QUASARS WITH GAIA: IDENTIFICATION AND ASTROPHYSICAL PARAMETERS

J.-F. Claeskens[1], **A. Smette**[*2], **J. Surdej**[†1]

[1]Institut d'Astrophysique et de Géophysique de Liège, Liège, Belgium
[2]European Southern Observatory, Santiago, Chile

ABSTRACT

GAIA will provide astrometric and photometric observations for about 500 000 quasars distributed over the whole sky. The latter would constitute an isotropic grid of fixed sources perfectly suited to determine the referential frame. However, they must first be properly identified among stars, whose population is about 2 000 times larger. Using broad and medium band photometry, we first compare the efficiency of two analysis methods (χ^2 fitting and Artificial Neurone Networks, ANNs) to produce the QSO catalog with the lowest amount of contamination by stars. We then investigate whether the *GAIA* photometry could also provide precise values of the QSO astrophysical parameters (APs, like the redshift, the continuum slope, the emission line strength and possibly the extinction). To reach that purpose, we also compare the performances of the Spectral Principal Component Analysis (SPCA) with those of the χ^2 fitting and ANN analysis.

Key words: Quasars, Astrophysical Parameters, Photometry.

1. INTRODUCTION

The end-of-mission *GAIA* database will contain information on variability, astrometric position, parallax, proper motion and photometry (in 5 broad bands [BBP] + 12 medium bands [MBP]) for about 500 000 quasars down to the magnitude $G = 20$. The bulk of the observed quasars is expected to be in the redshift range 1.5 – 2. This dataset must be faced to the highly heterogeneous set of 48 921 quasars gathered in the present version of the Véron QSO compilation (Véron-Cetty & Véron 2003, including the set of 23 338 QSOs from the last 2dF release by Croom et al. 2004) and to the sample of 100 000 QSOs to be observed at high latitudes in the north galactic hemisphere by the Sloan Digital Sky Survey (SDSS, York et al. 2000).

Although *GAIA* will only increase the total number of known quasars by less than one order of magnitude, and despite the fact that *GAIA* will not break the record of the highest QSO redshift, it has unrivalled advantages: *i*- it will scan the whole sky and provide the most *homogeneous* QSO survey (including low galactic latitudes where no optical survey has ever been carried out); *ii*- the combination of the three independent, complementary measurements that GAIA will produce (astrometry, variability, photometry) will likely allow us to build the most complete optically-selected survey down to G $\simeq$ 20. *iii*- the end-of-mission angular resolution is excellent (Söderhjelm 2002, DMS-SS-01); *iv*- an astrophysical diagnosis (e.g. photometric redshift) is possible to some extent, without requiring spectroscopic follow-up.

Advantage *i*- is not only crucial to obtain a dense, isotropic grid of fixed sources well designed to determine the *GAIA* optical celestial reference frame, but also to study the matter distribution traced by the luminous QSOs on cosmological scales and, through later spectroscopy, to detect intergalactic absorbing clouds located along their lines-of-sight.

Advantage *ii*- will help to identify the most peculiar types of QSOs and study their populations. Indeed, reaching a high degree of completeness in a traditional QSO survey is usually difficult because of the large differences between the quasar spectra (due to redshift, variable galactic reddening, variable absorption by intervening clouds along the line-of-sight, or due to intrinsic properties such as weaker emission lines [BL-Lac objects], broad absorption lines [BAL] QSOs, red continuum QSOs and type II objects with narrow emission lines).

Advantage *ii*- will also lead to a lower contamination rate of the QSO sample by specific stellar populations (e.g. white dwarfs at low redshifts, M stars at high redshifts and early F stars for $2 \lesssim z \lesssim 3$, where the Ly_α line is in the B filter) as compared with traditional broad band surveys. Given the fact that the QSO population does only represent 0.05% of the stellar population, building a *secure* sample with 100% QSOs and without the help of a spectroscopic follow-up requires a very efficient rejection algorithm. For comparison, the sample of color-selected QSO candidates in the SDSS does only contain 66% of true QSOs (Schneider et al. 2003).

