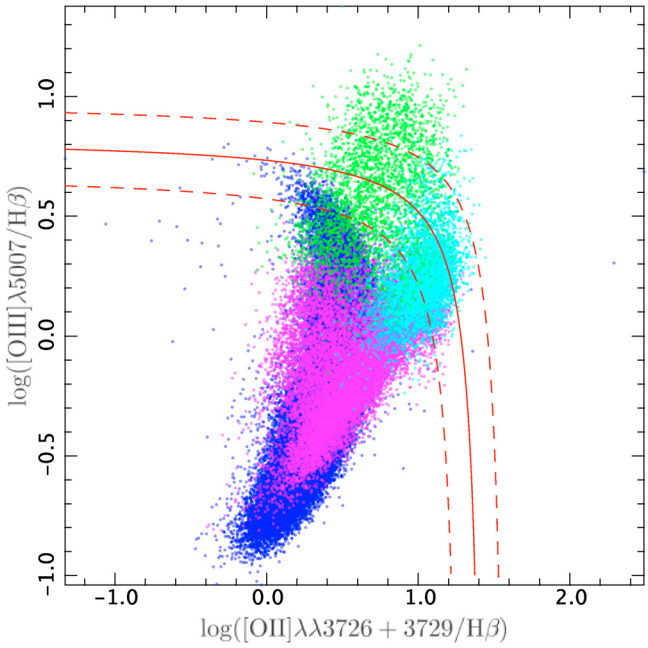

Fig. 3. The "blue" classification of emission-line galaxies at high redshift. The diagnostic diagram shows the relation between two line ratios: log([OIII]λ5007/Hβ) vs. log([OII]$\lambda\lambda$3726+3729/Hβ). According to the red classification (see Fig. 2), starforming galaxies are shown in blue, composites in magenta, Seyfert 2 in green, and LINERs in cyan. The red curves show the empirical separations defined by Lamareille et al. (2004): the solid curve is the separation between starforming galaxies and AGNs, the dashed curves show the uncertainty region.

starforming galaxies according to the red classification. We can reduce this contamination by excluding almost all LINERs and most of the Seyfert 2 with a more conservative separation.

The equation that minimizes the contamination is

$$\log([\mathrm{O\,III}]/\mathrm{H}\beta) = \frac{0.11}{\log([\mathrm{O\,II}]/\mathrm{H}\beta) - 0.92} + 0.85, \quad (1)$$

which corresponds to the solid curve in Fig. 4. Starforming galaxies are below this curve, and AGNs above. The contamination is minimized to 16%. The minimization was done by eye, making certain to maximize the success rate at the same time. Figure 3 shows that a majority of composites, and a number of Seyfert 2 cannot be excluded from the region of starforming galaxies, which explains why the contamination cannot be zero. We checked that 99.7% of the starforming galaxies, according to the red classification, are correctly classified with the new blue classification, which is quite satisfactory.

3.2. The mixed regions

Even if almost all starforming galaxies can be correctly classified using the blue diagram, we know that all starforming galaxies classified according to the blue classification are not actual starforming galaxies. As shown in the right panel of Fig. 4, a non-negligible number of Seyfert 2 galaxies fall into the region of starforming galaxies. From this plot, we easily define the boundary of the region where starforming galaxies become mixed with Seyfert 2:

$$\log([\mathrm{O\,III}]/\mathrm{H}\beta) > 0.3. \quad (2)$$

We call SF/Sy2 all the galaxies above this line. Counting the region of AGNs (as defined by Eq. (1)) and the region of SF/Sy2, 86% of actual AGNs are correctly classified with our new blue classification (59% as Seyfert 2, 26% as SF/Sy2). The region of SF/Sy2 is nevertheless dominated by starforming galaxies (74%). The left panel of Fig. 4 shows that, unlike Seyfert 2, LINERs do not significantly get mixed with starforming galaxies: 91% of them are correctly classified without the need to define an SF/LINER region.

