A realistic QSO Catalogue for the Gaia Universe Model

prepared by: E. Slezak, F. Mignard
affiliation : OCA

reference: GAIA-C2-TN-OCA-ES-001-1
issue: 1
revision: 0
date: 27/08/2007
status: Issued

Abstract

This note describes how the QSO Catalogue for the DPAC simulation has been produced and summarized its main statistical properties. The underlying statistical distributions in redshift, luminosity and colors have been derived from the most recent released of the SDSS. We have generated a list of sources with similar statistical properties as the SDSS, but extrapolated to $G = 20.5$ (the SDSS sample being complete to $i = 19.1$) and taking into account the flatter slope expected at the faint-end of the QSO luminosity distribution. The space density per bin of magnitude and the luminosity function should be very close to the actual sky distribution.

Document History

Issue	Revision	Date	Author	Comment
3	0	27-08-2007	FM	Final updates of Sect. 8
3	0	02-08-2007	FM	Additional editing
2	0	01-08-2007	ES	Detailed description of the procedure
1	0	15-06-2007	FM	Creation & astrometry

Contents

1 Introduction **6**

2 Building an homogeneous sample from the SDSS-DR5 data **6**

3 Simulating QSOs with $i < 19.1$ **8**

4 SDSS colours and Gaia photometry **14**

5 Simulating QSOs up to $i = 22.0$ **16**

6 QSO number counts and faint-end correction **19**

7 Nearby objects with $z < 0.1$ **21**

8 Astrometric parameters **23**

8.1 Positions 23

8.2 Distances 23

8.3 Proper motions 23

8.4 Radial velocity 25

9 File properties **26**

Acronyms

The following table has been generated from the on-line Gaia acronym list:

Acronym	**Description**
AGN	Active Galactic Nucleus
BP	Blue Photometer
DPAC	Data Processing and Analysis Consortium
RD-X	Data Release number X
ICRF	International Celestial Reference Frame
PDF	Probability Density Function
QSO	Quasi-Stellar Object
RA	Right Ascension
RP	Red Photometer
SDSS	Sloan Digital Sky Survey
SED	Spectral Energy Distribution
TN	Technical Note

1 Introduction

Gaia will contribute significantly to the knowledge of the quasars by providing for the first time an all-sky $V = 20$ flux-limited survey, very difficult to carry out from the ground. Gaia should observe more than 300 000 QSOs over the sky, from which a clean subsample will be extracted in order to build the primary reference frame. This zero-proper motion survey will provide a direct realization of the quasi-inertial celestial reference frame with a residual rotation less than $0.5\ \mu\mathrm{as\,yr}^{-1}$ and a space density at least a hundred time larger than that achieved by the radio version of the ICRF. Many more secondary sources (stellar or extragalactic) will also facilitate the access to this frame. Although the QSOs are observed like point-like stars, they have peculiar astrometric and spectral properties not included at the moment in the Gaia Universe Model. Since these properties are important in the automatic recognition of these extragalactic sources and given the importance of their astrometric solution, a specific population of simulated QSOs must be included in the Universe Model to test dedicated pieces of software related to their processing. This note explains how this set was selected and completed with astrometric and photometric data.

2 Building an homogeneous sample from the SDSS-DR5 data

Taking benefit of the fifth release of the Sloan Digital Sky Survey (SDSS) and of the fourth edition of the related Quasar catalog by Schneider et al. (2007), we decided to make use of this information to generate a QSO list for the Gaia Universe Model with up-to-date parameters for their redshift and luminosity distributions, but still ignoring their spatial correlation properties (equatorial coordinates will therefore be drawn uniformly in an independent way).

As stressed in Richards et al. (2006), extreme care must be taken when constructing a statistical sample from observed data. To do so, we followed their recommendations and the information they provide about the SDSS and useful corrections.

The fourth edition of the SDSS Quasar Catalog contains more than 77 000 objects. Among them, we selected those which are identified as a point-like source, designated as a “primary” object, with a redshift value in the range $0. < z < 5.5$, and with an available BEST photometry. We then applied a uniform magnitude limit of $i < 19.1$, accounting before for the Galactic absorption[1] and for emission-line effects. Unless indicated, all i-band magnitude indications in the present paper refer to such unreddened emission-line free flux measurements.

