Template-based classification of SDSS-*GALEX* point sources

AYESHA ANJUM[1,*], MOUSUMI DAS[2], JAYANT MURTHY[2], S. B. GUDENNAVAR[1],
RAJESH GOPAL[3] and S. G. BUBBLY[1]

[1]Department of Physics and Electronics, CHRIST (Deemed to be University), Bengaluru 560 029, India.
[2]Indian Institute of Astrophysics, II Block, Kormamangala, Bengaluru 560 034, India.
[3]Department of Physics, CMR Institute of Technology, Bengaluru 560 037, India.
*Corresponding author. E-mail: astro.pray@gmail.com

Abstract. We have classified a sample of 37,492 objects from SDSS into QSOs, galaxies and stars using photometric data over five wave bands (u, g, r, i and z) and UV *GALEX* data over two wave bands (near-UV and far-UV) based on a template fitting method. The advantage of this method of classification is that it does not require any spectroscopic data and hence the objects for which spectroscopic data is not available can also be studied using this technique. In this study, we have found that our method is consistent by spectroscopic methods given that their UV information is available. Our study shows that the UV colours are especially important for separating quasars and stars, as well as spiral and starburst galaxies. Thus it is evident that the UV bands play a crucial role in the classification and characterization of astronomical objects that emit over a wide range of wavelengths, but especially for those that are bright at UV. We have achieved the efficiency of 89% for the QSOs, 63% for the galaxies and 84% for the stars. This classification is also found to be in agreement with the emission line diagnostic diagrams.

1. Introduction

A peculiar problem in modern astronomy is the selection of objects of interest from large surveys with sources of many different types. Ideally, this would be done through spectroscopic surveys but these require exorbitant amounts of time. As a result, there have been many attempts to use broad band photometric magnitudes and spectroscopic observations to classify sources, particularly extragalactic sources which include active galactic nuclei (AGNs), starburst galaxies and normal galaxies (Croom *et al.* 2001; Ivezic *et al.* 2002; Rowan-Robinson *et al.* 2005; Yan *et al.* 2013; Chung *et al.* 2014; Monroe *et al.* 2016). Previous studies have shown that the application of near and far ultraviolet (NUV/FUV) photometry is very important for the classification of sources, especially for the separation of stars from quasars (Bovy *et al.* 2011; DiPompeo *et al.* 2015) at different redshifts. The low redshift quasars are known to be easily classified by their UV-excess emission (Marshall *et al.* 1984). However, for

a more detailed study, it is important to use the UV photometry to group the QSOs differently (Atlee and Gould 2007; Trammell *et al.* 2007; Hutchings and Bianchi 2008; Jimenez *et al.* 2009). There are studies which show that the source of UV-excess need not always be the AGN activity (Tadhunter *et al.* 2002). The selection of quasar candidates based on nonstellar colour is carried out by Richards *et al.* (2002) and the classification using optical IR colour by Richards *et al.* (2015). Colour-based studies have shown that the classification methods involving infrared bands can robustly separate AGN, normal galaxies and stellar sources (Assef 2010; Jarrett *et al.* 2011; Stern *et al.* 2012). Colour–colour diagrams with multiwavelength data have done an excellent job of rejecting stars from a sample of sources (Bianchi *et al.* 2005; Hutchings and Bianchi 1987). Variability studies have also been used to classify sources but require quality observations carried out multiple times. However, combining variability analysis with colour analysis to classify a well-sampled data with sufficient colour cuts would

help yield more confident quasar classification with high completeness and efficiency (Peters *et al.* 2015). A template-based method of classifying extragalactic sources has been studied by Assef (2010) using a set of three galaxies and one AGN template. They are especially constructed to improve photometric redshifts, but the templates are of low resolution. Template-fitting methods like that of *Lephare* (Arnouts *et al.* 1999; Ilbert *et al.* 2006) and Bayesian Photometric Redshift (BPR) (Benitez 2000) use models of galaxies and stars which are often used to estimate photometric redshifts rather than for classification. Empirical methods such as artificial neural networks (Collister and Lahav 2004; Brescia *et al.* 2015), boosted decision trees, for example, ArborZ (Gerdes *et al.* 2010), regression trees or random forests (Carliles *et al.* 2010; Carrasco and Brunner 2013), polynomial mapping (Budavari *et al.* 2005; Li and Yee 2012) are also some of the efficient methods for redshift estimation but prior knowledge of spectroscopic redshifts or assumptions of the input parameters is a requirement. Bayesian Markov Chain Monte Carlo method has been employed to model SEDs of AGN and classify them into Type 1 and Type 2 AGNs with an efficiency of 80% and 77%, respectively (Calistro *et al.* 2016). Machine-learning (Abraham *et al.* 2012), support vector machines (Wadadekar 2005) and combined supervised and unsupervised hybrid techniques (Fadely *et al.* 2012; Kim *et al.* 2015) have shown promise but require well-sampled training sets.

In this work, we report a template-based method of classifying point sources using multiwavelength photometric magnitudes obtained from the Sloan Digital Sky Survey (SDSS; York *et al.* (2000)) and the Galaxy Evolution Explorer (*GALEX*; Martin *et al.* 2005) with coverage from the far ultraviolet (FUV: 1516 Å) to the optical (z: 8931 Å). The present method differs from previous methods in the sense that it targets a sample of objects for classification without the use of spectroscopy and stringent colour cuts. Also, without the help of spectroscopy and the magnitude cuts, we have achieved efficiency comparable to the machine learning approaches carried out using spectroscopic data.

Nevertheless, we have separately used spectroscopic data only to compare our results with SDSS classification and validate our method. We would like to stress on the fact that this spectroscopic data is not part of the template fitting method. We have found that the classification efficiency is 89% for QSOs, 63% for galaxies and 84% for stars. The method and the results are discussed in the following sections.

2. Data: Sample selection

To implement the template-fitting method, we have employed two samples: first one is the test sample, based on the catalogs constructed by experts. The second one is the true sample based on SDSS spectroscopy. The test sample is used to test our method and validate it by comparing with SDSS classification while the true sample is used for the present work on classification.

2.1 *Test sample*

To test our method of classification, we have employed a test set of objects obtained from the vetted catalogs dedicated to the classification of QSOs, galaxies and stars. We have chosen the test sets keeping in mind their similarities with our actual data and the particularities such as the redshift and the availability of data in all the 7 wavebands. The details of the test set are given in Table 1.

Firstly, in order to test our method of classification, we have obtained data from Paris *et al.* (2014). They have constructed a catalog of QSOs, based on the visual inspection of SDSS spectra, which will add value and confidence in our classification beyond the automated classification of Bolton *et al.* (2012). From this catalog, we have chosen objects with $z < 2$, matching our true data set explained in the next section. This reduces the final number of QSOs in the test set to 25,075. Secondly, as reported by Munoz-Mateos *et al.* (2009), it is difficult to find a vetted sample of galaxies having data in both UV and optical. Having to select only the

Table 1. Details of the vetted sample used as the test data to test our method of classificaiton.

Class	Number of objects	References
QSO	25,075	Paris *et al.* (2014)
Galaxies	3,870	Hernandez-Fernandez *et al.* (2012)
	407	Fotopoulou (2012)
Stars	27,914	Bianchi *et al.* (2011)

point sources among these reduces the sample size even further. However, we found two catalogs which satisfy our requirements in the literature. We have obtained 4,277 galaxies from Hernandez-Fernandez *et al.* (2012) and Fotopoulou (2012). From Hernandez-Fernandez *et al.* (2012), we found 3,870 point sources confirmed as galaxies. There are 407 point sources confirmed to be galaxies in the catalog of Fotopoulou (2012). Finally, for stars, we have used objects from the catalog of Bianchi *et al.* (2011) which is mainly dedicated to the hot stars. As per our requirement, we need objects confirmed to be stars based on 7 magnitudes taken together from UV and optical regimes. So, this catalog fits our requirement and it contains 27,914 objects confirmed to be stars. We would like to stress on the fact that this data is not a part of classification effort, but only used for the sake of comparison and validation of the method.