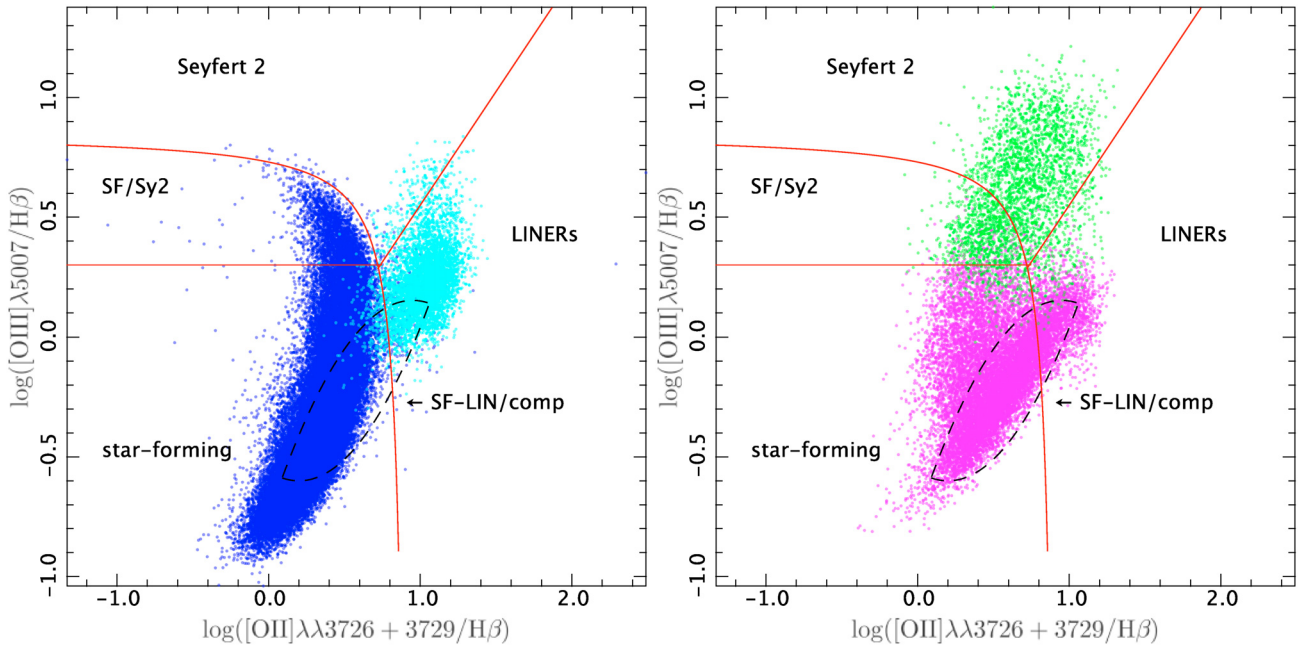

Fig. 4. New improved "blue" classification of emission-line galaxies. The two diagnostic diagrams show the relation between two line ratios: log([OIII]λ5007/Hβ) vs. log([OII]$\lambda\lambda$3726+3729/Hβ). According to the red classification (see Fig. 2), starforming galaxies are shown in blue, LINERs in cyan, composites in magenta, and Seyfert 2 in green. For clarity, the two first classes are shown only in the *left panel*, while the last two classes are shown only in the *right panel*. The red curves show the new empirical separations defined in the text: between starforming galaxies and AGNs (Eq. (1)), between Seyfert 2 and LINERs (Eq. (4)), between starforming galaxies and SF/Sy2 (Eq. (2)). The black dashed curves delimit the region where the majority of composites lie (SF-LIN/comp region, Eq. (3)).

We also need to consider the case of composites that fall in the region of starforming galaxies and LINERs in the blue diagram: 85% of the composites are classified as starforming galaxies and 16% as LINERs in our new classification. We see from the right panel of Fig. 4 that almost all the composites fall by chance below the line defined in Eq. (2). Thus this line can be used to define the region where one should expect to find composites. However, the majority of the composites fall in a much narrower region. We define this region with the two following inequalities:

$$\begin{cases} y \leq -(x-1.0)^2 - 0.1x + 0.25 \\ y \geq (x-0.2)^2 - 0.6 \end{cases} \quad (3)$$

where $y = \log([\mathrm{OIII}]/\mathrm{H}\beta)$, and $x = \log([\mathrm{OII}]/\mathrm{H}\beta)$. We call SF-LIN/comp all the galaxies in this region, which straddles the starforming galaxies and the LINERs. Sixty-four percent of the actual composites fall in our SF-LIN/comp region. The SF-LIN/comp region is composed of 79% of starforming galaxies, 19% of composites, and 2% of LINERs.

