

# Artificial neural networks for quasar selection and photometric redshift determination

Ch. Yèche<sup>1</sup>, P. Petitjean<sup>2</sup>, J. Rich<sup>1</sup>, E. Aubourg<sup>3</sup>, N. Busca<sup>3</sup>, J.-Ch. Hamilton<sup>3</sup>, J.-M. Le Goff<sup>1</sup>,  
I. Paris<sup>2</sup>, S. Peirani<sup>2</sup>, Ch. Pichon<sup>2</sup>, E. Rollinde<sup>2</sup>, and M. Vargas-Magaña<sup>3</sup>

<sup>1</sup> CEA, Centre de Saclay, IRFU, 91191 Gif-sur-Yvette, France  
e-mail: christophe.yeche@cea.fr

<sup>2</sup> Université Paris 6 et CNRS, Institut d'Astrophysique de Paris, 98bis Blvd. Arago, 75014 Paris, France

<sup>3</sup> APC, 10 rue Alice Domon et Léonie Duquet, 75205 Paris Cedex 13, France

Received 20 October 2009 / Accepted 22 May 2010

## ABSTRACT

**Context.** Baryonic acoustic oscillations (BAO) and their effects on the matter power spectrum can be studied using the Lyman- $\alpha$  absorption signature of the matter density field along quasar (QSO) lines of sight. A measurement sufficiently accurate to provide useful cosmological constraints requires the observation of  $\sim 10^5$  quasars in the redshift range  $2.2 < z < 3.5$  over  $\sim 8000$  deg<sup>2</sup>. Such a survey is planned by the Baryon Oscillation Spectroscopic Survey (BOSS) project of the Sloan Digital Sky Survey (SDSS-III).

**Aims.** We assess one of the challenges for this project, that of building from five-band imaging data a list of targets that contains the largest number of quasars in the required redshift range. In practice, we perform a stellar rejection of more than two orders of magnitude with a selection efficiency for quasars better than 50% to magnitudes as bright as  $g \sim 22$ .

**Methods.** To obtain an appropriate target list and estimate quasar redshifts, we develop artificial neural networks (ANNs) with a multilayer perceptron architecture. The input variables are photometric measurements, i.e., the object magnitudes and their errors in the five bands (*ugriz*) of the SDSS photometry. The ANN developed for target selection provides a continuous output variable between 0 for non-quasar point-like objects to 1 for quasars. A second ANN estimates the QSO redshift  $z$  using the photometric information.

**Results.** For target selection, we achieve a non-quasar point-like object rejection of 99.6% and 98.5% for a quasar efficiency of, respectively, 50% and 85%, comparable to the performances of traditional methods. The photometric redshift precision is on the order of 0.1 over the region relevant to BAO studies. These statistical methods, developed in the context of the BOSS project, can easily be extended to any quasar selection and/or determination of their photometric redshift.

**Key words.** methods: statistical – quasars: general

## 1. Introduction

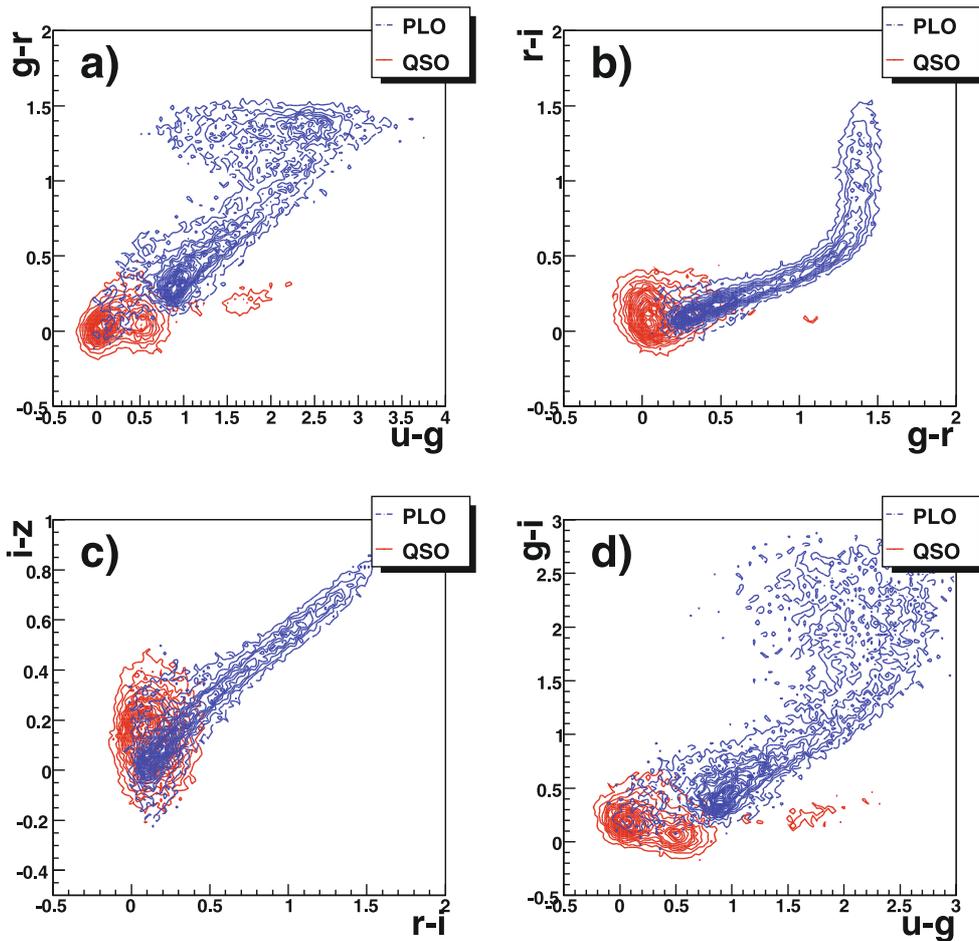
Since the first quasar was discovered (Schmidt 1963), methods have been developed to differentiate these rare objects from other astronomical sources in the sky. In the standard methods, it is assumed that QSOs have point-like morphology. They are then separated from the far more numerous stars by their photometric colors. The UVX selection, e.g., Croom et al. (2001), can be largely complete (>90%) for QSOs with  $0.3 < z < 2.2$  but this completeness drops at higher redshift. The selection purity was increased to 97% for  $g < 21$  by applying kernel density estimation techniques (KDE) to SDSS colors (Richards et al. 2004) and extended to the infrared by Richards et al. (2009a), implying that spectroscopy is not needed to identify a statistical sample of quasars at high Galactic latitudes. This led to the definition of a one-million-QSO catalog (Richards et al. 2009b) down to  $i = 21.3$  from the photometry of SDSS Data Release 6 (Adelman-McCarthy et al. 2008).