2.2 *True sample*

Budavari *et al.* (2009) have cross-matched sources using positions from SDSS data release 8 (SDSS-DR8; Aihara *et al.* 2011) and from *GALEX* general release 6 (*GALEX*-GR6; Morrissey *et al.* 2005). We have selected those sources from Budavari *et al.* (2009) which have been classified spectroscopically into QSOs, galaxies and stars by SDSS DR12 (Bolton *et al.* 2012). Our initial sample consisted of 3,628,720 sources with coverage in the two *GALEX* bands (FUV: 1516 Å; NUV: 2267 Å) and the five SDSS bands (*u*: 3551 Å; *g*: 4686 Å; *r*: 6165 Å; *i*: 7481 Å and *z*: 8931 Å). Since Budavari *et al.* (2009) only looked at spatial coincidence there were multiple *GALEX* sources for each SDSS source and vice versa. We discarded all of these to leave 1,656,728 unique point sources of which 70,653 were observed spectroscopically by SDSS. These were classified into 43,947 QSOs, 3,374 galaxies and 23,332 stars by Bolton *et al.* (2012). However, since we wanted to use both FUV and NUV wavebands, we could include only objects that had redshifts less than 2 (Fig. 1). This reduced our final sample size to 37,492 QSOs, 3,374 galaxies and 23,332 stars.

In order to carry out the emission line diagnostics, we have used the emission line spectra from the Skyserver of SDSS. We have chosen objects that have reliable emission line fluxes for Hα, Hβ, [O III], [N II], [S II] and [O I]. There are 14 objects classified as spiral galaxies from this method, for which all the required emission lines are available and similarly 119 starburst galaxies and 5 active galactic nuclei.
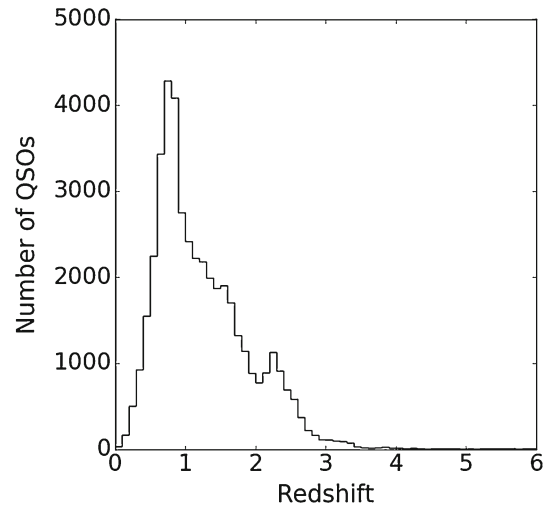

Figure 1. Histogram showing redshifts of objects classified as QSOs by SDSS.

3. Method

We follow Preethi *et al.* (2014) in fitting a set of templates to the 7 photometric bands of *GALEX* and SDSS and selecting models that had the lowest χ^2 value. The models are built using different templates (discussed below) with redshift and extinction as free parameters. This method can be easily extended to different classes of objects by simply adding the corresponding templates to the grid with no changes in the basic algorithm structure. The goal of this study is to fit the templates to each of the 37,492 QSOs, 3,374 galaxies and 23,332 stars as identified by Bolton *et al.* (2012) and compare the results with the SDSS classifications. The major advantage of our template fitting procedure is that we add the two UV bands which remove much of the possible degeneracies in the classification based on only optical bands or in the colour–colour method.

We have chosen templates for each of the different classes of objects (stars, galaxies and QSOs) and these are described below. Since we do not know *a priori* the amount of extinction within the Milky Way and the redshifts, we have left that as a free parameter with the shape given by the Milky Way extinction curve ($R = 3.1$) from Weingartner and Draine (2001). The final grid of photometric values is assembled by convolving the templates with different amounts of Milky Way reddening (Weingartner and Draine 2001) and with the calibration curves for *GALEX* (Morrissey *et al.* 2005) and SDSS (Gunn *et al.* 1998).

(1) Templates for main sequence and giant stars were taken from Castelli and Kurucz (2004). They

Table 2. Details of classification employed on the test data set.

Test class	Total number in test set	Template class			Efficiency (%)
		QSO	Galaxies	Stars	
QSO	25,075	21,795	2,122	1,158	87
Galaxies	4,277	636	3,203	438	75
Stars	27,914	4,298	730	22,886	82

have constructed stellar templates for a range of temperature, surface gravity and metallicity. We have used the recommended model for each spectral type from Castelli and Kurucz (2004) having metallicity ratio same as that of the Sun ([*M/H*] = 0).

(2) White dwarfs templates were taken from Bohlin *et al.* (1995) and comprise observations of four white dwarfs (G191-B2B, GD 71, GD 153 and HZ 43).

(3) Normal galaxy templates were taken from the spectral atlas by Kinney & Calzetti[1]. There are 5 templates in this atlas each of which was constructed from UV and optical spectra of elliptical, S0, Sa, Sb and Sc galaxies, respectively (Kinney *et al.* 1996)

(4) Starburst galaxies templates were taken from Calzetti *et al.* (1994) with 6 templates for each of the types, Starburst 1 to 6. These were made from a sample of multiwavelength observations of 39 starburst galaxies.

(5) We used thirteen different templates to fit AGNs and QSOs. Ten of these were from the SWIRE library (Polletta *et al.* 2007): three SEDs for AGNs (M82, Mrk231 and Arp220); four QSO templates (QSO1, QSO2, TQSO1, BQSO1) with a separate template for heavily reddened Type 2 QSOs; two Seyfert templates (Seyfert 1.8 and Seyfert 2). We added two QSO composite spectra from Francis *et al.* (1991) and Vanden Berk *et al.* (2001) and one template for Seyfert 1 galaxies from the AGN Atlas[2]

We created models corresponding to each of the templates explained above and ran our template fitting method for the QSO sample. We selected models that had the lowest χ^2 values as the best fitting model and fixed the class of the object corresponding to that model.

In order to check the influence of other templates on the models, we repeated the template fitting procedure, but this time, excluding templates belonging to different classes, taking one template at a time. The details are given in section 4.6

4. Results and discussion

4.1 *Test sample*

As described in section 2, we have used a sample of 25,075 QSOs, among which 21,795 objects are correctly classified as QSOs, 2,122 objects are classified as galaxies and 1,158 objects are classified as stars. Among 4,277 galaxies, 3,203 are correctly classified as galaxies whereas, 636 and 438 objects are classified as QSOs and stars, respectively. Finally, from a total of 27,914 stars, 22,886 objects are classified as stars and the rest 4,298 and 730 are classified as QSOs and galaxies, respectively. The efficiency of classification is given in Table 2. The efficiency of classification is defined as the ratio of the number of actual QSOs present in a sample of objects chosen as likely QSOs.

QSOs and AGNs are defined based on their nuclear activity. With spectroscopy, one can probe deeper into the nuclear activity which promotes more efficient classification. For example, a point source which is photometrically classified as a galaxy can be further subclassified as an active galaxy which is hosting an AGN only using spectroscopy. As we have completely relied on photometry, we have missed out the finer details of the signatures of AGN which are available only in the spectroscopic data. This is one of the reasons for the discrepancy between the well-vetted catalogs' spectroscopic classification and our template-based-photometric classification. Another reason for the discrepancy could be the method of classification. For example, Paris *et al.* (2014) have classified objects based on more rigourous emission line fitting, which again requires spectroscopy.

---

[1] http://www.stsci.edu/hst/observatory/crds/cdbs_kc96.html.

[2] ftp://ftp.stsci.edu/cdbs/grid/agn/Seyfert1_template.ascii.

Table 3. Details of classification employed on the photometric data set.