As indicated in Richards et al. (2006), emission lines make indeed quasar appear brighter relative to the continuum and subtracting this emission-line contribution from the observed fluxes is required for comparing luminosity functions at different redshifts. The emission-line compo-

[1] The Galactic extinction in the i-band A_i is indicated to be computed from the provided value in the u-band A_u as : $A_i = 0.405 \times A_u$.

nent $K_{\rm em}(z)$ of the cosmological i-band $K(z)$ correction is shown in Fig. 1. It can be derived from Table 4 of Richards et al. (2006) by subtracting the continuum component $K_{\rm cont}(z)$ from the listed total K-correction. For a power-law SED in the UV-optical with index α_ν, $f_\nu \propto \nu^{\alpha_\nu}$, ($\alpha_\nu = -0.5$ is the canonical value Vanden Berk et al. 2001), the correction at $z = 0$ is given as: $K_{\rm cont}(z) = -2.5(1 + \alpha_\nu)\log_{10}(1 + z)$.

The threshold at $i = 19.1$ is due to the flux limit for the targeted low-redshift part of the survey selected from the $ugri$ colour cube. Quasar candidates at $z < 3$ with $i > 19.1$ are in fact selected via criteria other than those defining the "primary" multicolour sample. A maximum brightness limit of $i = 15.0$ on quasar candidates is also imposed from the spectroscopic follow-up to avoid contamination between adjacent spectra.

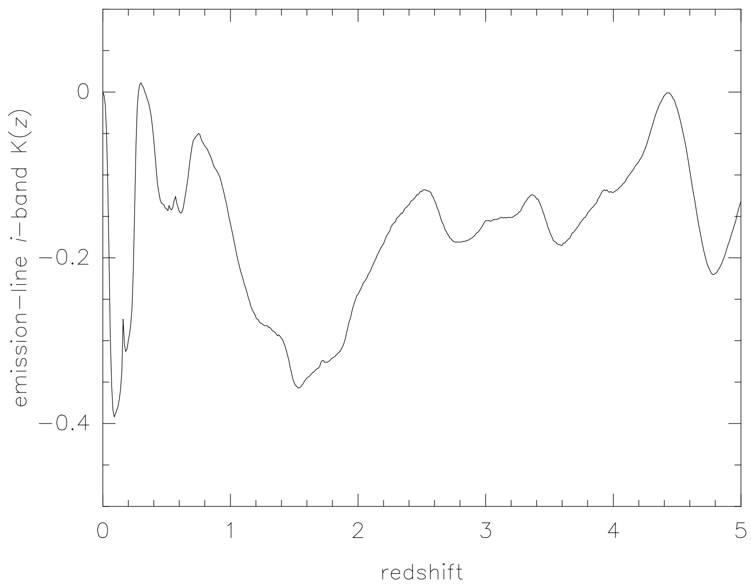

Figure 1: The emission-line component of the i-band K-correction (from Richards et al. 2006, Table 4).

This selection yields an homogeneous sample of 33 281 QSOs. However, before being used for statistical analyses, this smaller sample had still to be corrected for incompleteness effects. Fig. 2 shows the completeness of nonradio point sources averaged over $15.0 < i < 19.1$ in the *third release* of the SDSS data (cf. Richards et al. 2006, Table 1). The sharp drop around $z = 2.7$ and the dip at $z = 3.5$ arise because the SDSS colours of quasars at these redshifts are similar to the colours of stars, whence a reduced selection efficiency. We considered this function giving the completeness as a function of redshift and i-band magnitude to be still valid for the DR5 sample and following Schneider et al. (2007) we used it as in Richards et al. (2006). The correction to be applied at $z \sim 2.7$ being uncertain, a lower limit of 0.333 had been set on the selection function in the vicinity of this redshift to avoid any overcorrection.

Fig. 3 can be compared for checking to the same figure in Schneider et al. (2007, Fig. 3). It shows the redshift histograms of the 4th edition of the SDSS Quasar catalogue (upper curve)