Quasar selection at higher redshifts and fainter magnitudes is affected by several difficulties. For example, towards fainter magnitudes, galaxies begin increasingly to contaminate “point-like” photometric catalogs both because of the increasing photometric errors and because of non-negligible contributions from AGNs in certain bands. This is unfortunate because large catalogs of high-redshift QSOs would be useful not only to study the AGN population but also to use the quasars to study the foreground absorbers. In particular, studies of the spatial correlations

of the IGM from the Lyman- $\alpha$  forest and/or metal absorption lines need relatively high target density at high redshift (Petitjean 1997; Nusser & Haehnelt 1999; Pichon et al. 2001; Caucci et al. 2008).

It has been realized that the effects of baryonic acoustic oscillations (BAO) may be detectable in the Lyman- $\alpha$  forest. These oscillations occur in the pre-recombination universe because at that time the baryons, electrons, and photons formed a nearly perfect fluid supporting acoustic waves. In the post-recombination epoch, these waves have been fossilized as an enhanced matter correlation at a distance corresponding to the sonic horizon at recombination, i.e. the distance that a sound wave can travel before stopping at recombination. Equivalently, “wiggles” are seen in the matter power spectrum corresponding to harmonics of the horizon. To date, BAO effects have been seen using galaxies of redshift  $z < 0.4$  to sample the matter density (Eisenstein et al. 2005; Cole et al. 2005; Percival et al. 2009). The Baryon Oscillation Spectroscopic Survey (BOSS) (Schlegel et al. 2009) of the Sloan Digital Sky Survey (SDSS-III) (SDSS-III Coll. 2008) proposes to extend these studies using galaxies of higher redshifts,  $z < 0.9$ . The BOSS project will also study BAO effects in the range  $2.2 < z < 3.5$  using Lyman- $\alpha$  absorption towards high redshift quasars (QSOs) to sample the matter density, as proposed by McDonald & Eisenstein (2007).

The power spectrum has already been measured at  $z \sim 2.5$  using the one-dimensional matter power spectrum derived from quasar spectra (Croft et al. 1999). The observation of BAO



**Fig. 1.** 2D distributions of colors ( $u - g$ ,  $g - r$ ,  $r - i$ ,  $i - z$  and  $g - i$ ) for objects classified as PLO in SDSS photometric catalog (blue lines for contours) and for objects spectroscopically classified as QSO (red solid lines for contours). The PSF magnitudes ( $ugriz$ ) have been corrected for Galactic extinction according to the model of Schlegel et al. (1998).

effects will require a full three-dimensional sampling of the matter density, requiring a much higher number of quasars than previously available. BOSS aims to study around 100 000 QSOs over 8000 square degrees. The requirement that the Lyman- $\alpha$  absorption fall in the range of the BOSS spectrograph requires that the quasars be in the redshift range  $2.2 < z < 3.5$ .

The quasars to be targeted must be chosen only using available photometric information, mostly from the SDSS-I point-source catalog. The target selection method must be able to reject the non-quasar point-like objects (PLOs; mainly stars) by more than two orders of magnitude with a selection efficiency of QSOs better than 50%. The BOSS project aims to achieve a high density of  $z > 2.2$  QSOs ( $\sim 20$  QSOs per square degree) and therefore requires the selection to be extended to  $g \sim 22$ .