SDSS class	Total number in SDSS class	Template class			Efficiency (%)
		QSO	Galaxies	Stars	
QSO	37,492	33,213	3,563	716	89
Galaxies	3,374	356	2,114	904	63
Stars	23,332	1,906	1,798	19,628	84
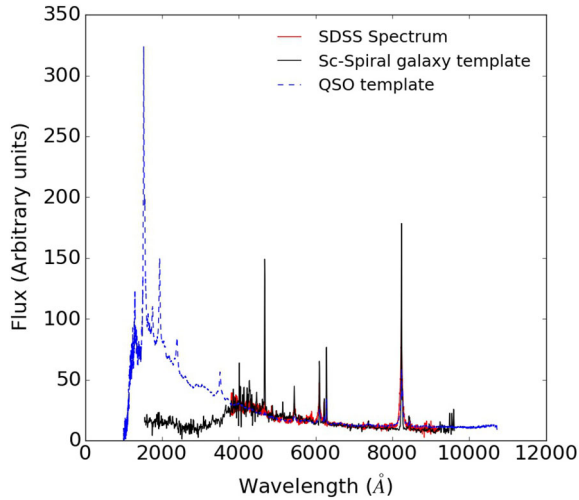
Figure 2. The Sc-type galaxy template from Kinney *et al.* (1996) and the QSO template from Vanden Berk *et al.* (2001) super-imposed on the SDSS spectrum of an SDSS QSO (SDSS J143117.45+492630.2) with a spectroscopic redshift of 0.255, classified as a Sc-type spiral galaxy by our template fitting method.
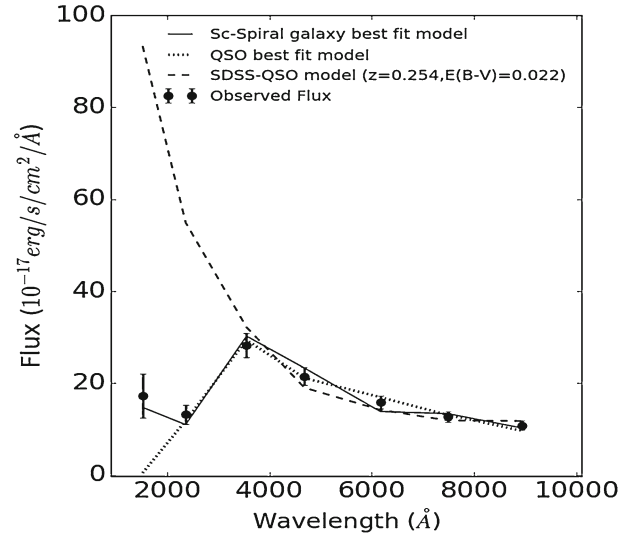
Figure 3. A sample of best-fitting model of an SDSS QSO (SDSS 143117.45+492630.2) classified as Sc-type spiral galaxy by our template fitting method, using UV and optical bands. The QSO model is from Vanden Berk *et al.* (2001) and the spiral galaxy model is from Kinney *et al.* (1996).

4.2 *True sample*

As mentioned earlier, there are 37,492 objects classified as QSOs according to SDSS method. Out of these, 33,213 (89%) are classified as QSOs/AGNs, 3,563 (9%) as galaxies and the rest 716 (2%) as stars. Out of 3,374 galaxies, 2,114 (63%) are classified as galaxies, the rest 356 (10%) and 904 (27%) objects are classified as QSOs and stars, respectively. Similarly, out of 23,332 stars, 19,628 (84%) objects are classified as stars, 1,906 (8%) objects are classified as QSOs and the rest 1,798 (8%) objects are classified as galaxies. The details of the classification is given in the Table 3.

As a next step, we have chosen QSOs to understand the classification better and to carry out detailed analysis. In most of the cases where objects have been misclassified, the discrepancy was because the models differ only in the two UV bands, leading to a degeneracy if only the five SDSS bands are used. This is illustrated in Fig. 2, where we have superimposed the templates for an Sc-type spiral galaxy and a QSO template for SDSS J143117.45+492630.2. This source is identified by SDSS as an unreddened QSO with a redshift of 0.254. The spectra are identical in the optical but the galaxy has a much lower UV flux than expected for the QSO best fit model. This upturn in UV wavebands could be due to the AGN activity or any other factors explained in Tadhunter *et al.* (2002). We have plotted the SEDs in Fig. 3 (dashed line) with both the QSO (dotted line) and Sc spiral galaxy model (solid line). In the optical band, all three fits work well. But we find that the UV flux is too high for the un-reddened QSO model. The best-fit QSO model has more extinction ($E(B - V) = 0.11$) and a higher redshift (1.4) but then has no FUV flux. Rather the observed SED is well-fit by an Sc-spiral galaxy with a redshift of 0.24. Observing the spectra of this object, we find that this object is a spiral galaxy hosting an AGN at its center.
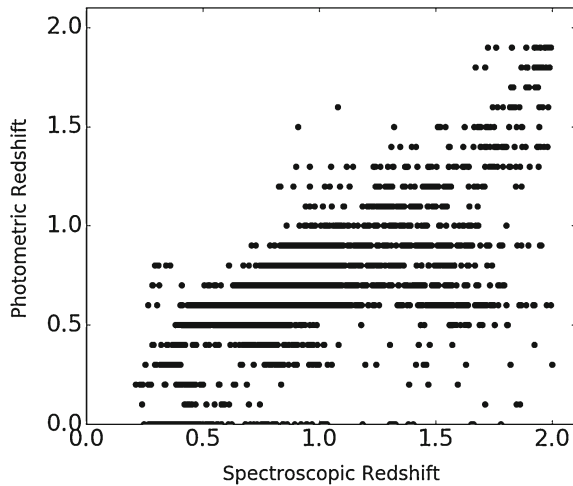
Figure 4. A scatter plot showing the comparison of robust spectroscopic redshifts of the objects and the photometric redshifts obtained from our template fitting procedure.
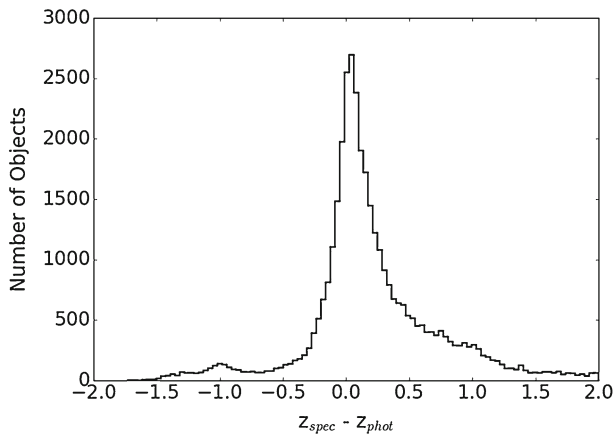

Figure 5. A histogram showing the difference in spectroscopic redshifts and photometric redshifts of the objects.

Since we have taken redshift as one of the parameters, we compared photometric redshifts of the objects under study with that of the spectroscopic redshifts. The comparison between the two are given in Fig. 4 and the histogram showing the difference in spectroscopic redshifts and photometric redshifts of the objects is given in Fig. 5. As redshift estimation is not our primary goal, during the model fitting process, we have incremented the value of redshift in steps of 0.1. So, in Fig. 4, the redshifts seem to be binned. The qualitative representation of differences in redshift is given in Fig. 4.

4.3 *QSOs and stars*

Colour–colour plots can also be used to classify objects. For example, Trammell *et al.* (2007) and Bianchi *et al.* (2007) have shown that the comparison of broad band
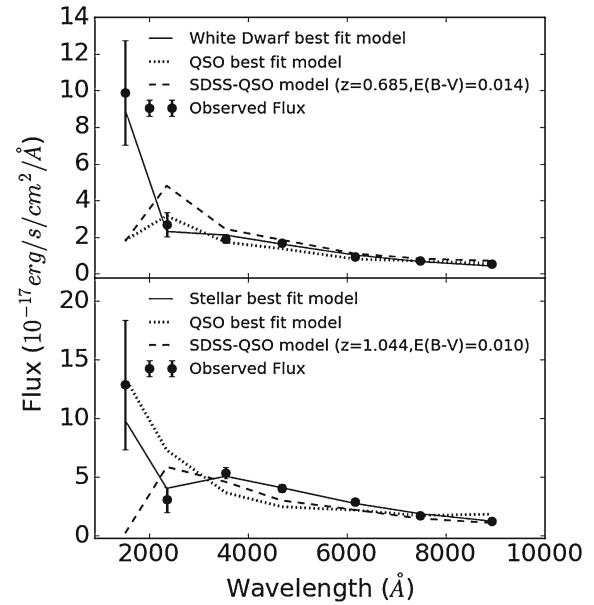
Figure 6. Plots showing the difference between the best-fit model for a QSO, created using a template from Vanden Berk *et al.* (2001) and a white dwarf (SDSS J130401.40+565741.2) and star (SDSS J135110.04+425830.9). The white dwarf model is created using a template from Bohlin *et al.* (1995) and the stellar template is for a B0I star from Castelli and Kurucz (2004).

photometry in the UV and optical is able to separate stars from QSOs. However, the converse is not true as proved by Preethi *et al.* (2014) who showed that it is not possible to find a sample free of extragalactic objects using only colour–colour plots.

Our procedure fits all 7 bands simultaneously and should therefore provide a better discrimination than by the colour–colour plots alone. We have found that the primary discriminator is the FUV flux. We have compared models of a white dwarf and a B0I star with that of a QSO in Fig. 6. If we force the QSO to match the optical bands, we find that the FUV flux is close to zero. We can only match the observed FUV if the source is galactic. We have shown colour–colour plots of the stars and QSOs extracted by our template fitting method in Figures 7 and 8, respectively. We are unable to separate the two sets of objects using only colour–colour plots.