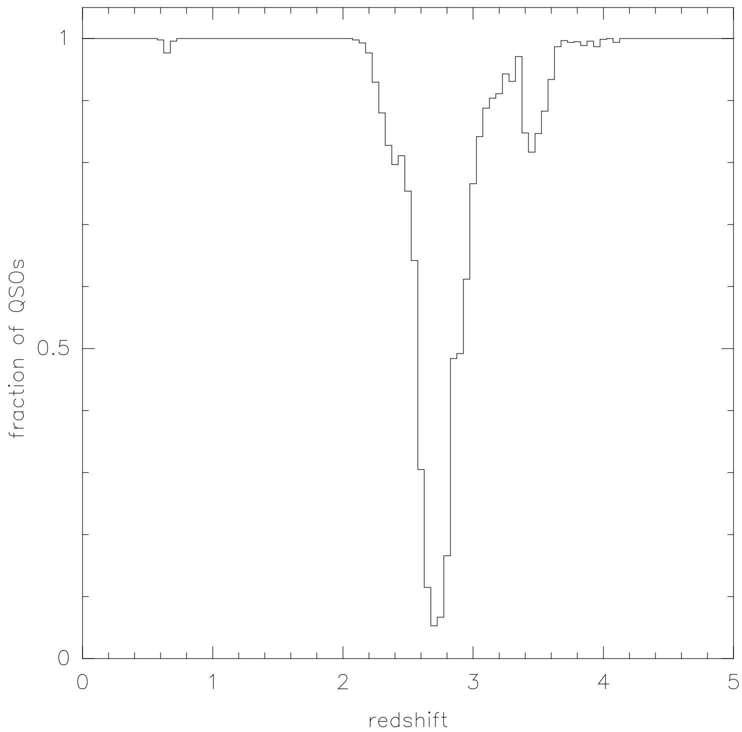

Figure 2: Quasar target selection completeness of nonradio point-like sources in the SDSS-DR3 averaged over $15.0 < i < 19.1$ (from Richards et al. 2006, Table 1).

with its marked structures around $z = 2.7$ and $z = 3.5$, a mode around $z = 1.7$ and most objects below $z = 2$. Also displayed are the redshift distributions of the homogeneous smaller sample we extracted (lower histograms) before and after having included the selection effects. In contrast with the histogram for the full catalogue, the redshift distribution of our SDSS-DR5 statistical subsample is smooth as expected and does not show deficits at $z = 2.7$ and $z = 3.5$. This histogram appears to be in very good agreement with the curve shown by Schneider et al. (2007) for their own subsample.

The distribution of the observed i-band magnitude (corrected for Galactic extinction and emission-line effects) as a function of redshift is displayed in Fig. 3 for the quasars included in our statistical subsample. The density of objects increases from black to white. The cut-off at very low z is due to the minimum luminosity criterion linked to the quasar definition ($M_i < -22.0$ measured in the rest frame). This feature is handled specifically in Sect. 7.

3 Simulating QSOs with $i < 19.1$

We tested our simulation strategy by building a catalogue of bright QSOs only with the same photometric properties as the homogeneous subsample of QSOs extracted from the SDSS-DR5 data. To do so, we adopted pragmatic solutions.

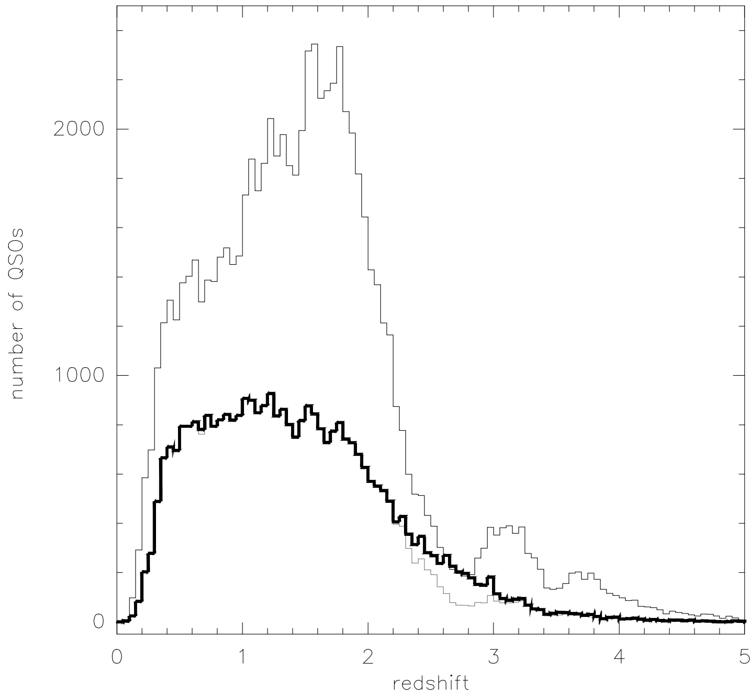

Figure 3: The redshift histograms of the QSO samples. The upper curve is the redshift distribution of the $\sim$ 77,400 objects in the 4th edition of the SDSS Quasar catalogue. The dips near $z \sim$ 2.7 and $z \sim$ 3.5 are due to the lower efficiency of the selection algorithm at these redshifts. The other two histograms show the redshift distributions of the $\sim$ 33,300 QSOs with $i < 19.1$ and labeled as point sources, before (grey line) and after (heavy line) correcting for selection effects. The width of the redshift bins is 0.05.