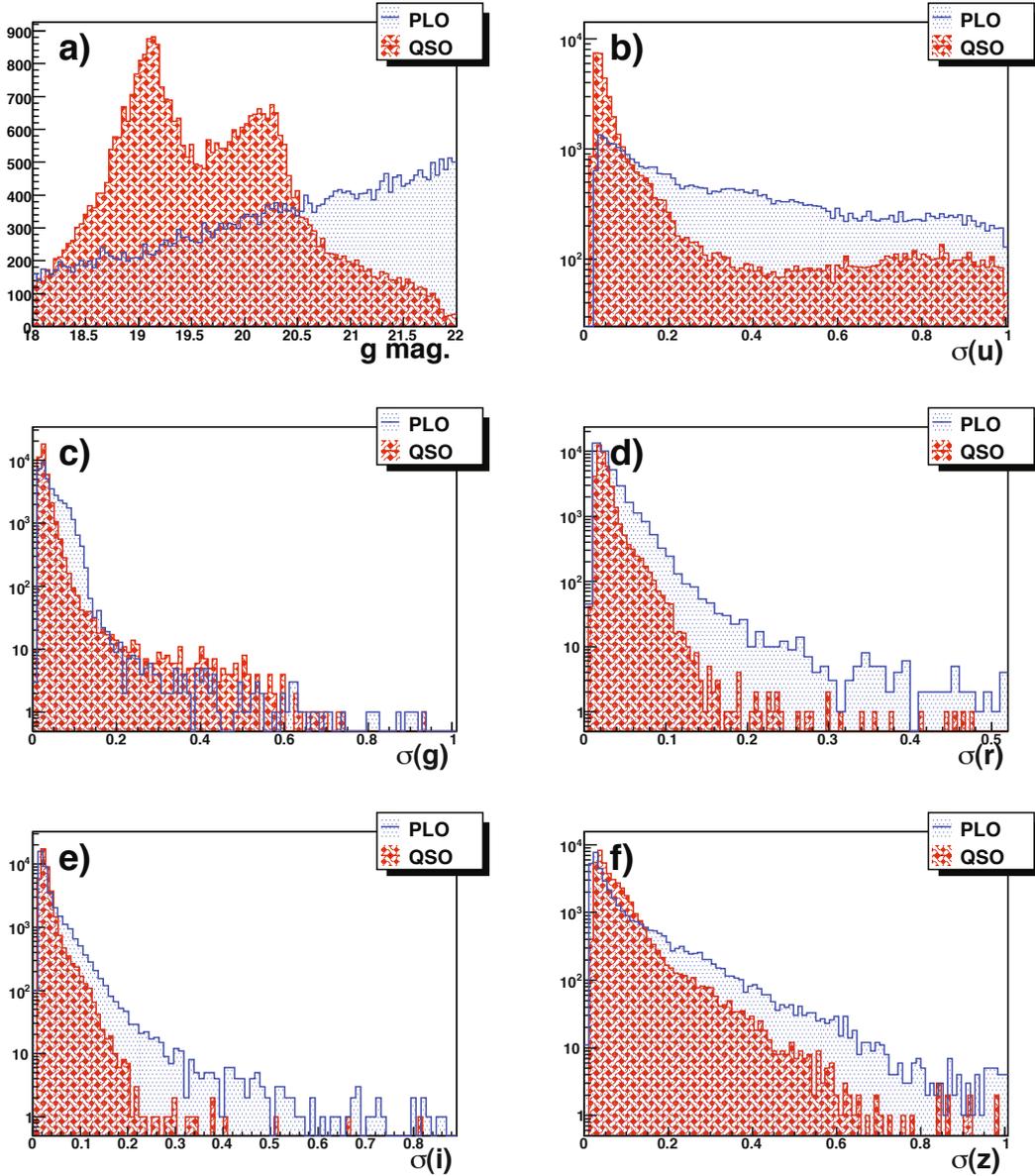
The classification of objects can be most simply performed by applying cuts to various distributions that distinguish signal objects from background objects. This approach is not optimal because all the information (the shapes of the variable distributions, the correlations between the variables) is not exploited, which leads to a loss in classification efficiency. Statistical methods based on multivariate analysis have been developed to tackle this kind of problem. For historical reasons, these methods have focused on linear problems which are easily tractable. To deal with nonlinearities, artificial neural networks (ANNs) have been

shown to be powerful tools in the classification task (see for instance Bishop 1995).

By combining photometric measurements such as the magnitude values and their errors for the five bands ( $ugriz$ ) of SDSS photometry, an ANN approach will allow us both to select the QSO candidates and predict their redshift. Similar methods such as KDE (Richards et al. 2004, 2009b) already exist to select QSOs. Our approach based on ANN will use information not commonly used such as photometric errors and apparent magnitudes rather than only colors. Moreover, we propose to estimate the emission redshift at the same time with the same tool. This approach contrasts with conventional methods for computing photometric redshifts, which employ  $\chi^2$  minimization techniques (Richards et al. 2001; Weinstein et al. 2004).