The objects belonging to different classes as per the template fitting method overlap in $(g - r)$ and $(u - g)$ colours. Nevertheless, stars occupy a different locus when UV bands are included compared to when only optical bands are used, as shown in Fig. 7. Few of the objects which have extreme colours have been ignored to improve the clarity of the plot. In the colour–colour method, the QSOs lie within the boundary set by
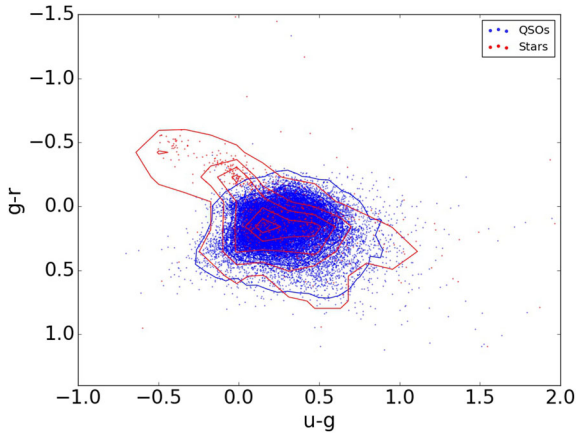
Figure 7. Colour–colour diagrams for the SDSS QSOs classified as QSOs (blue points and contours) and stars (red points and contours) by our template method. All the objects shown in the plots hereafter are the point sources spectroscopically classified as QSOs by SDSS.
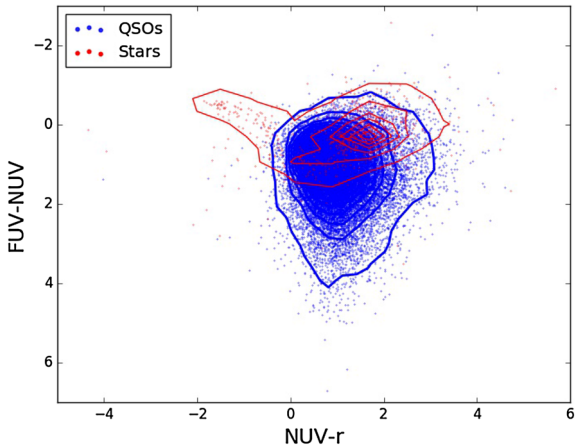

Figure 8. Colour–colour diagram of (FUV − NUV) vs. (NUV − r) for objects classified as QSOs (blue points and contours) and stars (red points and contours) by the template method. Nonetheless, including both the UV bands to separate objects by colours is appreciable compared to only optical colours.
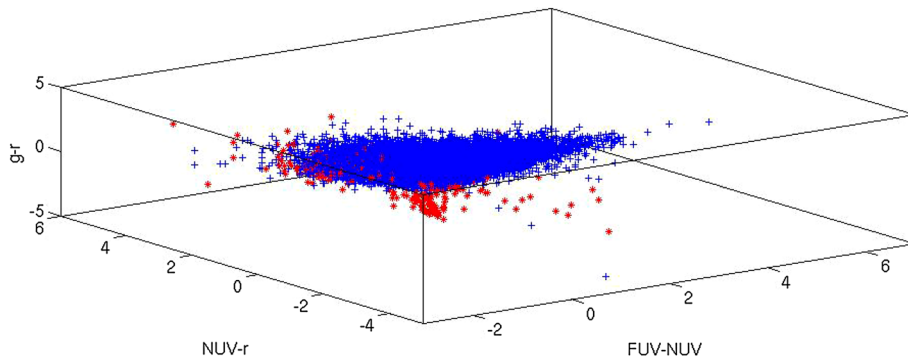
Agueros *et al.* (2005), that is, the QSOs have ($g − r$) colours greater than −0.2.

Bianchi *et al.* (2007) have shown that, including FUV-band in the colour space will help separate stars from extra galactic objects. This is in agreement with Fig. 8, where we have shown (FUV − NUV) vs. (NUV − r) colours of QSOs and stars. Again, in this (FUV − NUV) vs. (NUV − r) plot, objects having extreme colours are also not shown to improve clarity. The problem of slight overlap in this colour space is resolved when all the available bands are considered at once. Though the QSOs are following the boundary limits put forth by the conventional colour–colour methods, it is difficult to avoid overlap between colours of different classes in the colour–colour plots.

Furthermore, in Fig. 9 we have shown a three dimensional plot of (NUV − FUV) vs. ($NUV − r$) vs. ($g − r$) colours with an aim to differentiate between QSOs and stars. We have found that the three dimensional colour plot removes ambiguity between QSOs and stars more clearly than the two-dimensional plot of (FUV − NUV) vs. (NUV − r). Hutchings and Bianchi (1987) have used three-colour index of NUV, g and i vs. r-magnitude to differentiate between stars and QSOs. However, this expression does not seem to do any better when applied to our sample set as our sample consists of objects whose class is not confidently defined.

4.4 *QSOs and active galaxies*

AGNs are powered by the accretion of matter onto a super-massive black hole at the center of the host galaxy and emit radiation over a wide range of wavelengths. Among a wide variety of AGNs (Antonucci 1993; Urry & Padovani 1995; Netzer 2015), QSOs are interpreted as distant and highly luminous counterparts of Seyfert
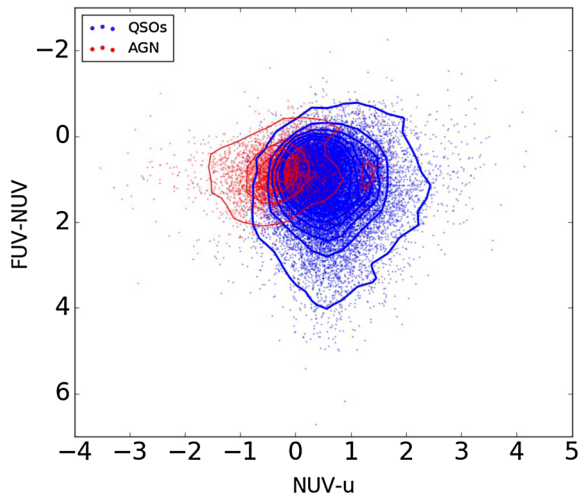

Figure 9. Colours of QSOs (blue) and stars (red) projected in a three-dimensional plot of diagram of (FUV − NUV) vs. (NUV−r) vs. ($g−r$). An object which is classified as a star by the template-based method, has extreme colour (NUV−r ∼ −9) and so it is not shown in the plot to improve clarity.

Figure 10. Colour–colour diagram of (FUV − NUV) vs. (NUV − *u*) for objects classified as QSOs (blue points and contours) and active galaxies (red points and contours) by the template method. To improve the clarity of the plot, objects having (NUV − *u*) colours less than −4 and greater than 5 have not been included.
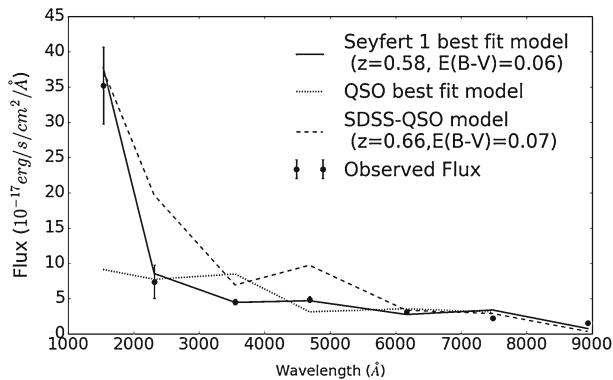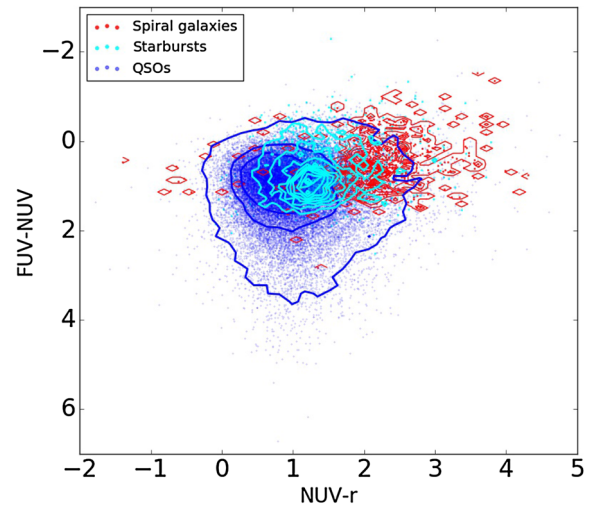