Finally, we would like to stress that, thanks to advantages

[*]research associate, f.n.r.s. (belgium)
[†]research director, f.n.r.s. (belgium)

i-, *ii-*, *iii-*, it will be possible to identify with *GAIA* more than 900 gravitationnaly lensed quasars with an angular separation $\Delta\theta \leq 1''$ between the images (provided that an image of that size is telemetered to the ground). This number is an order of magnitude larger than the presently known number of lensed QSOs. Given the importance of strong gravitational lensing to probe the dark matter distribution in distant galaxies and given its sensitivity to cosmological parameters (see e.g. Claeskens & Surdej 2002 for a review), *GAIA* will significantly contribute to the field by providing the largest and most uniform optical survey of lensed QSOs.

In the present study, we only deal with the *photometric* signature of QSOs present in the end-of-mission *GAIA* database. We use it *i-* to select QSOs among stars as efficiently as possible (see Section 4) and *ii-* to derive their astrophysical parameters (see Section 5). In Section 2, we present the libraries of synthetic spectra we have built (and which are available via the *GAIA* Photometric Working Group (PWG) Webpage[1]). In Section 3, we briefly describe the analysis methods (χ^2, ANNs, SPCA) and we summarize our conclusions in Section 6.

2. SPECTRAL LIBRARIES

The diversity of the QSO spectra must be taken into account to realistically test the *GAIA* performances. However, there are no observed QSO spectral libraries covering such a diversity over the full wavelength range sampled by *GAIA* ($\lambda\lambda 2400 - 10\,500$Å). On the other hand, a check of the robustness of the classification algorithms is only possible if two totally *independent* sets of objects are available. Thus, we generated synthetic librairies based on two different approaches: the *"Modified Templates"* and the use of *"Spectral Principal Components"*.

2.1. Modified Templates (MT)

We first built a composite spectrum in the rest frame range 310 - 8000Å, using four averaged observed spectra available in the literature (see Fig. 1). We then subtracted the continuum by fitting the relation (Ferland 1996):

$$C(\lambda) \propto \left[\left(\frac{\lambda}{\lambda_0} \right)^{-\alpha} \exp\left(-\lambda_{\rm X}/\lambda\right) \exp\left(-\lambda/\lambda_{\rm IR}\right) + a \right], \quad (1)$$

where $\alpha = 0.03$ is the *slope* of the continuum, and we were left with an emission line spectrum $E(\lambda)$, whose *total equivalent width* $W \simeq 10\,000$ Å. New spectra may then be generated by changing α and W in the observed ranges $-4 < \alpha < 3$ and $2 < \log W < 6$ respectively[2].

[1] http://gaia.am.ub.es/PWG/index.html

[2] The observed ranges of α and W have been derived from the fitting of $C(\lambda)$ and $E(\lambda)$ to 3411 QSO spectra extracted from the SDSS DR1β. Most of the QSOs are found in the range $-1 < \alpha < 1$ and $3000 < W < 10\,000$.

The resulting spectra is finally redshifted ($0 < z < 5.5$) and intergalactic absorption is added using the method described in Royer et al. 2000.

We then built two spectral libraries with 20 000 (resp. 17 325) spectra, corresponding to uniformly and randomly (resp. regularly) distributed values of the parameters z, α and W. Those libraries are available on the PWG Webpage.

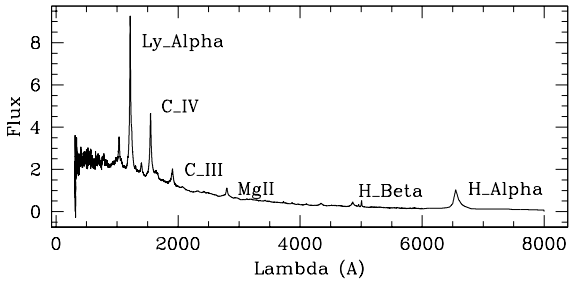


Figure 1. Composite QSO spectrum built from Cristiani & Vio (1990), Francis et al. (1991) and Zheng et al. (1997).

2.2. Spectral Principal Components (SPCs)

The (spectral) principal component analysis is a technique useful to identify spectral features that correlate with each other. Different studies indicate that any given rest frame QSO spectrum can indeed be reasonably well described by a linear combination of a mean spectrum $M(\lambda)$ and a set of principal components $P_i(\lambda)$ (Francis et al. 1992, Shang et al. 2003):

$$S(\lambda) \propto \left[M(\lambda) + \sum_{i=1,10} w_i \, P_i(\lambda) \right]. \quad (2)$$

We made use of this property to create an independent library of 20 000 synthetic QSO spectra. Here we only used w_i obtained by adjusting the Shang et al. SPCs to the SDSS DR1β QSO spectra. However, the library contains spectra uniformly covering the range $0 < z < 5.5$ instead of having the redshift distribution of the QSOs present in the SDSS DR1β. This library is also available on the PWG Webpage.