## 2. QSO and background samples

As mentioned above, quasar candidates should be selected from a photometric catalog of objects including real quasars and what we refer to as background objects. Here, for both the background and QSO samples, the photometric information comes from the SDSS-DR7 imaging database of point-like objects (Abazajian 2009), hereafter PLOs. We apply the same quality cuts to the photometry for the two samples and select objects with  $g$  magnitude in the range  $18 \leq g \leq 22$ . In the following,



**Fig. 2.** Distributions of the variables used as input in the ANN for objects classified as PLO in SDSS photometric catalog (blue dotted histogram) and for objects spectroscopically classified as QSO (red slashed histogram): **a)** Distribution of the PSF  $g$  magnitude, **b)**, **c)**, **d)**, **e)**, and **f)**. Distributions of, respectively,  $\sigma(u)$ ,  $\sigma(g)$ ,  $\sigma(r)$ ,  $\sigma(i)$ , and  $\sigma(z)$ , the errors in the corresponding PSF magnitudes. Note that the  $g$  distribution for the QSOs used for the ANN training is biased by the spectroscopic selection. This issue will be addressed in the future with the first observations of BOSS, which would provide a “less biased” training sample.

we note magnitudes are point spread function (PSF) magnitudes (Lupton et al. 1999) in the SDSS pseudo-AB magnitude system (Oke & Gunn 1983).

### 2.1. Background sample

For the background sample, we would like to use an unbiased sample of spectroscopically confirmed SDSS point-like objects *that are not QSOs*. Unfortunately, we have no unbiased sample of these objects because the spectroscopic target selection of SDSS-I was biased to particular types of objects. Fortunately, the number of QSOs among PLOs is sufficiently small that using all PLOs as background does not affect the ANN’s ability to identify QSOs. We verified that this strategy works by using the synthetic PLO catalog of Fan (1999). We degraded the star sample by adding a few percent of QSOs. We then retrained the ANN

and compared our results with those for an ANN trained with a pure star sample. We did not observe any significant worsening of the ANN performances.

The background sample used in the following was drawn from the SDSS PLO sample. We used objects with Galactic latitude  $b$  around  $45^\circ$  to average the effect of Galactic extinction. In the future, we may consider the possibility of having a different ANN for each stripe of constant Galactic latitude. The final sample contains 30 000 PLOs: half of them constituting the “training” sample, the other half the “control” sample, as explained in the next section.

### 2.2. QSO sample

For the QSO training sample, we used a list of 122818 spectroscopically-confirmed quasars obtained from the 2QZ

quasar catalog (Croom et al. 2004), the SDSS-2dF LRG and QSO Survey (2SLAQ) (Croom et al. 2009), and the SDSS-DR7 spectroscopic database (Abazajian 2009). These quasars have redshifts in the range  $0.05 \leq z \leq 5.0$  and  $g$  magnitudes in the range  $18 \leq g \leq 22$  (Galactic extinction-corrected). Since quasars will be observed over a limited blue wavelength range (down to about 3700 Å), we targeted only quasars with  $z > 2.2$ . Therefore, the sample of known quasars includes 33918 QSOs with  $z \geq 1.8$ , half of which constitute the effective “training” sample, the other half the “control” sample. For the determination of the photometric redshift, we used a wider sample of 95266 QSOs with  $z \geq 1$ .

In order to compare QSOs with background objects from different regions of the sky, the QSO magnitudes were corrected for Galactic extinction with the model of Schlegel et al. (1998).

### 2.3. Definition of the variables

The photometric information is extracted from the SDSS-DR7 imaging database (Abazajian 2009). The 10 elementary variables are the PSF magnitudes for the 5 SDSS bands ( $ugriz$ ) and their errors. As explained in Richards et al. (2009b), the most powerful variables are the four usual colors ( $u - g$ ,  $g - r$ ,  $r - i$ ,  $i - z$ ) combining the PSF magnitudes. Figure 1 shows the 2D color-color distributions for the QSO and PLO samples.

These plots give the impression that it is easy to distinguish the two classes of objects but one needs to keep in mind that the final goal is to obtain a 50% efficiency for QSOs with a non-quasar PLO efficiency of the order of  $\sim 10^{-3}$ . Therefore to improve the ANN performances, we added the absolute magnitude  $g$  and the five magnitude errors. Their distributions for the two classes are given in Fig. 2. An improvement can be expected from the additional variables and also from the correlations between the variables. For example, it is indeed expected that errors be larger for compact galaxies than intrinsic point-like objects.

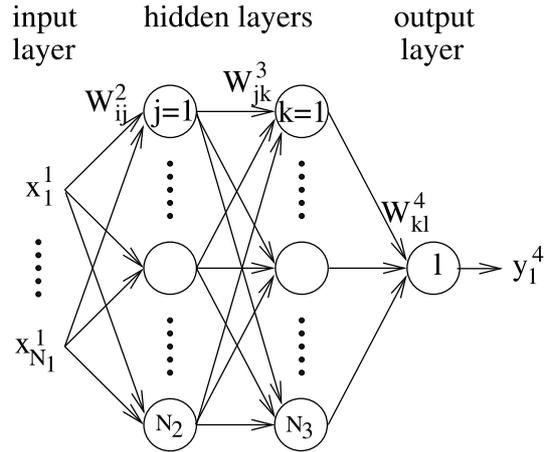
We note that the  $g$  distribution of the QSOs is biased by the spectroscopic selection. This issue will be addressed in the future with the first observations of BOSS for which the photometric selection of QSOs is based on loose selection criteria. They should provide a “less biased” catalog of spectroscopically confirmed quasars, close to completeness to magnitudes as faint as  $g = 22$ .