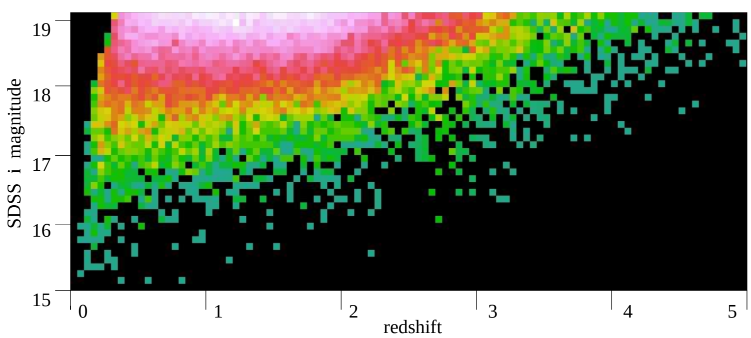

Figure 4: The i magnitude as a function of redshift for the 33,281 objects in the statistical subsample of QSOs with $M_i < -22.0$ extracted from the SDSS-DR5 catalogue. The magnitude is corrected from the galactic extinction and from emission-line effects. The incompleteness near $z \sim 2.7$ and $z \sim 3.5$ is accounted for. The curved cut-off on the left-hand side of the graph is due to the minimum luminosity criterion.

First, the observed redshift distribution of our SDSS-DR5 QSO subsample has been modeled in order to remove remaining specific features. Due to the shape of the redshift histogram, a best fit was found by considering the cumulative distribution of the real data for $z < 0.4$, a 2nd-order polynomial solution for data with $0.4 < z < 1.8$, and an exponential form for $z > 1.8$. Special care has been taken to properly connect together the different pieces of the solution. The resulting histogram is plotted in Fig. 5 with a smooth continuous distribution superimposed.

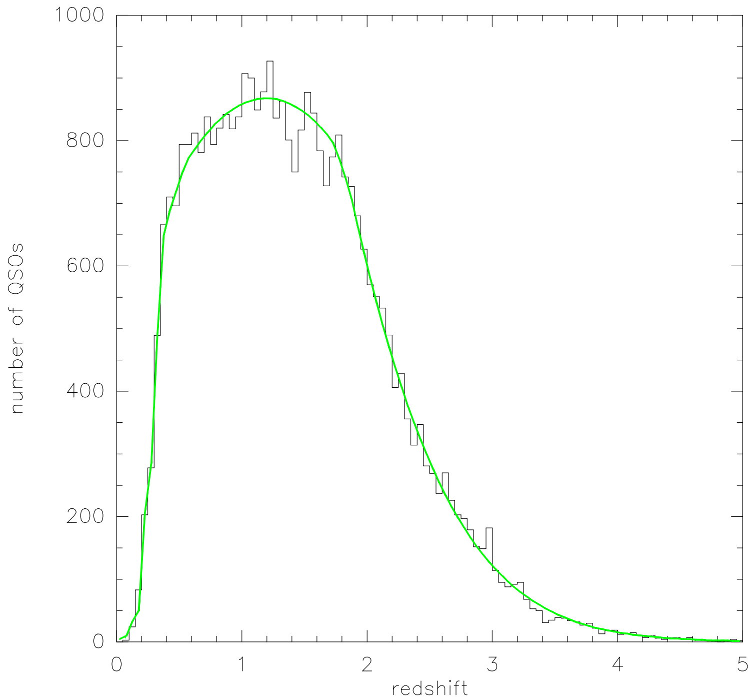

Figure 5: The redshift distribution for the QSOs in our statistical subsample with $i < 19.1$ (magnitudes are corrected for extinction and emission-line effects). The solid line shows the fit obtained by considering the cumulative distribution computed from the data for $z < 0.4$, a parabolic solution for $0.4 < z < 1.8$, and an exponential form for $z > 1.8$.

Then, we considered the relationship between the i-band magnitude and the redshift z. For computing accurate PDF (Probability Density Function), we transformed the discrete (z, i) histogram into a continuous distribution keeping most of the details of the original data. Such a result has been obtained by applying a noise cheating enhancement to these data with a small but large enough smoothing parameter. This more regular (z, i) distribution is displayed in Fig. 6. When compared to the original data (Fig. 4), one can verify that the chosen setting allowed us to smooth out regions where bin values are close to zero while not modifying bins with higher intensities.

Instead of trying to determine the QSO luminosity function and to look at its evolution with redshift, we simply decided to fit the quasar luminosity distribution for each available redshift bin of size 0.05. Differential number counts can be fitted by a pure exponential form providing that there is no evidence for a break or strong curvature in the data. Even if it is the case for

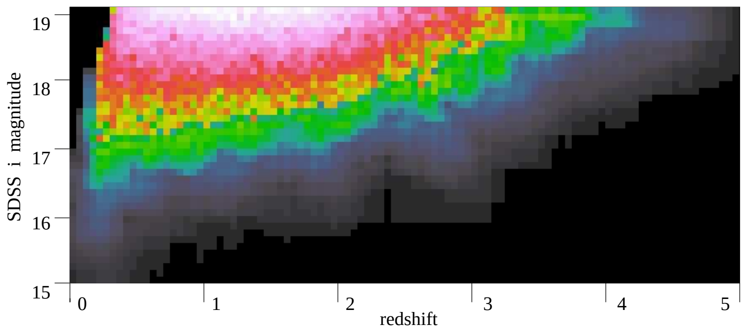

Figure 6: A noise cheating enhancement smoothing with a parameter $N = 20$ of the (z, i) distribution displayed in Fig. 4.