2.3. Stellar Libraries

Stellar contaminants must also be modeled. We considered single 'normal' stars, binaries and white dwarfs (WDs). Again, we used synthetic spectra, respectively from the Basel Stellar evolution Library BaSeL2.2 (Westera et al. 2002) with $2000\text{K} < T < 50\,000\text{K}$, $-1.02 < \log g < 5.5$, $-5 < [M/H] < 1$, from Malkov (*GAIA* blind testing cycle 2, UB-PWG-014) and from

Table 1. Summary of the photometric databases.

Name	Object	Origin	N	$N_{\rm sim}$
QSO-RAN	QSOs	MT	20 000	1
QSO-REG	QSOs	MT	86 625	25
QSO-PCA	QSOs	SPC	20 000	25
STAR-RAN1	Stars	Basel	10 000	1
STAR-RAN2	Stars	Basel	10 000	25
BIN-RAN1	Binaries	Malkov	1 000	1
BIN-RAN2	Binaries	Malkov	1 000	25
WD-REG1	W.Dwarfs	Koester	972	1
WD-REG2	W.Dwarfs	Koester	162	25

pure-hydrogen atmosphere models for degenerate stars (Koester, private communication), with $7000\mathrm{K} < T < 80\,000\mathrm{K}$ and $7.5 < \log g < 8.5$.

2.4. Photometric Databases

From the synthetic spectra, the expected numbers of photo-electrons collected at the end of the *GAIA* mission are computed with the *GAIA* II revised simulator (June 2003; C. Jordi) in the 2B BBP system (*GAIA*-LL-045), the 1X and the 2F MBP systems (*GAIA*-VV-008 and UB-PWG-01)[3]. Dust extinction is added at this stage with a uniform distribution in the range $0 < A_v < 9$ for normal and binary stars, and in the range $0 < A_v < 2.5$ for QSOs and WDs.

For all sets of spectra listed in Tab. 1, databases have been generated without noise as well as for G=18, 19 and 20 with realistic errors computed from the counts already affected by noise. Limiting magnitudes have also been introduced. For test sets (TS), $N_{\rm sim} = 25$, i.e. 25 realisations of the noise have been made for each entry in order to determine classification probabilities and uncertainties on the astrophysical parameters.

3. ANALYSIS METHODS

3.1. Nearest Neighbour (χ^2)

The simplest approach for classification and parameter determination consists in comparing the observed flux vector f_i with the flux vectors F_{ij} of each object j present in the database and selecting the one for which

$$\chi^2 = \sum_{\mathrm{filter}_i=1}^{n} \left(\frac{f_i - F_{ij}}{\sigma_i} \right)^2 \qquad (3)$$

is minimum. σ_i is the photometric error in filter f_i. This is the most popular minimum distance estimator, also called "template fitting". This is a *local* fitting in the sense that the result only depends on the local information in the color space. This method is powerful since it provides physical informations on the associated template, it identifies degeneracies between templates, it works with missing data and it provides hints on *unknown* objects from their large distance to any known object locus. However, the finite density of templates in the color space is the dominating source of errors on the APs for bright objects (requires interpolation) and is the dominating source of misclassification for faint objects. Given the number of operations, this method is *slow*.

[3]The choice of the BBP and MBP was not yet finalized at the time of the simulations. The present design (4B and 2X or 3F) is not drastically different. However, there is one more filter and the wings of the 3F filters are steeper; this can only improve the efficiency of the algorithms proposed in this study.

3.2. Artificial Neural Networks (ANNs)

A supervised ANN can be considered schematically as a black box able to learn during a "training" phase a non linear relation between known inputs (here the flux vectors) and known outputs (here the type of the objects or the value of the APs), contained in a *learning set* (LS). Once the relation is learnt, new outputs can be *quickly* and reliably derived from new inputs, provided the latter are statistically compatible with the inputs of the LS (see e.g. Bishop 1995 for more details on ANNs). ANNs are complementary to the χ^2-like approach in the sense that they belong to the class of *global* fitting methods; ANNs are basically interpolaters whose characteristics are defined by the global, or statistical properties of the LS. This may seem more robust, but ANNs suffer from the so-called *regularization problem*. In practice, the training must be tuned to learn the details of the distribution in the large dimensional color space while not "memorizing" the *exact* distribution of the LS (overfitting). In such a case, the ANN would not recognize a slightly different object belonging to the same family. To reach that goal, we found three empirical rules: i- complex problems should be broken into several simpler ones (e.g. successive binary tests to reject well defined populations of stars and WDs instead of one "big classifier"); ii- the data in the LS should be slightly noisier than in the test set (and therefore than in the sample that *GAIA* will observe); iii- objects in the LS located on the edges of the parameter space are duplicated. However two weaknesses remain: ANNs are sensitive to missing input data, such as the lack of flux information in a filter, and they are unable to flag *unknown* objects.