## 3. Artificial neural network approach

The basic building block of the ANN architecture<sup>1</sup> is a processing element called a neuron. The ANN architecture used in this study is illustrated in Fig. 3, where each neuron is placed on one of four “layers”, with  $N_l$  neurons in layer  $l$ ,  $l = 1, 2, 3, 4$ . The output of each neuron on the first (input) layer is one of the  $N_1$  variables defining an object, e.g., magnitudes, colors, and uncertainties. The inputs of neurons on subsequent layers ( $l = 2, 3, 4$ ) are the  $N_{l-1}$  outputs of the previous layer, i.e., the  $x_j^{l-1}$ ,  $j = 1, \dots, N_{l-1}$ . The inputs of any neuron are first linearly combined according to “weights”,  $w_{ij}^l$  and “offsets”  $\theta_j^l$

$$y_j^l = \sum_{i=1}^{N_{l-1}} w_{ij}^l x_i^{l-1} + \theta_j^l \quad l \geq 2. \quad (1)$$

<sup>1</sup> For this study, both for target selection and redshift determination, we use a C++ package, TMultiLayerPerceptron developed in the ROOT environment (Brun et al. 1995).



**Fig. 3.** Schematic representation of the artificial neural network used here with  $N_1$  input variables, two hidden layers, and one output neuron.

The output of neuron  $j$  on layer  $l$  is then defined by the non-linear function

$$x_j^l = \frac{1}{1 + \exp(-y_j^l)} \quad 2 \leq l \leq 3. \quad (2)$$

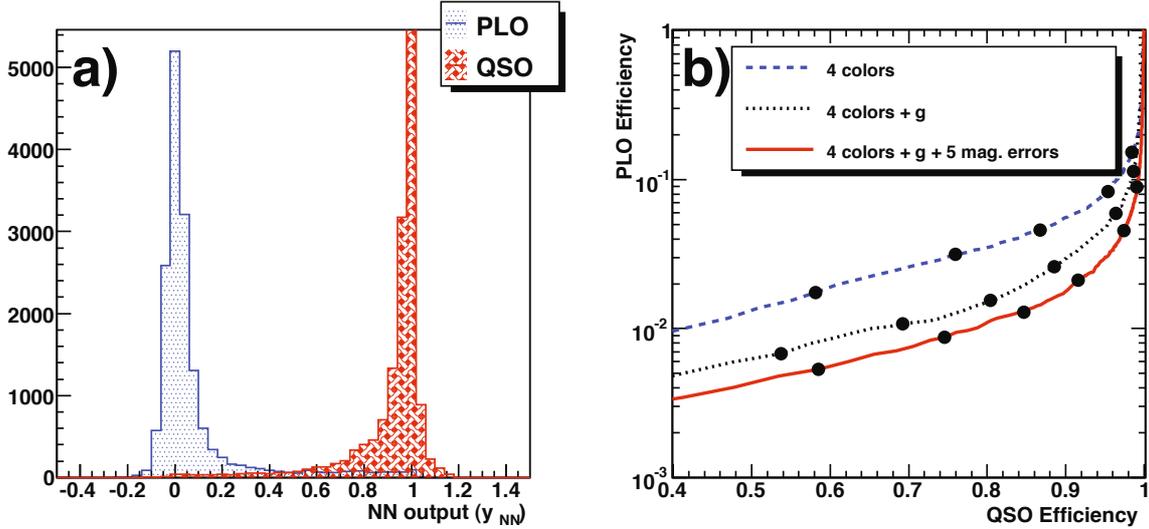
The fourth layer has only one neuron giving an output  $y_{\text{NN}} \equiv y_1^4$ , reflecting the likelihood that the object defined by the  $N_1$  input variables is a QSO.

Certain aspects of the ANN procedure, especially the number of layers and the number of nodes per layer, are somewhat arbitrary and are chosen by experience and for simplicity. On the other hand, the weights and offsets must be optimized so that the ANN output,  $y_{\text{NN}}$ , correctly reflects the probability that an input object is a QSO. The ANN must therefore be “trained” with a set of objects that are known to be either QSOs or not QSOs (background objects). More precisely, the weights and offsets are determined by minimizing the “error” function

$$E = \frac{1}{2n} \sum_{p=1}^n (y_{\text{NN}}(p) - y(p))^2, \quad (3)$$

where the sum is over  $n$  objects,  $p$ , and where  $y(p)$  is a discrete value defined as  $y(p) = 1$  (or  $y(p) = 0$ ) if the object  $p$  is a QSO (or is not a QSO). When the ANN is developed to estimate a photometric redshift, the targeted value  $y(p)$  is a continuous value equal to the true spectrometric redshift,  $z_{\text{spectro}}$ . We note that in the ANN architecture used for this study, the activation function, defined in Eq. (2), is not applied to the last neuron, allowing the output variable to vary in a range wider than  $[0; 1]$ .