Figure 12. This plot shows the colour space occupied by QSOs (blue points and contours), spiral galaxies (red points and contours) and starbursts (cyan points and contours). Starbursts have occupied colour space in between spiral galaxies and QSOs.
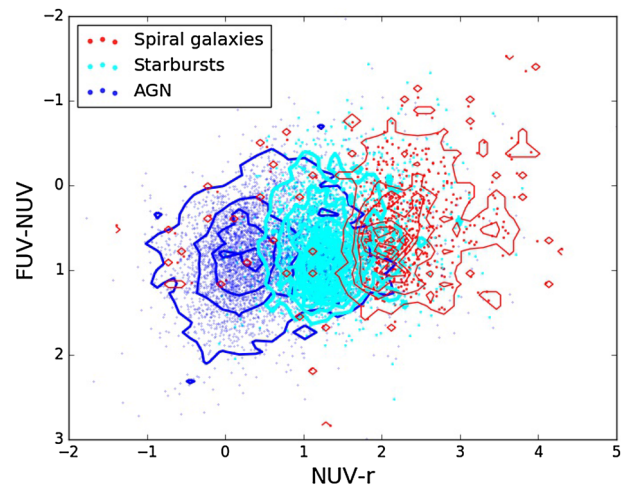

Figure 11. Plots showing the difference between the best-fit model for a QSO, created using a template from Vanden Berk *et al.* (2001) and an object classified as Seyfert 1 galaxy from our template method (SDSS J165705.59+464151.2). The starburst galaxy model is created using a template from Polletta *et al.* (2007).

Figure 13. This plot shows that the active galaxies (blue points and contours) are brighter in NUV than the spiral galaxies (red points and contours). This is possible as the UV luminosity in Seyferts could be enhanced by the AGN at the center. The starbursts (cyan points and contours) have again occupied the colour space in between that of Seyferts and spiral galaxies.

galaxies (Weedman 1977; Bradley 1997) and appear as point sources in optical images. Their luminosity is sometimes so high that they outshine their host galaxies. In such conditions, classification on photometric grounds alone is a challenging task which the SDSS pipeline does not attempt. They use the composite QSO spectrum of Vanden Berk *et al.* (2001), which itself is constructed using the spectra of QSOs and active galaxies selected based on the presence of at least one broad emission-line.

We have added a number of individual templates for active galaxies distinct from the QSO templates and have found 4,296 Seyferts, 3 resembling Arp 220, 20 resembling M82 and 21 resembling the emission line galaxy, Mrk 231. In our template method of classification, AGN refers, in general, to Seyferts and objects such as Arp 220, M 82 and Mrk 231. We have plotted the colours of these objects in Fig. 10 with the

active galaxies lying in the upper left of the region. As an example we have shown 7-band template fit for the object (SDSS J165705.59+464151.2) spectroscopically classified as a QSO, with redshift of 0.66 and extinction of 0.07. According to our template fitting method, this object is best fit to a Seyfert galaxy template yielding similar fitting parameters (redshift = 0.58 and extinction = 0.06). The template fitting is shown in Fig. 11.

4.5 *QSOs, spiral galaxies and starburst galaxies*

In our method of classification of point sources, it was found that 711 out of the 37,492 (2%) SDSS-QSOs have been classified as spiral galaxies and 2 objects as

Figure 14. Plot of (FUV − NUV) vs. (NUV − B) in order to distinguish between active galaxies (blue points and contours) and spiral galaxies (red points and contours). There is a clear separation between the QSOs and spiral galaxies in (NUV − B) colour space.

lenticular galaxies with their colour–colour plot shown in Fig. 12. The bluer (NUV − r) colours of QSOs show that these objects are brighter in ultraviolet than in optical wavebands, in agreement with Bianchi *et al.* (2007).

We then included 2,859 (8%) objects classified as starburst galaxies by our template method. Starbursts show photometric properties intermediate between QSOs and normal galaxies, and it is therefore difficult to distinguish between them purely on the basis of their colours. This class of objects, i.e. spiral and AGN show the most separation in the (FUV − NUV) versus (NUV − r) colour plot. Figure 12 shows that the population of starburst galaxies lies in between AGN and spirals, clearly indicating that starbursts play an important role in triggering the AGN activity in galaxies (Combes 2003). Many starburst galaxies also show AGN activity and there is thought to be a strong AGN-starburst connection (Sanders *et al.* 1988). It is only when we fit all the available photometric bands that we can separate them.

Finally, we classify 4,340 out of the 37,492 (12%) sources as active galaxies. Despite all these objects being unresolved point sources, they are well-separated from normal galaxies when UV colours are added. Starbursts fill in the space between normal galaxies and AGN (Fig. 13). In starbursts, star formation quenching could be either due to the supernova explosions (Vivienne *et al.* 2010) or due to feedback from an AGN (Tremonti *et al.* 2007). The spiral galaxies and Seyfert galaxies occupy different colour space when *B*-magnitudes are used (Fig. 14). The *B*-magnitudes are calculated for quasars with $z \leq 2$ from Jester *et al.* (2005).

Again, for completion, an object, SDSS J151637.38 +140747.5 spectroscopically classified as a QSO with a redshift of 0.84 and extinction of 0.10 is classified
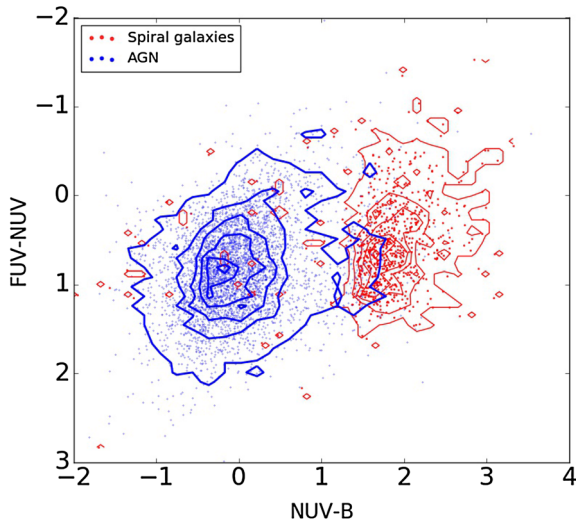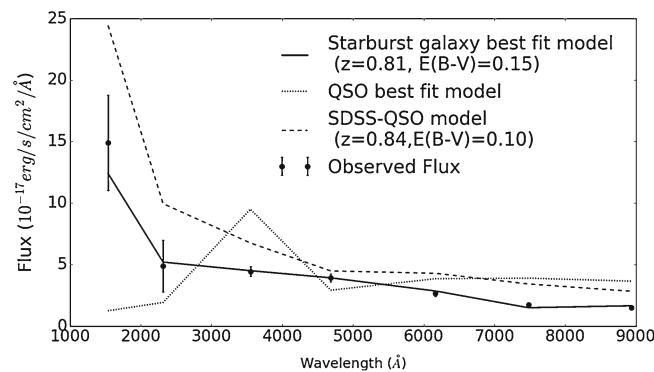

Figure 15. Plots showing the difference between the best-fit model for a QSO, created using a template from Vanden Berk *et al.* (2001) and an object classified as starburst galaxy from our template method (SDSS J151637.37+140747.5). The starburst galaxy model is created using a template from Kinney *et al.* (1996).

as starburst galaxy by our template fitting method. The best fit model has a redshift of 0.81 and an extinction of 0.15 and is classified as a starburst galaxy. The 7-band template fitting is shown in Fig. 15.