this subsample of QSOs brighter than $i = 19.1$, there is however some indication for a small curvature beyond $i = 18$ for some redshift bins (a flatter slope is indeed expected for the faint-end of the QSO luminosity function). We decided therefore to fit 2nd-order polynomials to the observed counts (in logarithmic units). Fig. 7 shows examples of the best fit we obtained for some redshift bins.

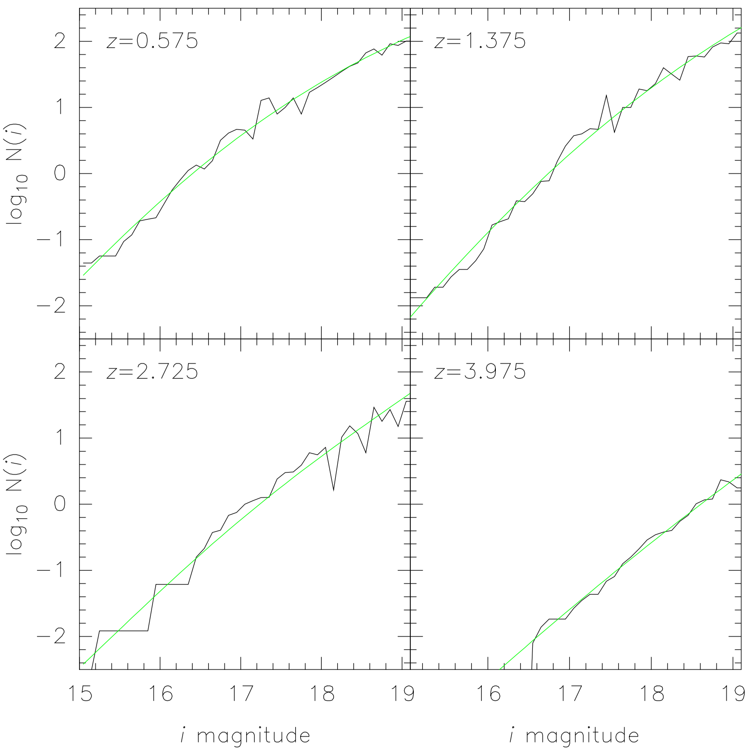

Figure 7: Examples of the quasar luminosity distribution for some redshift values. The width of the redshift and of the magnitude bins is 0.05 and 0.1 mag, respectively. Superimposed in each panel as a green solid line is the result of a 2nd-order polynomial fitting of the data set.

It was then straightforward with these results to compute the probability of having a given i-band

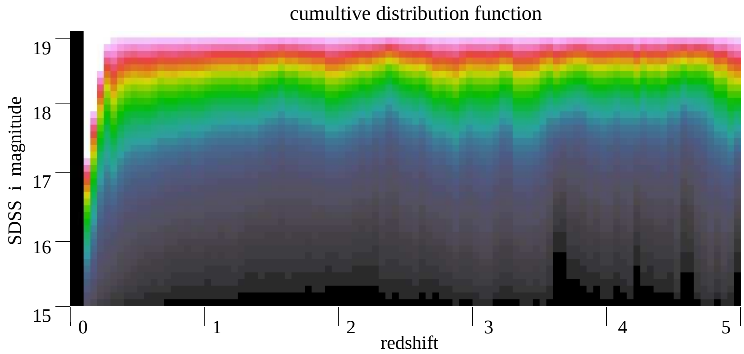

Figure 8: The cumulative distribution function $P_z(i)$ as a function of redshift for the statistical subsample of QSOs extracted from the SDSS-DR5 catalogue. Values increase from zero to one as the colour goes from black to white.

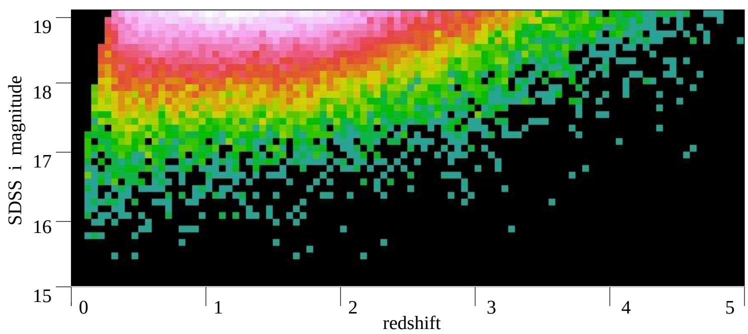

Figure 9: The i magnitude as a function of redshift for the simulated distribution of QSOs with $i < 19.1$ based on the statistical properties derived from the SDSS-DR5 homogeneous subsample.

magnitude for a peculiar redshift bin. This cumulative distribution function $P_z(i)$ is displayed in Fig. 8. Only objects with $z > 0.1$ have been considered here since very few data are available for $z < 0.1$.