3.3. Spectral Principal Component Analysis (SPCA)

Equation 2 shows how the QSO spectral information is contained in only a small number of w_i. In turn, for a QSO for which GAIA produces a set of fluxes in $N_{\rm filter}$ filters, we can find the values of w_i by resolving the overdetermined linear system:

$$A_{ki}\, w_i = f_k \qquad (4)$$

where f_k is the observed flux in filter k and A is a $[N_{\rm filter} \times (N_{\rm PC}+1)]$ matrix with element ki given by the computed flux of the principal component i integrated in

Table 2. Percentage of the different objects present in the TS to be found in the QSO class (1X+2B/2F+2B)

Method	G	True STAR(%)	True QSO(%)	True WD(%)
χ^2	18	00.00/00.01	92.4/91.6	00.00/00.00
	19	00.01/00.01	83.9/84.8	00.00/00.00
	20	00.39/00.22	68.4/71.1	00.00/00.00
ANNs	18	00.00/00.03	68.5/83.0	00.00/01.23
	19	00.00/00.03	62.5/59.2	00.00/00.00
	20	00.00/00.02	31.8/40.4	00.00/00.00

the filter k. This computation is done after the averaged spectra $M(\lambda)$ and each $P_i(\lambda)$ have been redshifted and multiplied by a function representing the absorption by the intergalactic medium. In practice, the inversion of Eq. 4 must be done under the constraint of positivity of the resultant spectrum.

This method is aimed at recovering the maximum of spectral information contained in the MBP. It may also be used to determine the redshift of the QSO. Indeed, when the assumed value of z is wrong, the system of equations (4) often can not be solved or yields unrealistic values of w_i or χ^2. The final adopted value for z is the one for which a solution can be found for the w_i and which minimizes

$$\chi^2(z) = \sum_{\mathrm{filter}_k=1}^{n} (\hat{f}_k(z) - f_k)^2/\sigma_k^2 \qquad (5)$$

where $\hat{f}_k$ is the reconstructed flux in filter k.

As shown in Section 5, this methods works quite well when a good signal to noise ratio is achieved in many filters, i.e. for $z \lesssim 2$.

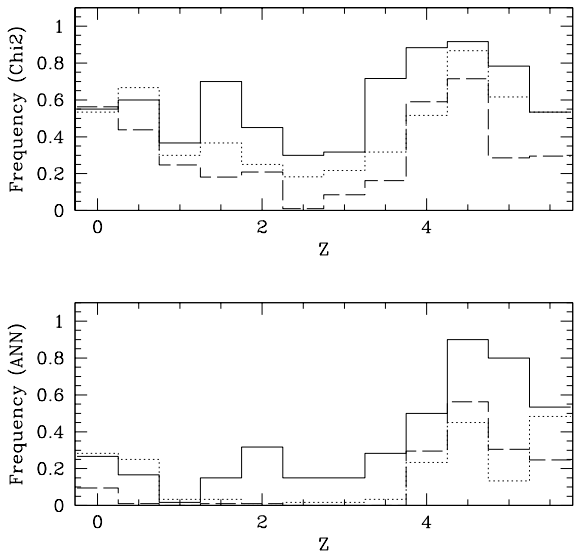


Figure 2. Redshift distributions of the QSOs found with the χ^2 method (top) and with ANNs (bottom) for G=20, $-0.5 \leq \alpha \leq 0.5$. Solid: $2500 \leq W \leq 10\,000$, $A_V = 0$; dotted: $2500 \leq W \leq 10\,000$, $A_V = 2$; dashed: $W < 2500$, $A_v = 0$.