3.3. The new Seyfert 2 – LINER boundary

We define an empirical boundary that allows one to distinguish Seyfert 2 from LINERs in the AGNs region of the blue diagram. It is shown as a solid diagonal line in Fig. 4 and follows the equation

$$\log([\mathrm{OIII}]/\mathrm{H}\beta) = 0.95 \times \log([\mathrm{OII}]/\mathrm{H}\beta) - 0.4. \quad (4)$$

This separation minimizes the number of misclassifications between these two classes. Only 8% of the Seyfert 2 and 4% of the LINERs, according to the red classification, are misclassified as LINERs or Seyfert 2 respectively, according to the new blue classification.

4. Example of an application

Our new classification will be useful for building samples of starforming galaxies at high redshift. As an example, we show updated results obtained for VVDS data with our new classification in Fig. 5. We refer the reader to Lamareille et al. (2009) for details. The left panel shows that we now obtain less starforming galaxies than with the previous classification scheme. In the right panel, this leads to slightly different estimates of the mass-metallicity relation in the $0.5 < z < 0.6$ and $0.6 < z < 0.8$ redshift ranges. With our improved classification, we now confirm and even strengthen the conclusion of Lamareille et al. (2009) that the metallicity evolution of starforming galaxies is less significant as a function of redshift for low-mass galaxies than for high-mass galaxies.

For galaxies of masses $\sim 10^9\ M_\odot$, the metallicity evolution is 0.06 dex and 0.10 dex lower in the two above mentioned redshift ranges than the results obtained with the old classification of starforming galaxies. The difference between the old and the new classification is not significant for galaxies of masses $\sim 10^{10.2}\ M_\odot$.

5. Conclusion

Table 1 summarizes the distribution of the objects in the new blue classification, compared to the reference red classification. Please note that the SF/Sy2 and SF-LIN/comp are already counted in one of the three main classes (i.e. starforming galaxies, Seyfert 2 or LINERs) which may be summed to get the total number of objects. A large majority of starforming galaxies are correctly classified with the new blue classification. But the new blue classification also suffers from a non negligible contamination of the starforming regions by composites, which should be taken into account in studies of starforming galaxies. We note,