Using this information we constructed a catalogue covering the same area of 5740 square degrees as the SDSS and spanning the same ranges in redshift and in magnitude as our homogeneous subsample. The distribution of this simulated data set in the (z, i) plane is displayed in Fig. 9 with the same look-up table as Fig. 4. From a statistical point both distributions appear very alike of view as expected.

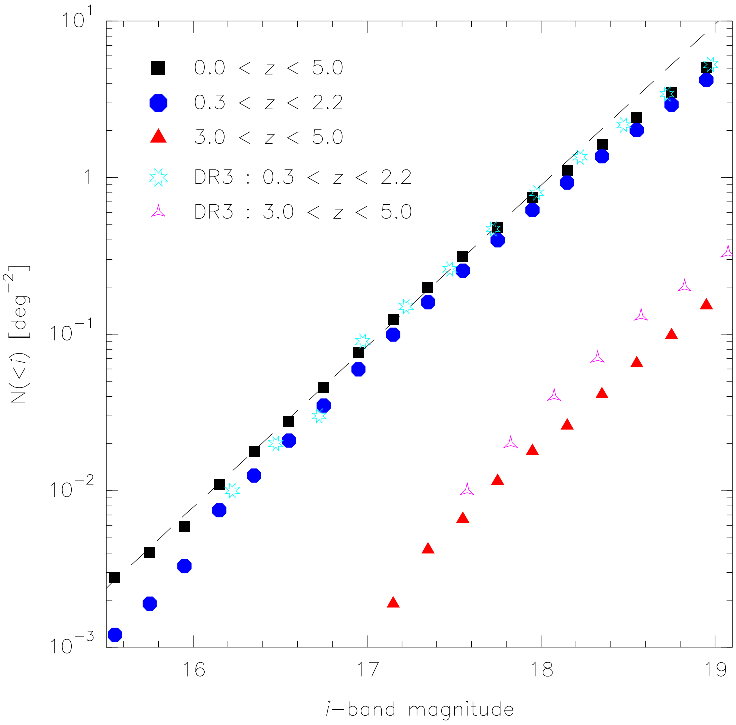

Figure 10: Cumulative i-band number counts with $0 < z < 5$ (squares), $0.3 < z < 2.2$ (circle) and $3 < z < 5$ (triangles) for our simulated distribution of QSOs with $i < 19.1$. A power-law fit over the bright range $16 < i < 18$ shown by a dashed line has a slope of 1.03 ± 0.01. The corresponding SDSS-DR3 data from Richards et al. (2006) are shown by starred circles and triangles, respectively, for comparison purposes only.

As a further check, the cumulative i-band number counts have been computed. As shown in Fig. 10, the bright-end $(16 < i < 18)$ of the whole simulated data set is well fitted by an exponential form with a slope of 1.03 ± 0.01. The value of this slope appears to be nearly the same as that found by Richards et al. (2006) for a least-square fit between $i = 16$ and 19 of the SDSS-DR3 $0.3 < z < 2.2$ QSO sample. For comparison purposes with this paper, our $0.3 < z < 2.2$ and $3.0 < z < 5.0$ cumulative counts are also displayed together with those coming from the SDSS-DR3 samples.

4 SDSS colours and Gaia photometry

Beyond a redshift and a i-band magnitude, each QSO simulated data point has been given $(g-i)$ and $(g-r)$ colours according to the statistical behaviour evidenced for our statistical subsample. As shown by Fig. 11, real data exhibit a typical colour for a given redshift. Taking into account these facts, we binned the data in redshift with a bin width of 0.05 and we calculated for each bin the average colour and the rms dispersion. As clearly shown in Fig.12, bluer colours are found for redshift values greater than $z = 3.$, but with intrinsic large error bars due to the small populations being involved. This bluer trend may reflect the change occurring at this redshift in the quasar target selection algorithm used in the SDSS, going from distinctive $ugri$ colours for $z < 3.0$ where most sources are UV-excess to a search for outliers from the stellar locus in a $griz$ colour space for objects with $z > 3.0$. Taking the colours at face value, each data point has been affected with such two colours according to a Gaussian distribution with mean and rms values related to the experimental data for its redshift. No specific correlation between colours has been taken into account for this first simulation.