4. QUASAR IDENTIFICATION

Whatever the method, the photometric identification of quasars among stars rely on the signature either of their strong emission lines (UV-excess like, for moderate redshifts) or of the Ly$_\alpha$ break absorption (large value of a specific color index: (B-V) for $3 \lesssim z \lesssim 4$, (V-R) for $4 \lesssim z \lesssim 4.5$, (R-I) for $4.5 \lesssim z \lesssim 6$).

In the following, the LS consists of the databases STAR-RAN1 + BIN-RAN1 + WD-REG1 + subset of 4 000 QSOs from QSO-RAN, while the test set (TS) is built with the databases STAR-RAN2+BIN-RAN2+WD-REG2+QSO-REG (see Tab. 1). The final class attribution follows the rule:

-QSO if $P_{QSO} > P_{STAR} + c$ and $P_{QSO} > P_{WD}$;
-STAR if $P_{STAR} > P_{QSO} - c$ and $P_{STAR} > P_{WD}$;
-WD if $P_{WD} > P_{QSO} - c$ and $P_{WD} > P_{STAR}$,

where P_{X} is the relative frequency of attribution of the class X among 25 simulations and $0 < c < 1$ is a *contrast parameter* intended to minimize the stellar contamination.

In Tab. 2, the results are compared for different G magnitudes, for both the χ^2 and the ANN methods and for both photometric systems. The general trends are the following: *i*- the χ^2 method leads to a higher completeness level (in the range ∼70-90% depending on the magnitude) but the ANNs are much better at rejecting contaminating stars, virtually reaching a 0% contamination rate even at $G = 20$; *ii*- white dwarfs are efficiently rejected with both methods; *iii*- both photometric systems give similar results, although the 1X+2B is slightly better in terms of stellar rejection with the ANNs.

Figure 2 shows the redshift distributions of the QSOs found with each method at $G = 20$. The tribute to pay for a high stellar rejection rate is the loss of QSOs looking too much like a star, which is especially the case for $2 \lesssim z \lesssim 3$ at "faint" magnitudes. The QSO depletion is particularly severe with ANNs, as expected from their higher stellar rejection rate. Figure 2 also shows that with both methods the completeness is even smaller for reddened QSOs ($A_v = 2$) as well as for weak emission line objects. Finally, very high redshift QSOs ($z \gtrsim 3.5$) will

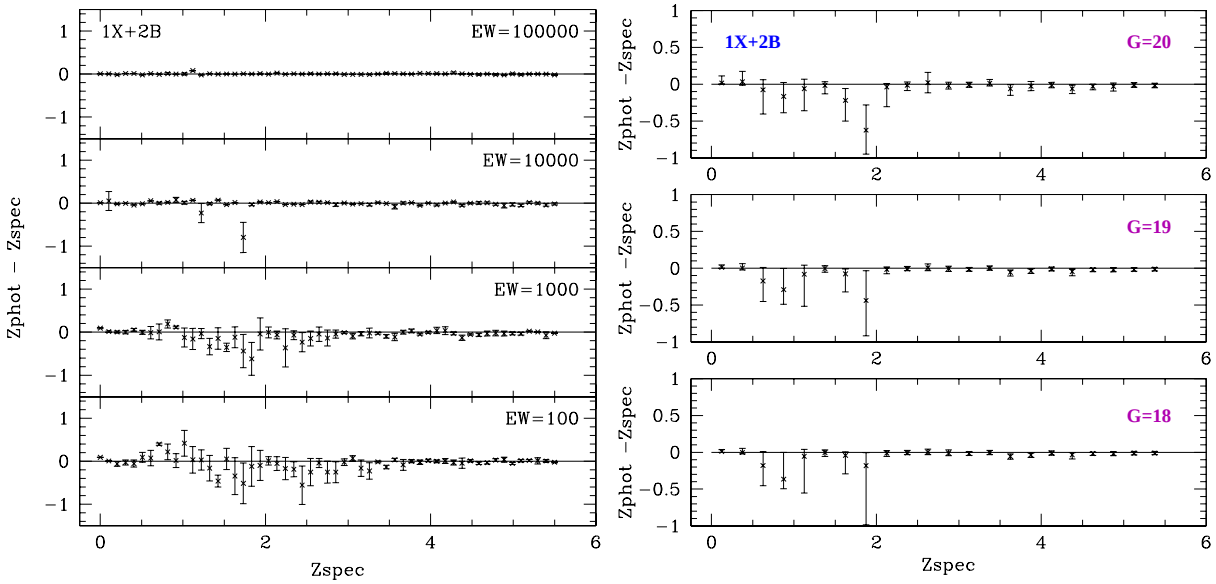


Figure 3. Left: error on the photometric redshift obtained with the χ^2 method, as a function of spectroscopic redshift for typical QSOs in the QSO-REG database (G=19, $\alpha = 0$ and $A_v = 0$ and different values of W; averaged values and 1σ error bars from the dispersion of 25 simulations). Right: same for the QSO-PCA database (G=18,19 and 20; median values and quartiles from the dispersion within redshift-bins of $\Delta z_{\rm bin} = 0.25$).