In this kind of classification analysis, the major risk is the “over-training” of the ANN. It occurs when the ANN has too many parameters ( $w_{ij}$  and  $\theta_j$ ) determined by too few training objects. Over-training leads to an apparent increase in the classification efficiency because the ANN learns by heart the objects in the training sample. To prevent this behavior, the QSO and background samples are divided into two independent subsamples called “training” and “control” samples. The determination of the ANN parameters ( $w_{ij}$  and  $\theta_j$ ) is obtained by minimizing the error  $E$ , computed over the QSO and background training samples. The minimization is halted as soon as the error in the control samples stops decreasing even if the error continues to decrease in the training samples. We followed this procedure for both the target selection and the determination of the photometric redshift.



**Fig. 4.** **a)** ANN output for objects classified as PLO in the SDSS photometric catalog, i.e. background objects, (blue dotted histogram) and for objects spectroscopically classified as QSO (red slashed histogram) in the control samples, using 10 variables: 4 colors,  $g$  magnitude, and errors in the five ( $u, g, r, i$  and  $z$ ) magnitudes. **b)** PLO efficiency as a function of the QSO efficiency for three ANN configurations. Blue dashed line: 4 colors ( $u - g, g - r, r - i, i - z$ ). Black dotted line: 4 colors +  $g$  magnitude. Red solid line: 4 colors +  $g$  magnitude + errors in the five ( $u, g, r, i$  and  $z$ ) magnitudes. The curves are obtained by varying the cut value,  $y_{\text{NN}}^{\text{min}}$  for the two distributions of Fig. 4a. Efficiency is defined as the ratio of the number of objects with a ANN output greater than  $y_{\text{NN}}^{\text{min}}$  to the number of objects in the sample. The dots correspond, from left to right, to  $y_{\text{NN}}^{\text{min}}$  equal to, respectively, 0.2, 0.5, 0.8, 0.9, 0.95, and 0.98.

The result of the ANN training procedure is shown in Fig. 4a. The histograms of  $y_{\text{NN}}$  for the control QSO and background samples are overplotted. Most objects have either  $y_{\text{NN}} \sim 1$  (corresponding to QSOs) or  $y_{\text{NN}} \sim 0$  (corresponding to background objects). QSO target selection is achieved by defining a threshold value  $y_{\text{NN}}^{\text{min}}$  to be chosen between  $y_{\text{NN}} = 1$  and  $y_{\text{NN}} \sim 0$ . The optimal value of the threshold is obtained by balancing the number of accepted QSOs against the number of accepted background objects. A plot of the QSO efficiency versus the background efficiency is shown in Fig. 4b.

#### 4. Photometric selection of quasar

For illustration, we considered three ANN configurations that differ in terms of the number of variables. The first one uses only the four standard colors ( $u - g, g - r, r - i, i - z$ ). In the second configuration, we add the absolute magnitude  $g$  and finally in the third configuration, the errors in the five PSF magnitudes are also taken into account. For each configuration, we optimized the number of neurons in the hidden layers and the number of iterations in the minimization to derive the optimal “PLO efficiency-QSO efficiency” curve. The three curves are superimposed on Fig. 4b. Adding information, i.e. variables, clearly improves the classification performance. For instance, for a QSO efficiency of 50%, the PLO rejection fraction increases from 98.8%, to 99.4%, and then to 99.6% when the number of variables increases respectively from 4 to 5, and then to 10. In the region of QSO efficiency in which we wish to work, between 50% and 80%, the PLO background is reduced by a factor 3 by adding 6 variables to the four usual colors. The small improvement found by using photometric errors is not fully understood, but we think that it may be caused by a small contamination of the PLO catalog by compact galaxies.

It is therefore apparent that the 10-variable ANN should be used for the purpose of selecting quasars in any photometric catalog. In this case, the PLO rejection factors are 99.6%,

99.2%, and 98.5% for QSO efficiencies of 50%, 70%, and 85%, respectively.