4.6 *Confidence level for different magnitudes and redshifts*

Our method of classification is completely dependent on the photometric magnitudes and the classification efficiency (i.e., the fraction of objects classified as QSOs
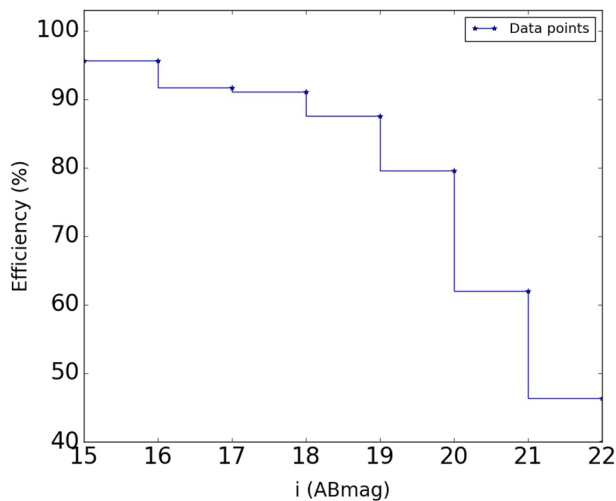
to the total number of objects present in the sample at a given range of magnitudes). The efficiency is found to be good (95%) at brighter magnitudes and decreases (45%) at fainter magnitudes. As an example, we have used i-band magnitudes to show the influence of brightness on the efficiency of classification. In Fig. 16, we have shown the i-band AB magnitudes plotted against the classification efficiency. The plot of redshift versus mean χ^2 values per 0.1 redshift bin is shown in Fig. 17, exhibiting the confidence level at different redshifts. From this, we found that the efficiency of classification is not affected by the redshift. Also, the dependence of efficiency of classification on the redshift is shown in Figures 18 and 19. We have also compared the success rates of Hutchings and Bianchi (1987). They have achieved 90% of match rate for the objects with $i < 20$. We have plotted a cumulative distribution of the best χ^2_{Red} values and the next best χ^2_{Red} values for the whole sample (Fig. 20). We found that the minimum second best χ^2_{Red} is more than 1. Also, almost all the objects have the second best χ^2_{Red} more than 10. This increased the confidence level in finding the best fitting model.

To check the influence of other templates on the classification and selection of models, we started by excluding templates corresponding to active galaxies. A major percentage of the objects among the QSO classes are falling into the composite QSO template by Vanden Berk *et al.* (2001). When this template is excluded, out of 2961 objects, 3 were shifted to spiral galaxies class, 248 were shifted to starburst galaxies group, one was shifted to Mrk231 class and the remaining 2637 were
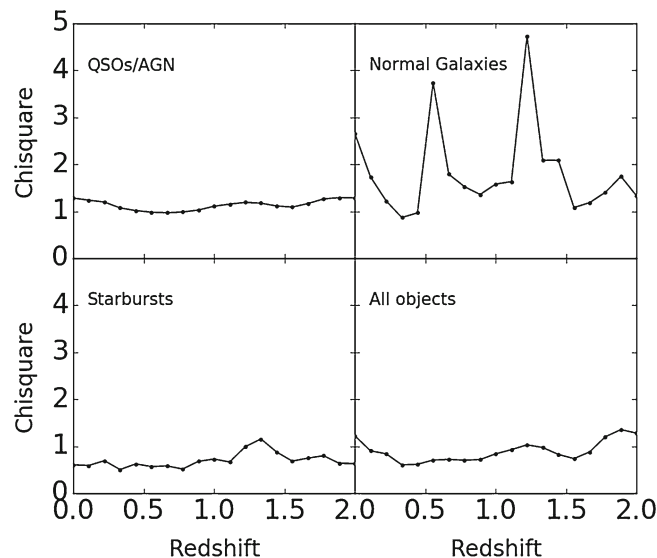

Figure 16. Plot showing the efficiency of classification at different magnitude ranges. It is defined as the ratio of objects classified as QSOs in a certain magnitude range to the total number of objects in that magnitude range.

Figure 17. Plot showing the redshift versus mean value of χ^2 for the objects with a redshift bin of 0.1. This shows confidence of classification and the χ^2 distribution of QSOs/AGNs, normal galaxies and starbursts.
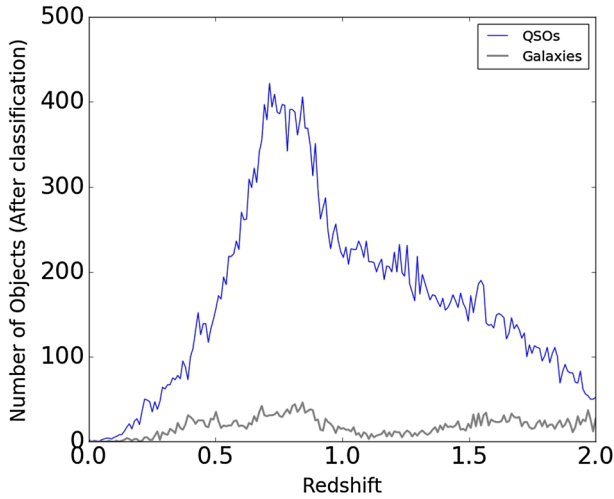
Figure 18. Plot showing the number of objects classified as QSOs (blue) and galaxies (grey) following the present method at different redshifts.
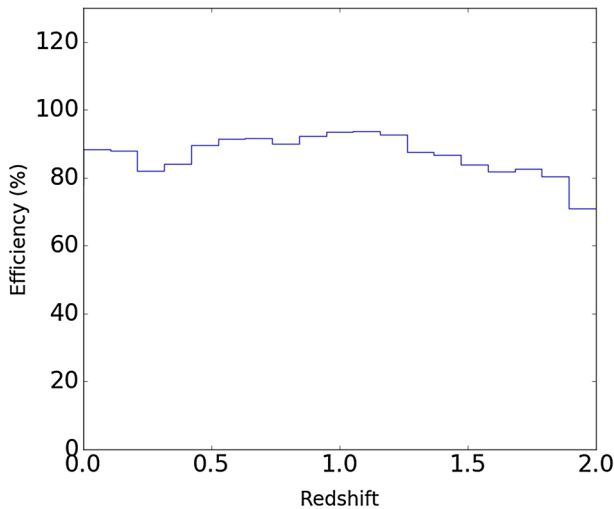

Figure 19. Plot showing efficiency of classification of QSOs at different redshifts with a redshift bin of 0.1.

still being classified as one or the other type of AGN (2431 as QSOs, 165 as Seyferts and 77 were shifted to a torus model) and 36 objects were shifted to stellar models. Then, we excluded templates of Arp220. Out of the 3 resembling Arp220, 2 objects were shifted to starburst class, and the remaining one was shifted to a Seyfert galaxy template. Among 21 that had been classified as Mrk231 type, two were classified as a spiral galaxy, one was classified as a starburst and 10 were shifted to AGN classes, and the rest were classified into Seyferts. Among 21 objects resembling M82, one was shifted to the torus model, 5 were classified as starburst galaxies, 4 as Seyferts and the rest 11 were
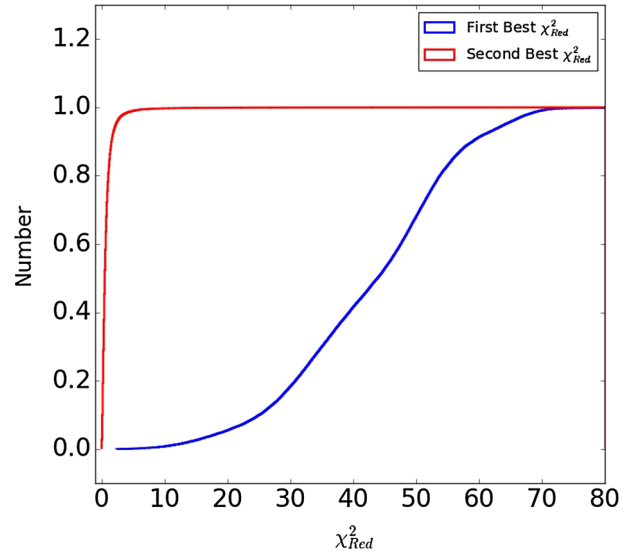

Figure 20. Plot showing cumulative distribution of the first best χ^2_{Red} values and the second best χ^2_{Red} values for the whole sample.

classified as QSOs. Finally, we excluded Seyfert galaxy templates. Out of 4296 objects, 4233 objects were again classified as QSOs, 47 were classified as AGNs, 12 as starbursts, 2 as spiral galaxies and 2 as lenticular galaxies.