be more easily recognized, thanks to the strong and peculiar signature of the Ly$_\alpha$ break; unfortunately, they will not be very numerous because of the relatively bright limiting magnitude of *GAIA*.

With the χ^2 method, the few contaminants at $G = 18$ are hot stars confused with BL-Lac QSOs while most of the contaminants at $G = 20$ are *low metallicity reddened stars* ($[Fe/H] < -1$;$A_v > 2$) confused with reddened QSOs whose intrinsic continuum is flat or red ($\alpha < -1$). When relaxing the contrast parameter with ANNs, the first contaminants are cold, low metallicity stars.

The *observed* completeness of the QSO catalog and the *observed* stellar contamination rate depend on the intrinsic QSO and stellar AP distributions. The contamination rate also depends on the stellar to QSO population ratio, which is a function of the galactic latitude b. To get an idea of the expected results, we adopted the QSO APs distribution observed in the SDSS DR1β, we made use of the Besançon Galactic Model (Robin et al. 2003) to predict the distribution of the stellar APs and we estimated the population ratio from the observed QSO number count relation (Hartwick & Schade 1990) and the predicted stellar number count as a function of b from the Bahcall & Soneira (1980) model. With the χ^2 (resp. ANN) method, the *observed* QSO completeness is found to range between 90% and 55% (resp. between 47% and 16%) for $18 < G < 20$. The lower ANN completeness is due to the non-uniform observed redshift distribution[4]. On the other hand, the stellar contamination should be maintained below 10^{-3}% with the ANNs for any values of G and b, while it ranges from a few percent at large b to 95% towards the bulge with the χ^2 method at G=20.

Finally, an independent check using the QSO-PCA database as TS leads to a completeness level in the range 92 – 70% (resp. 62 – 10%) with the χ^2 (resp. ANNs) and $18 < G < 20$. The drop in completeness at $G = 20$ (for a uniform z-distribution) with ANNs indicates the danger of working only with synthetic libraries.

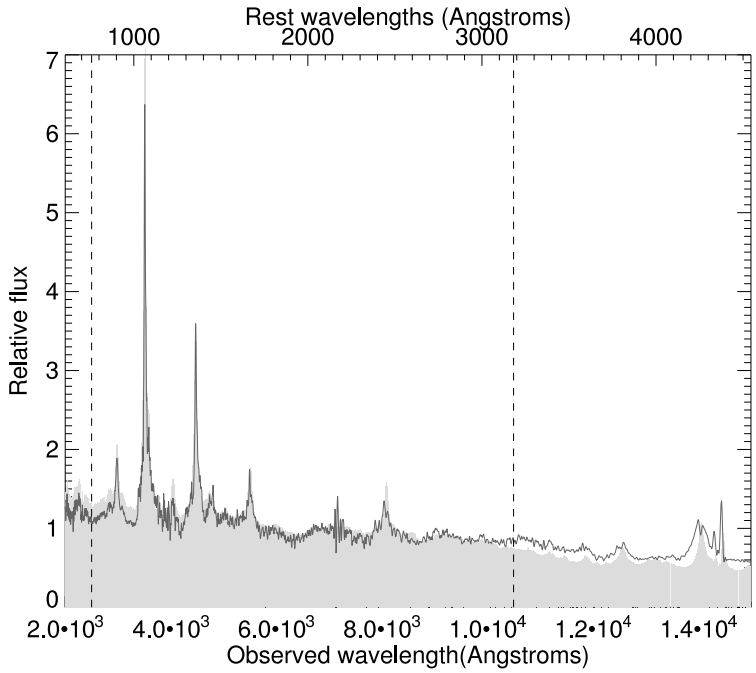
5. ASTROPHYSICAL PARAMETERS

In this Section, the LS only consists of the QSO-RAN database, while QSO-REG is the default TS and QSO-PCA is the independent TS. We shall not present results based on the ANN technique, since they are close to, but generally less good than, those obtained with the χ^2 analysis (see Claeskens et al. 2005 for more details). On the other hand, since the results are very similar between both photometric systems, we only show those relative to the 1X+2B system.