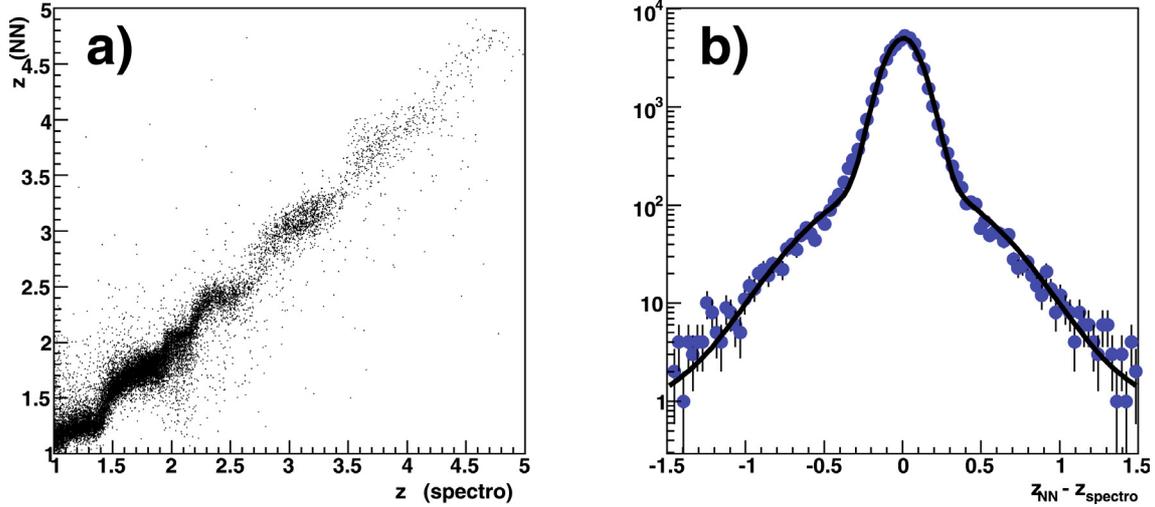
According to the McDonald & Eisenstein (2007) computation based on the Jiang et al. (2006) survey of faint QSOs, we expect to observe  $\sim 20$  QSOs per  $\text{deg}^2$ , with  $g < 22$  and  $2.2 \lesssim z \lesssim 3.5$ . For a Galactic latitude  $b \sim 45^\circ$ , the number of objects selected in the SDSS-DR7 imaging database is  $\sim 4000$ . Thus, with a QSO efficiency of 70% and a PLO efficiency<sup>2</sup> of 0.8%, we will select 32 objects per  $\text{deg}^2$  including  $\sim 14$  “true” QSOs. These numbers corresponds roughly to the requirements of the BOSS project.

#### 5. Photometric redshift of quasar

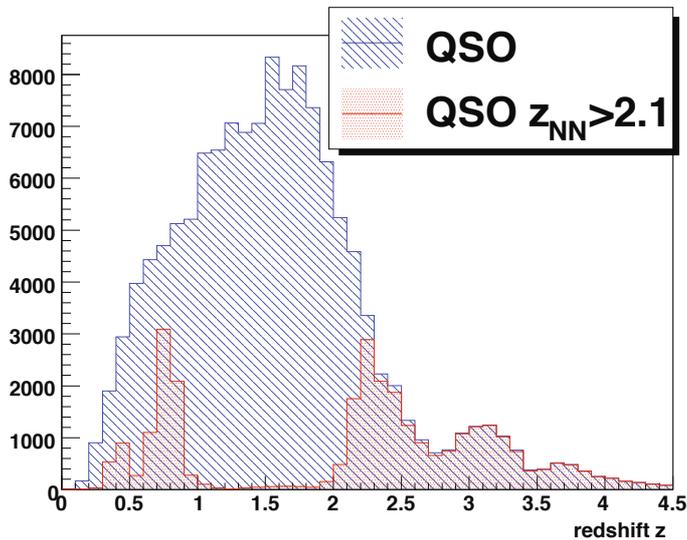
For the BOSS project, only quasars with a redshift in the range  $2.2 \lesssim z \lesssim 3.5$  are useful. In the definition of the training sample, we applied a cut to the redshift,  $z \geq 1.8$ , to enhance the selection of high- $z$  QSOs. But it is useful to add an additional constraint and select only QSOs with  $u - g > 0.4$ . This a posteriori color cut helps us to remove QSOs in the region  $0.8 \lesssim z \lesssim 2.2$ . However, we propose a more elegant method that consists of estimating the redshift of the QSO from the photometric information with another ANN.

To determine the photometric redshift, we used the same 10 variables employed in the ANN for target selection. The difference is that in defining the error  $E$ , in Eq. (3), the targeted value  $y(p)$  is a continuous value equal to the true spectrometric redshift,  $z_{\text{spectro}}$ . Except for this difference, the ANN architecture is the same as for target selection with two hidden layers with the same number of hidden neurons. The minimization is computed with a single “training” sample of spectroscopically-confirmed QSOs and it is halted as soon as the error  $E$  for the QSO “control” sample stops decreasing.

<sup>2</sup> Note that according to its definition in Sect. 2.1, the PLO sample contains QSOs.



**Fig. 5.** **a)** Photometric redshift determined with the ANN ( $z_{\text{NN}}$ ) as a function of the redshift measured from spectroscopy ( $z_{\text{spectro}}$ ). **b)** The  $z_{\text{NN}} - z_{\text{spectro}}$  distribution is fitted with three Gaussians contributing 93.4%, 6.4%, and 0.2% of the histogram and of width  $\sigma = 0.1, 0.4,$  and  $1.0,$  respectively. The rms of the  $z_{\text{NN}} - z_{\text{spectro}}$  distribution is 0.18 and its mean is 0.00.



**Fig. 6.** Spectrometric redshift distribution in the QSO sample (blue slashed histogram). The distribution for the QSO passing the cut  $z_{\text{NN}} > 2.1$  is overplotted (red dotted histogram). After this cut, 90.0% of the QSOs with  $z < 2.2$  are removed and only 5.3% of the QSOs in the  $2.2 < z < 3.5$  region are lost.