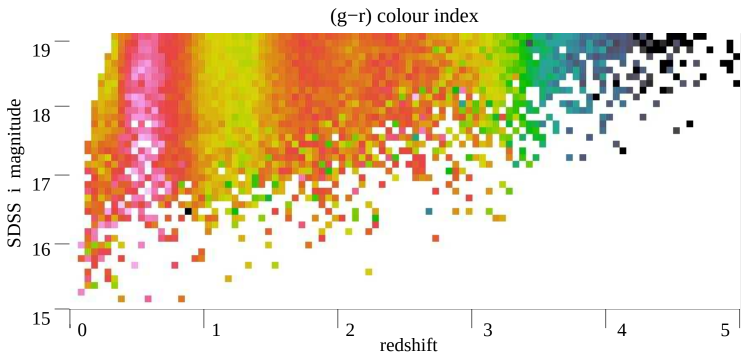

Figure 11: The $(g - r)$ colour index as a function of redshift and i magnitude for the statistical subsample of QSOs extracted from the SDSS-DR5 catalogue. The colour is unreddened.

Magnitudes in the Gaia photometric system were then computed using the relationships between the Gaia system and the SLOAN magnitudes and colours $C_1 = (g - i)$ and $C_2 = (g - r)$ according to the coefficients of the colour-colour polynomial fittings given in Table 3 of Jordi

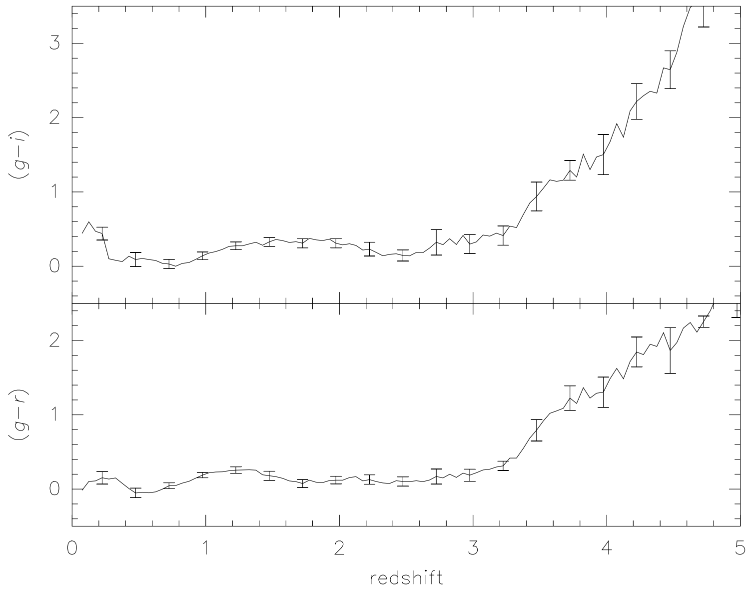

Figure 12: The average $(g-r)$ and $(g-i)$ colour index values as a function of redshift (bottom and top panel respectively) for the statistical subsample of QSOs extracted from the SDSS-DR5 catalogue. Error bars are given by the rms values and are displayed for every 10 bins for sake of clarity.

(2007):

$$
\begin{aligned}
g-G &= +0.0946 \\
&\quad +0.5318\times C_1+0.2788\times C_1^2-0.0004\times C_1^3 \\
&\quad -0.0230\times C_2+0.7085\times C_2^2-0.0200\times C_2^3 \\
&\quad -0.7341\times C_1\times C_2 \\
g-G_{rvs} &= +0.5300 \\
&\quad +2.0714\times C_1-0.0703\times C_1^2+0.0085\times C_1^3 \\
&\quad -1.3449\times C_2+0.5800\times C_2^2-0.0353\times C_2^3 \\
&\quad -0.2519\times C_1\times C_2 \\
g-BP &= -0.0330 \\
&\quad +0.0523\times C_1-0.0121\times C_1^2+0.0009\times C_1^3 \\
&\quad +0.2750\times C_2+0.1773\times C_2^2-0.0161\times C_2^3 \\
&\quad -0.0009\times C_1\times C_2 \\
g-RP &= +0.3644 \\
&\quad +1.3594\times C_1+0.1842\times C_1^2+0.0057\times C_1^3 \\
&\quad -0.5273\times C_2+0.8469\times C_2^2-0.0244\times C_2^3 \\
&\quad -0.7962\times C_1\times C_2
\end{aligned}
\tag{1}
$$

5 Simulating QSOs up to $i = 22.0$

Even if we succeeded in getting a simulated data set with statistical properties similar to the homogeneous subsample of QSOs from the SDSS-DR5 catalogue, this preliminary data set was however restricted to objects with $15.0 < i < 19.1$. Since the Gaia mission is planned to observe objects as faint as $G = 20$, we had therefore to generate objects with apparent luminosities fainter than $i = 19.1$ (but with absolute magnitude $M_i < -22.0$) to fulfill the requirements.