Among all QSO APs, the redshift is the most important one since it governs the QSO distance and luminosity through the cosmological model. It has the largest effect on the QSO spectrum, so this parameter should also be the easiest one to determine on the basis of the *GAIA* photometry.

Figure 3 displays the expected errors on the photometric redshift $z_{\rm phot}$ as a function of the spectroscopic redshift $z_{\rm spec}$. The left graph shows that, unsurpinsingly, $z_{\rm phot}$ is best determined when $z_{\rm spec} \gtrsim 3$ (i.e. when the Ly$_\alpha$ break enters the bluest filter) and for QSOs with strong emission lines. The larger errors in the range

[4] A better completeness should certainly be achieved with ANNs by including information on the redshift distribution in the LS, but we did not want to bias the searching algorithm at this level.



cmcm

Figure 4. Example of a spectral reconstruction using the SPCA method. The shaded region represents the input spectrum (G=19, $z_{\rm spec} = 1.89$, $\alpha = 0.03$, $W = 10\,000$) while the reconstructed spectrum is displayed with the solid line. The vertical dashed lines delineate the GAIA wavelength range.

$0.5 \lesssim z_{\rm spec} \lesssim 2.5$ reflects degeneracies coming from the interplay between the limited wavelength coverage and coincidental wavelength shifts between strong emission lines (e.g. $\lambda_{\rm CIV}/\lambda_{\rm Ly_\alpha} \simeq \lambda_{\rm C[III}/\lambda_{\rm CIV} \simeq 1.25 \pm 0.2$). Those trends are amplified in the right graph, independently of the magnitude, which is a sign of systematic errors arising from the differences between the LS and the independent TS (QSO-PCA). Those differences cannot be avoided with synthetic libraries. A robust estimate of the error on $z_{\rm phot}$ can only be derived in that context. We find $|\Delta z|_{\rm Median} \simeq 0.2$ for $0.5 \leq z_{\rm spec} \leq 2$ and $|\Delta z|_{\rm Median} \simeq 0.03 - 0.05$ for $z_{\rm spec} > 2.5$.

The other QSO APs, namely the slope α, the total emission line equivalent width W and the extinction A_V have only second order effects on the photometry. We thus expect their determination to be less accurate. We can only make predictions within the QSO-REG TS. For G=19, $-1 \leq \alpha \leq 1$, $z_{\rm spec} < 2.5$ and $A_{\rm V} = 0$, we find $|\Delta\alpha|_{\rm Median}$ = 0.4 ($W \leq 3000$) – 1.1 ($W > 11000$), $|\Delta W/W|_{\rm Median}$ =1.8 ($W \leq 3000$) – 0.6 ($W > 11000$) and $|\Delta A_v|_{\rm Median}$=0.4 ($W \leq 3000$) – 0.2 ($W > 11000$). Those results confirm that those APs will be poorly constrained with the *GAIA* MBP and BBP.

Finally, Fig. 4 shows how a QSO spectrum (G=19) can be retrieved from *GAIA* photometry by using the SPCA method and solving Eqs. 4 and 5. Although the results might be too optimistic because the spectra belong to the set used to define the PCs, the method looks promising when the the S/N is good. It should be noticed that if the PCs are representative of the whole QSO family, spectral features may be constrained even out of the wavelength range covered by the *GAIA* photometric systems. Preliminary results indicate that the abolute error on $z_{\rm phot}$ averaged over the range $0 < z_{\rm spec} < 2.3$ is of the order of 0.12 for typical QSO spectra.

6. CONCLUSIONS

Given the importance of creating from the data to be collected by *GAIA* the largest and purest sample of QSOs with their photometric redshifts (see Sect. 1), we investigated in this first-step study the possibility to identify and characterize them, on the basis of their *photometry only* (1X+2B or 2F+2B systems). To those ends, we built synthetic spectral libraries and compared the χ^2 template fitting and the ANN methods. We also introduced the Spectral Principal Component Analysis. The results do not strongly depend on the adopted photometric system.