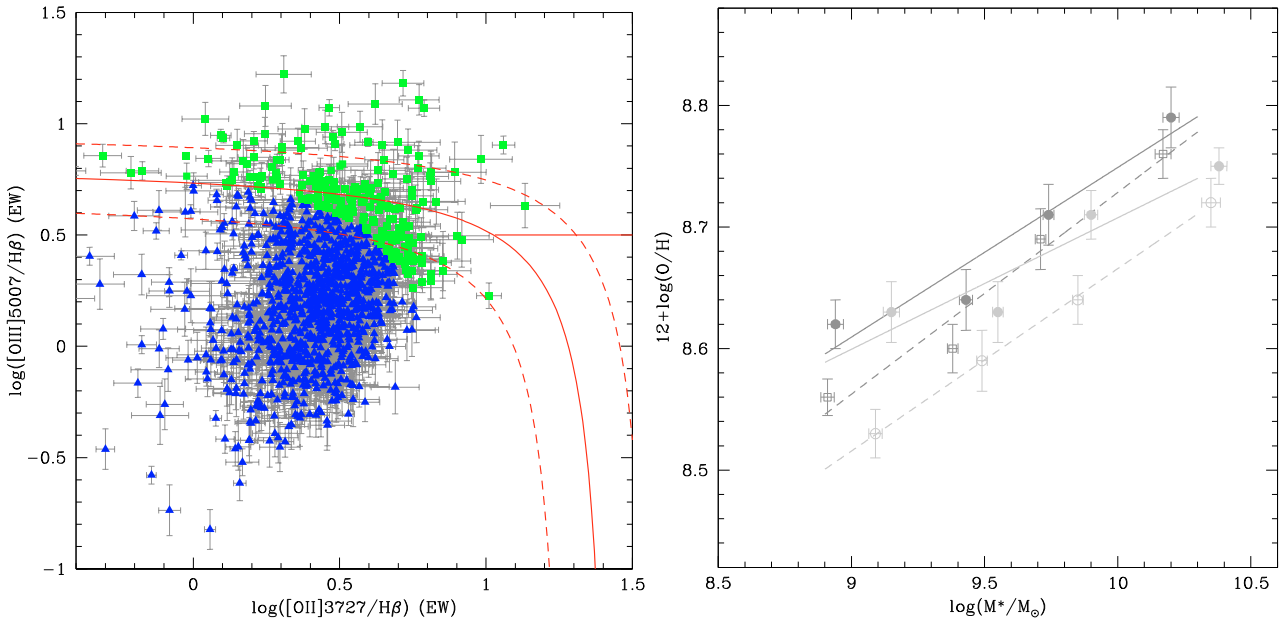

Fig. 5. *Left*: spectral classification of VVDS galaxies (Lamareille et al. 2009). According to our new classification given in Eq. (1), we classify these objects as starforming galaxies (blue triangles) and AGNs (green squares). The red curves show the older classification scheme by Lamareille et al. (2004, see also Fig. 3) *Right*: relation between the logarithm of the stellar mass and the gas-phase oxygen abundance of starforming galaxies in the VVDS sample. Dashed lines and open symbols show the old results as given in Lamareille et al. (2009). Solid lines and filled symbols show the results with our new classification. The results are given in the $0.5 < z < 0.6$ (dark gray) and $0.6 < z < 0.8$ (light gray) redshift ranges.

Table 1. The number of galaxies in each class of the reference classification (columns) as a function of each class of the new blue classification (lines).

	Total	SFG	Sey. 2	LINER	Comp.
total	89 379	67 778	2949	4912	13 740
SFG	80 312	67 539	952	270	11 551
Sey. 2	2016	55	1748	187	26
LINER	7051	184	249	4455	2163
SF/Sy2	3603	2668	774	6	155
SF-LIN/comp	47 461	37 669	17	988	8787

SFG: starforming galaxies; Sey. 2: Seyfert 2.

however, that a number of composites, according to the classification of Kauffmann et al. (2003), may actually only be starforming galaxies. Kewley et al. (2001) have shown from theoretical modeling that pure starforming spectra can be expected in the region defined as the composite region by Kauffmann et al. (2003). Stasińska et al. (2006) also show that the composites allow an AGN contribution up to 20%, but this does not mean that this contribution cannot be less than 20%, or even be zero. True composites may only be confirmed from far infrared or X-ray observations.

We define the region of SF-LIN/comp, which contains the majority of actual composites, but is dominated by actual starforming galaxies and LINERs. The composites also contaminate the region of LINERs in our new classification. However, most of the actual LINERs are correctly classified with our new classification.

Finally, the region of Seyfert 2 in our new classification is almost only composed of actual Seyfert 2 with no significant contamination. But about a third of the actual Seyfert 2 are classified as just SF/Sy2 in our new classification, this class is unfortunately dominated by actual starforming galaxies. The DEW classification proposed by Stasińska et al. (2006) actually complements our classification in classifying correctly SF/Sy2 galaxies. This will be discussed in greater detail in the second paper of this series. Our new classification can be used not only to define samples of starforming galaxies (see Sect. 4), but also samples of Seyfert 2 or LINERs (e.g. to compute luminosity functions) in a much more accurate way than the previous blue classification scheme. However, it cannot be used to derive samples of composites since they get mixed with starforming galaxies and LINERs.

Acknowledgements. The author thanks "La Cité de l'Espace" for financial support while this paper was being written. The data used in this paper were produced by a collaboration of researchers (currently or formerly) from the MPA and the JHU. The team is made up of Stéphane Charlot (IAP), Guinevere Kauffmann, and Simon White (MPA), Tim Heckman (JHU), Christy Tremonti (Max-Planck for Astronomy, Heidelberg – formerly JHU), and Jarle Brinchmann (Sterrewach Leiden – formerly MPA). I thank in particular J. Brinchmann for always keeping these data up to date and for his quick answers to my questions. I also thank E. Hache and J. Marocco, whose work is partly at the origin of the idea to write this paper. I finally thank the referee, G. Stasinska, for useful comments and suggestions, and E. Davoust for improving the English. All data presented in this paper were processed with the JClassif software, part of the Galaxie pipeline.

References

Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5
Heckman, T. M. 1980, A&A, 87, 152
Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003, MNRAS, 346, 1055
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., et al. 2006, MNRAS, 372, 961
Lamareille, F., Mouhcine, M., Contini, T., Lewis, I., & Maddox, S. 2004, MNRAS, 350, 396
Lamareille, F., Brinchmann, J., Contini, T., et al. 2009, A&A, 495, 53
Rola, C. S., Terlevich, E., & Terlevich, R. J. 1997, MNRAS, 289, 419
Stasińska, G., Cid Fernandes, R., Mateus, A., Sodré, L., & Asari, N. V. 2006, MNRAS, 371, 972
Tresse, L., Rola, C., Hammer, F., et al. 1996, MNRAS, 281, 847
Veilleux, S., & Osterbrock, D. E. 1987, ApJS, 63, 295