5. Emission line diagnostics

The three Baldwin–Phillips–Terlevich (BPT) diagrams are widely in use to classify the emission-line galaxies: [N II]λ6584/Hα vs. [O III]λ5007/Hβ (hereafter [N II]/Hα vs. [O III]/Hβ), [S II] λ6718/Hα vs. [O III]λ5007/Hβ (hereafter [S II]/Hα vs. [O III]/Hβ) and [O I] λ6302/Hα vs. [O III]λ5007/Hβ (hereafter [O I]/Hα vs. [O III]/Hβ) (e.g. Veilleux & Osterbrock 1987), using the calibrations obtained by Kewley *et al.* (2001), Kauffmann *et al.* (2003) and Kewley *et al.* (2006). Specifically, these diagrams help us distinguish between normal star-forming galaxies and QSOs/AGNs using emission lines ratios mentioned above (Baldwin *et al.* 1981). These emission line fluxes are calculated by the SDSS pipeline and are made available from the explore tool of SDSS-III Sky Server[3]. We made use of these fluxes to construct emission line diagnostic diagrams for a set of objects which are classified as spiral galaxies and starbursts galaxies as per our template-based method. With an aim to retain
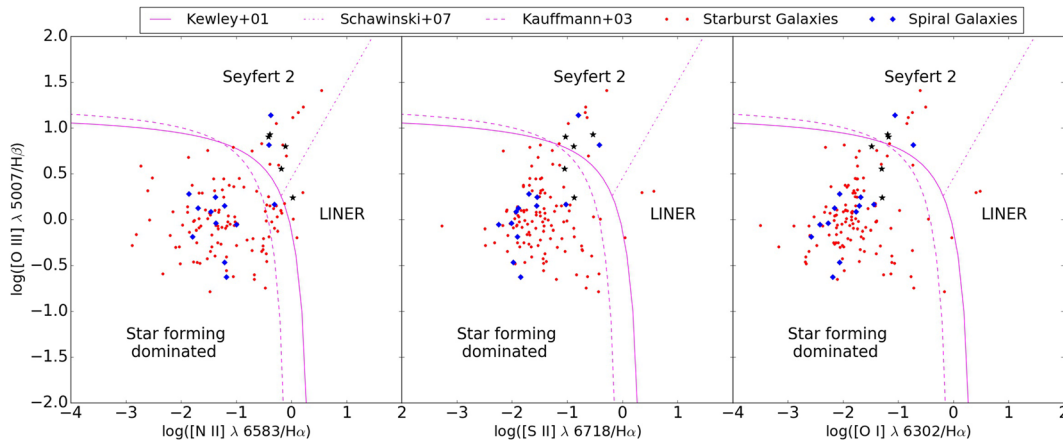
---

[3]http://skyserver.sdss.org/dr12/en/tools/explore/Summary.aspx?.

Figure 21. Plots showing the classification of objects based on the emission line flux ratios. These emission line diagnostic diagrams show that the objects that are classified as QSOs according to the SDSS pipeline are being classified as starburst galaxies (red points), star-forming galaxies (blue diamonds) and AGN (black stars). This is in agreement with our template fitting method of classification.

maximum possible number of objects for comparison, we have not limited our sample-based on their signal to noise ratio.

Starburst galaxies with different metallicities and ages are studied using BPT diagrams (Levesque *et al.* 2010). There are 625 objects which are classified as starburst galaxies. Among these, emission line fluxes were available for only 119 objects (red circles; Fig. 21). The remaining objects were not included in the diagnostic diagrams as they were either too weak to be detected or were out of the redshift range required for their detection. According to the BPT-diagrams, a sample of 119 objects that are used for analysis 16 ([N II]/Hα vs. [O III]/Hβ), 10 ([S II]/Hα vs. [O III]/Hβ) and 8 ([O I]/Hα vs. [O III]/Hβ) have indicated the presence of AGN activity and the rest 87%, 92% and 93% have shown starburst activity in all the three respective panels of the diagram.

Similarly, there are 705 objects classified as spiral galaxies by the template-fitting method. But the Hβ, [O III], [O I], Hα, [N II] and [S II] emission line fluxes were available for only 14 objects (blue diamonds; Fig. 21). The rest of the objects were either not detected as they fell outside the redshift limit ($z \sim 0.5$) or the emission line fluxes were unreliable due to a bad model-fit by SDSS. 12 (85%) objects out of these 14 objects are classified as star-forming galaxies, one as a composite type using the [N II]/Hα vs. [O III]/Hβ emission line ratio and two objects as AGNs in both [S II]/Hα vs. [O III]/Hβ and [O I]/Hα vs. [O III]/Hβ panels of the diagram.

Four objects in our sample are classified as Seyfert 2 galaxies, which are obscured AGN (Antonucci &

Miller 1985; Smith *et al.* 2002; Zakamska *et al.* 2003; Brandt & Hasinger 2005; Reyes *et al.* 2008). Emission line fluxes are not available for any of these 4 objects due to the low signal-to-noise ratio of spectra. Emission lines of obscured QSOs show dominant broad line component. Whereas in Type 2 AGN, the nucleus is heavily obscured by the dusty torus which hides the features of the active galactic nuclei (Antonucci 1993; Urry & Padovani 1995) and thus reduces the flux in UV/optical bands. In such Type 2 objects, the narrow line region is exposed. Due to these reasons, majority of our sample from the SDSS consists of broad emission line sources or objects with active star formation (Mullaney *et al.* 2013). Furthermore, as our sample is limited to low redshifts and point sources, significant fraction of objects in our sample is filtered out. 648 objects are classified as Seyfert 1 galaxies and for the above said reasons, we are unable to project all these objects in the BPT diagram (Menzel *et al.* 2016). Detailed studies and differentiation between obscured and unobscured sources are carried out using different methods (Treister *et al.* 2008; DiPompeo *et al.* 2014; Donoso *et al.* 2014). However, we could find five objects which are narrow line AGN (black stars; Fig. 21) among which two objects show AGN activity and the other three are ambiguous in BPT diagrams.

6. Conclusions

(1) We have taken two sample sets: Firstly, the test set containing objects confirmed to be QSOs, galaxies and

stars selected from the well-vetted catalogs. Secondly, the actual photometric data set from SDSS containing objects spectroscopically classified as QSOs, galaxies and stars. Completeness of classification for both the sets is given in the Table 2 and Table 3.

(2) As this work is focused on the identification of QSOs, in order to carry out detailed analysis, we took a sample of 37,492 objects which are spectroscopically classified as QSOs by SDSS. We then, classified them into QSOs, galaxies and stars, by using multiwavelength broad band photometric magnitudes all the way from FUV (1516 Å) to optical (8931 Å). For classification, we followed a template based method which uses templates from the SWIRE template library, the AGN atlas, the Kinney–Calzetti spectral atlas of galaxies, Castelli and Kurucz atlas and the QSO composite spectra.

(3) It is difficult to separate stars and QSOs using only optical colours, because, the colour space of QSOs overlaps with that of the stars when optical–optical colours like $(g - r)$ are plotted versus $(u - g)$. However, including FUV band we obtain a better separation between QSOs and stars as in $(FUV - NUV)$ vs. $(NUV - r)$ colour–colour plot. There is still a slight overlap, but it is not significant. Thus we have found that the stars and QSOs are separated best when all the bands are involved rather than only optical or UV colours are used (Fig. 8).

(4) We have found that the QSOs can be distinguished from the Sc-type spiral galaxies using a colour–colour plot of $(FUV - NUV)$ vs. $(NUV - r)$, which comes out to be even better when their SEDs are compared. The normal galaxies and QSOs are distinguishable in $(NUV - r)$ colour space, where the spiral galaxies have redder $(NUV - r)$ colours than the QSOs.

(5) Similarly, $(FUV - NUV)$ vs. $(NUV - r)$ colour-colour plot helps separate out spiral galaxies and starburst galaxies from QSOs, though there is a slight overlap between starburst galaxies and the other two classes of objects. This can be overcome using template fitting method.

(6) The colour–colour plot of $(FUV - NUV)$ vs. $(NUV - r)$ helps us clearly discriminate between spiral galaxies and Seyfert galaxies with the starbursts lying intermediate between these two classes of objects. Also, spiral galaxies and Seyfert galaxies can be separated well using $(NUV - B)$ colour space.

(7) Our results show that including the UV wavebands to the photometric method of classification improves the efficiency. This method is especially advantageous in cases where spectroscopic data is not available for the sample. Photometric classification is important as it is often more easily available as compared to spectroscopy. Even though the colour–colour method is widely in use, we have found that it is important for photometric classification only when all the five optical bands (u, g, r, i and z) and two UV bands (FUV and NUV) are used.

(8) The spectroscopic information of these objects is used only for comparative study and to estimate efficiency of our classification method. Comparison with spectroscopic information has ensured that, given only photometric data, we can carry out classification with an efficiency of 89%.

(9) We have also classified the objects taken from the dedicated quasar catalog (DR10Q) and found that the classification efficiency has changed by 3% which has added additional confidence to our classification.