Figure 5a compares the photometric redshift,  $z_{\text{NN}}$ , determined by the ANN with the spectroscopic redshift of the spectroscopically-confirmed QSOs. Most of the objects are distributed along the diagonal illustrating the close agreement between the two measurements. This can be quantified by plotting the difference  $z_{\text{NN}} - z_{\text{spectro}}$  (Fig. 5b). The fit of this distribution with three Gaussians contains 93.4% and 6.4% of the objects in core and wide Gaussians, respectively. The fraction of outliers, determined by fitting the third Gaussian, is only 0.2%.

The corresponding distribution can be fitted with three Gaussian functions comprising, respectively, 93.4%, 6.4%, and 0.2% of the distribution and of width,  $\sigma = 0.1, 0.4,$  and  $1.$

Therefore, as shown in Fig. 6, by applying a conservative cut to the photometric redshift,  $z_{\text{NN}} > 2.1$ , we can remove 90.0% of the QSOs with  $z < 2.2$ . The fraction of lost QSOs with a redshift

in the relevant region,  $2.2 < z < 3.5$ , remains at a reasonable level of 5.3%.

## 6. Conclusions

We have presented a promising new approach to select quasars from photometric catalogs and to estimate their redshift. It involves the application of an artificial neurone network with a multilayer perceptron architecture. The input variables are photometric measurements, i.e., magnitudes and their errors for the five bands (*ugriz*) of the SDSS photometry.

In the target selection, we have achieved a PLO rejection factor of 99.6% and 98.5% for a quasar efficiency of 50% and 85%, respectively. The rms of the difference between the photometric redshift and the spectroscopic redshift is of the order of 0.15 to the region relevant to BAO studies. These new statistical methods developed in the context of the BOSS project can easily be extended to any other analysis requiring QSO selection and/or determination of their photometric redshift.

*Acknowledgements.* We thank N. P. Ross and D. H. Weinberg for triggering our interest in QSO target selection in the context of the BOSS project and for many interesting discussions. The authors are also grateful to G. T. Richards, A. D. Myers and E. Sheldon for important discussions and for providing the QSO catalog developed for the target selection in BOSS and used in this paper. We like also to thank Fan X. who has provided us some synthetic catalogs of PLOs.

## References

- Adelman-McCarthy, J. K., Jennifer, K., Agüeros, M. A., et al. 2008, ApJS, 175, 297
- Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, The Seventh Data Release of the SDSS, ApJS, 182, 543
- Bishop, C. M. 1995, Neural Networks for pattern recognition, Oxford University Press
- Brun, R., et al. (the ROOT Team) <http://root.cern.ch>
- Cauci, S., Colombi, S., Pichon, C., et al. 2008, MNRAS, 386, 211
- Cole, S., Percival, W. J., Peacock, J. A., et al. (the 2dFGRS Team) 2005, MNRAS, 362, 505
- Croft, R. A. C., Weinberg, D. H., Pettini, M., Hernquist, L., & Katz, N. 1999, ApJ, 520, 1
- Croom, S. M., Smith, R. J., Boyle, B. J., et al. 2001, MNRAS, 322, 29
- Croom, S. M., Smith, R. J., Boyle, B. J., et al. 2004, MNRAS, 349, 1397
- Croom, S. M., Richards, G. T., Shanks, T., et al. 2009, MNRAS, 392, 19

- Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. (the SDSS Collaboration) 2005, *ApJ*, 633, 560
- Fan, X. 1999, *AJ*, 117, 2528
- Jiang, X., Fan, X., Cool, R. J., et al. 2006, *AJ*, 131, 2788
- Lupton, R. H., Gunn, J. E., & Szalay, A. S. 1999, *AJ*, 118, 1406
- McDonald, P., & Eisenstein, D. J. 2007, *Phys. Rev. D*, 76, 063009
- Nusser, A., & Haehnelt, M. 1999, *MNRAS*, 303, 179
- Oke, J. B., & Gunn, J. E. 1983, *ApJ*, 266, 713
- Percival, W. J., Reid, B. A., Eisenstein, D. J., et al. 2010, *MNRAS*, 401, 2148
- Petitjean, P. 1997, *The Early Universe with the VLT*, ed. J. Bergeron (Berlin: Springer), 266
- Pichon, C., Vergely, J. L., Rollinde, E., Colombi, S., & Petitjean, P. 2001, *MNRAS*, 326, 597
- Richards, G. T., Weinstein, M. A., Schneider, D. P., et al. 2001, *AJ*, 122, 1151
- Richards, G. T., Nichol, R. C., Gray, A. G., et al. 2004, *ApJS*, 155, 257
- Richards, G. T., Deo, R. P., Lacy, M., et al. 2009a, *AJ*, 137, 3884
- Richards, G. T., Wilcock, W. S. D., Barclay, A. H., et al. 2009b, *ApJS*, 180, 67
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, 500, 525
- Schlegel, D., White, M., & Eisenstein, D. 2009, unpublished SDSS-III white paper for the US Decadal Survey, [arXiv:0902.4680]
- Schmidt, J. A. 1963, *Nature*, 197, 1040
- SDSS-III Collaboration, <http://www.sdss3.org/collaboration/description.pdf>
- Shanno, D. F. 1970, *Math. Comp.*, 24, 647
- Weinstein, M. A., Richards, G. T., Schneider, D. P., et al. 2004, *ApJS*, 155, 243