First, we found that building a *secure* QSO catalogue based on photometry alone is possible, although the incompleteness can be severe, especially in the galactic plane where the reddening is expected to be higher. To that aim, ANN are more efficient than the χ^2 approach, but the latter has a better completeness and could be used at high galactic latitude when the population ratio is more favourable to QSOs. It is however clear from this study that the photometry alone is not sufficient to reach a high degree of completeness, in particular for QSOs with $2 \lesssim z \lesssim 3$ located close to the galactic plane, and

it should be complemented with the variability and astrometric constraints to be provided by *GAIA*. However, adopting $\sigma_\pi = 160\mu$as at G=20, a 3σ measurement of the parallax is possible only within a distance of about 2 kpc. Since all stars hotter than M stars can be detected beyond that limit, their lack of apparent parallax will not distinguish them from QSOs. A model of the Milky Way will thus be necessary to quantify the probability for a given star to have no apparent parallax. The constraint on the proper motion looks more promising since most of the stars following the galactic rotation have a proper motion which will be detectable by *GAIA*. The galactic model should thus also include the stellar motion in order to estimate the probability that a given star has no apparent proper motion, which depends on the direction of the line-of-sight. This will be the aim of a second-step study.

Second, this study has shown that the photometric redshift of quasars can be reasonnably well retrieved from the BBP+MBP photometry. The median absolute error on $z_{\rm phot}$ is the largest for $0.5 \leq z_{\rm spec} \leq 2$ ($|\Delta z|_{\rm Median} \simeq 0.2$ using the QSO-PCA independent TS). Unfortunately, this is the redshift range where most of the QSOs are expected! The reason is to be found in the interplay between the limited wavelength coverage and the existence of casual multiplication factors between the wavelengths of several QSO strong emission lines. This degeneracy should be alleviated by adding photometric data farther in the ultra-violet. Such data are being obtained over the whole sky down to $m_{\rm AB}$=20.5 by the *GALEX* satellite (Martin et al. 2003). They should thus be included into the analysis of the *GAIA* photometry. Finally, we noted that the other QSO APs (α, W, A_V) are much less constrained by the *GAIA* photometry. However, recovering at the same time the redshift and the weights of the spectral principal components seems promising for high S/N objects.

ACKNOWLEDGMENTS

It is a pleasure to thank Prof. L. Wehenkel, who gave us access to the *GTDIDT* software, and L. Vandenbulcke who made the first investigations with ANNs and other automatic classification algorithms. We also thank P. Francis and Z. Shang for kindly providing us with their QSO SPCs.

REFERENCES

[1] Bahcall, J. N. & Soneira, R. M. 1980, ApJS, 44, 73

[2] Bishop, C.M., 1995, "Neural Networks for Pattern Recognition", Oxford University Press

[3] Claeskens, J.-F., Surdej, J., 2002, A&AR, 10, 263

[4] Claeskens, J.-F., Smette, A., Surdej, J., 2005, A&A, in prep.

[5] Cristiani, S., Vio, R., 1990, A&A 227, 385

[6] Croom, S. M., Smith, R. J., Boyle, B. J. et al. 2004, MNRAS, 349, 1397

[Ferland 1996] Ferland, G.J. 1996, Hazy, a Brief Introduction to CLOUDY 90, Univ. of Kentucky Physics Department Internal Report

[7] Francis, P.J., Hewett, P.C., Foltz, C.B. et al. 1991, ApJ 373, 465

[8] Francis, P. J., Hewett, P. C., Foltz, C. B., & Chaffee, F. H. 1992, ApJ, 398, 476

[9] Hartwick, F. D. A. & Schade, D. 1990, ARAA, 28, 437

[10] Martin, C. & GALEX Science Team 2003, American Astronomical Society Meeting, 203, #96.01

[11] Robin, A. C., Reylé, C., Derrière, S., & Picaud, S. 2003, A&A, 409, 523

[12] Royer, P., Manfroid, J., Gosset, E., Vreux, J.-M. 2000, A&AS 145, 351

[13] Schneider D.P., Fan., X., Hall, P.B. et al.,2003, AJ 126, 2573

[14] Shang, Z., Wills, B. J., Robinson, E. L. et al., 2003, ApJ, 586, 52

[15] Véron-Cetty, M.-P., Véron, P., 2003, A&A 412, 399

[16] Westera, P., Lejeune, T., Buser, R. et al., 2002 A&A 381, 524, 2002

[17] York, D.G., Adelman, J., Anderson, J.E. et al., 2000, AJ 120, 1579

[18] Zheng, W., Kriss, G.A., Telfer R.C. et al. 1997, ApJ 475, 469