(10) The present classification method is compared with the conventional and more effective emission line diagnostic diagrams, from which we see that our classification is efficient to 87% for starburst galaxies, 85% for star-forming galaxies and 40% for the narrow line AGN. This decrease in efficiency for narrow line AGN could be due to their intrinsic properties that could hamper optical observations due to the presence of obscuring torus.

(11) Thus, we argue that the automated fitting algorithm of SDSS is reliable but may not be 100% accurate. Thus, we used the SDSS sample to implement template fitting method of classification for further study. From this followed by the emission line diagnostic diagrams, we manually checked and observed that some of the objects classified as QSOs by SDSS are being classified as star-forming galaxies. These are also dominated by star formation in the BPT diagrams. We have manually checked their spectra and have found traces of broad lines, which force them to qualify for AGN class. Thus, this classification scheme that includes both optical and UV wavebands is more efficient and can clearly differentiate between stars, galaxies and QSOs compared to the one which uses only optical colours, without the use of spectroscopic data for the classification.

References

Abraham S., Philip N. S., Kembavi A., Wadadekar Y. G., Sinha R. 2012, MNRAS, 419, 80

Agueros M. A. *et al.* 2005, AJ, 130, 1022

Aihara H. *et al.* 2011, ApJS, 193, 29

Antonucci R. 1993, ARA&A, 31, 473

Antonucci R. R. J., Miller J. S., 1985, ApJ, 297, 621

Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, MNRAS, 310, 540

Assef R. J. *et al.* 2010, ApJ, 713, 970

Atlee D. W., Gould A., 2007, ApJ, 664, 53

Baldwin J. A., Phillips M. M., Terlevich R. 1981, PASP, 93, 5

Benitez N. 2000, ApJ, 536, 571

Bianchi L. *et al.* 2005, ApJ, 619, L27

Bianchi L. *et al.* 2007, ApJS, 173, 659

Bianchi L. *et al.* 2011, MNRAS, 411, 2770

Bohlin R. C., Colina L., Finley D. S. 1995, AJ, 110, 1316

Bolton A. S. *et al.* 2012, AJ, 144, 144

Bovy J. *et al.* 2011, ApJ, 729, 141

Bradley M. P. 1997, An introduction to Active Galactic Nuclei, Cambridge University Press

Brandt W. N., Hasinger G. 2005, ARA&A, 43, 827

Brescia M., Cavouti S., Longo G., 2015, MNRAS, 450,3893

Budavari T. *et al.* 2005, ApJ, 619, L31

Budavari T. *et al.* 2009, ApJ, 694, 1281

Calistro Rivera G., Lusso E., Hennawi J. F., Hogg D. W. 2016, ApJ, 833, 98

Calzetti D., Kinney A. L., Storchi-Bergmann T. 1994, ApJ, 429, 582

Carliles S., Budavari T., Heinis S., Priebe C., Szalay A. S. 2010, ApJ, 712, 511

Carrasco K. M., Brunner R. 2013, MNRAS, 432, 1483

Castelli F., Kurucz R. L. 2004, preprint (astro-ph/0405087v1)

Chung S. M. *et al.* 2014, ApJ, 790, 54

Collister A., Lahav O. 2004, PASP, 116, 345

Combes F. 2003, ASPC, 290, 411

Croom S. M., Smith R. J., Boyle B. J., Shanks T., Loaring N. S., Miller L., Lewis I. J. 2001, MNRAS, 322, L29

DiPompeo M. A., Myers A. D., Hickox R. C., Geach J. E., Hainline K. N. 2014, MNRAS, 442, 3443

DiPompeo M. A., Bovy J., Myers A. D., Lang D. 2015, MNRAS, 452, 3124

Donoso E., Yan L., Stern D., Assef R. J. 2014, ApJ, 789, 44

Fadely R., Hogg D. W., Willman B. 2012, ApJ, 760, 15

Fotopoulou S. 2012, ApJS, 198, 1

Francis P. J., Hewett P. C., Foltz C. B., Chaffee F. H., Weymann R. J., Morris S. L. 1991, ApJ, 373, 465

Gerdes D. W., Sypniewki A. J., McKay T. A., Hao J., Weis M. R., Wechsler R. H., Busha M. T. 2010, ApJ, 715, 823

Gunn J. E. *et al.* 1998, AJ, 116, 3040

Hernandez-Fernandez J. D., Iglesias-Paramo J., Vilchez J. M. 2012, ApJS, 199, 22

Hutchings J. B., Bianchi L. 2008, PASP, 120, 275

Hutchings J. B., Bianchi L. 2010, AJ, 140, 1987

Ilbert O. *et al.* 2006, A&A, 457, 841

Ivezic Z. *et al.* 2002, AJ, 124, 2364

Jarrett T. H. *et al.* 2011, ApJ, 735, 112

Jester S. *et al.* 2005, AJ, 130, 873

Jimenez R., Spergel D. N., Niemack M. D., Menanteau F., Hughes J. P., Verde L., Kosowsky A. 2009, ApJS, 181, 439

Kauffmann G. *et al.* 2003, MNRAS, 346, 1055

Kewley L. J., Dopita M. A., Sutherland R. S., Heisler C. A., Trevea J. 2001, ApJ, 556, 121

Kewley L. J., Groves B., Kauffmann G., Heckman T. 2006, MNRAS, 372, 961

Kim E. J., Brunner R. J., Carrasco Kind M. 2015, MNRAS, 453, 507

Kinney A. L., Calzetti D., Bohlin R. C., McQuade K., Storchi-Bergmann T., Schmitt H. R. 1996, ApJ, 467, 38

Li I., Yee H. 2012, ApJ, 749, 2

Levesque E. M., Kewley L. J., Larson K. L. 2010, AJ, 139, 712

Martin D. C. *et al.* 2005, ApJ, 619, L1

Marshall H. L., Avni Y., Braccesi A., Huchra J. P., Tananbaum H., Zamorani G., Zitelli V. 1984, ApJ, 283, 50

Menzel M. L. *et al.* 2016, MNRAS, 457, 110

Monroe T. R., Prochaska J. X., Tejos N., Worseck G., Hennawi J. F., Schmidt T., Tumlinson J., Shen Y. 2016, AJ, 152, 25

Morrissey P. *et al.* 2005, ApJ, 619, L7

Mullaney J. R., Alexander D. M., Fine S., Goulding A. D., Harrison C. M., Hickox R. C. 2013, MNRAS, 433, 622

Munoz-Mateos J. C. *et al.* 2009, ApJ, 703, 1569

Netzer H. 2015, ARA&A, 53, 365

Paris I. *et al.* 2014, A&A, 563, 54

Peters C. M. *et al.* 2015, ApJ, 811, 95

Polletta M. *et al.* 2007, ApJ, 663, 81

Preethi K., Gudennavar S. B., Bubbly S. G., Murthy J., Brosch N. 2014, MNRAS, 437, 771

Reyes R. *et al.* 2008, AJ, 136, 2373

Rowan-Robinson M. *et al.* 2005, AJ, 129, 1183

Richards G. T. *et al.* 2002, AJ, 123, 2945

Richards G. T. *et al.* 2015, ApJS, 219, 39

Sanders D. B., Soifer B. T., Elias J. H., Madore B. F., Matthews K., Neugebauer G., Scoville N. Z. 1988, ApJ, 325, 74S

Smith J. A. *et al.* 2002, AJ, 123, 2121

Stern D. *et al.* 2012, ApJ, 753, 30

Tadhunter C., Dickson R., Robinson T. G., Wills K., Martin V. M., Hughes M. 2002, MNRAS, 330, 977

Trammell G. B., Vanden Berk D. E., Schneider D. P., Richards G. T., Hall P. B., Anderson S. F., Brinkmann J. 2007, AJ, 133, 1780

Treister E., Krolik J. H., Dullemond C. 2008, ApJ, 679, 140

Tremonti C. A., Moustakas J., Diamond A. M. 2007, ApJ, 663, L77

Urry C. M., Padovani P. 1995, PASP, 107, 803

Vanden Berk D. E. *et al.* 2001, AJ, 122, 549

Veilleux S., Osterbrock D. E. 1987, ApJS, 63, 295

Wadadekar Y. 2005, PASP, 117, 79

Weedman D. W. 1977, ARA&A, 15, 69

Weingartner J. C., Draine B. T. 2001, ApJ, 548, 296

Yan L. *et al.* 2013, AJ, 145, 55

York D. G. *et al.* 2000, AJ, 120, 1579

Zakamska N. L. *et al.* 2003, AJ, 126, 2125