The best strategy would have been to make use of an analytical form for the QSO luminosity function and of its evolution with redshift, as given in the literature for deep surveys covering part of the sky. This is planned for our next simulation run. At present, in a first attempt to get quickly a catalogue including faint enough objects, we decided to follow an approach similar to the one we used in the previous section for obtaining the cumulative distribution function $P_z(i)$. This approach consists first in extrapolating at fainter magnitudes the quasar luminosity distributions we can get for different redshift bins from the SDSS-DR5 homogeneous subsample, and then to check the obtained QSO number counts and to correct them in an appropriate way since a change in slope is expected at the faint-end of the luminosity distribution.

In order to account for significative magnitude variations between the SDSS and the Gaia photometric systems for objects with large colour indices while getting a complete catalogue down to $G \sim 20.5$, we decided to simulate objects as faint as $i = 22.0$. Such a magnitude limit extends by 75% the $i = 19.1$ limiting magnitude of our statistical SDSS-DR5 QSO sample, which is challenging. In fact, extrapolation is always a tricky problem, especially far from the available set of data.

Fitting and extrapolating the differential luminosity distributions of QSOs with a 2nd-order polynomial (in logarithmic units) wouldn't lead to a satisfying solution since the relative increase in the differential number counts might be to slow at faint magnitudes. An exponential form $N = N_\circ\, 10^{\alpha i}$ leads to larger increases, but it is a less good maximum likelihood fit to the available data and the increases might be too large. Since our goal in this first study was not to determine the exact shape of the QSOs number density, we ended up with a compromise. We decided to model the differential number counts with redshift as a linear combination of the two solutions. with most weights equal to 0.9 and 0.1 respectively. Such a weighting scheme insures an almost best fit to data with $15.0 < i < 19.1$ while providing a fast enough increase in the number counts for fainter objects. Some examples of this first guess for the differential luminosity distribution at a given redshift are given in Fig.13. The quality of the extrapolation is not perfect for sure, with too steep slopes at the luminosity faint-end, but it appears to be good enough for our purpose.

The QSO luminosity distribution was processed this way separately for each 0.05 bin in redshift. Hence, no consistency is guarantee between extrapolated data points at close redshift values, which is very unlikely in any physical distribution. To insure such a consistency and to remove

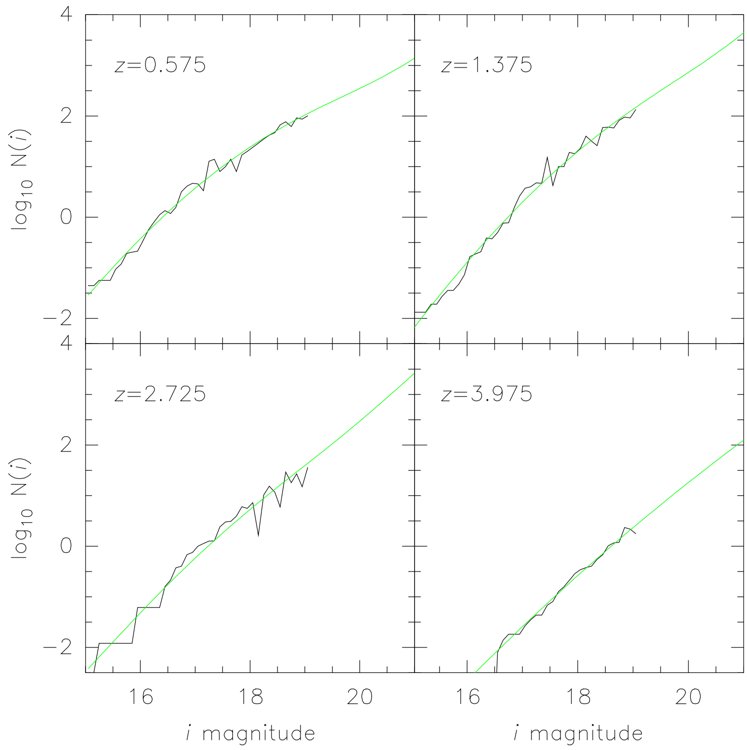

Figure 13: Some examples of the extrapolated luminosity distribution of QSOs from the homogeneous subsample for some redshift values. Superimposed in each panel as a green solid line is the result of a linear combination of a 1st-order and of a 2nd-order polynomial fitting of the data set we used as a first guess.

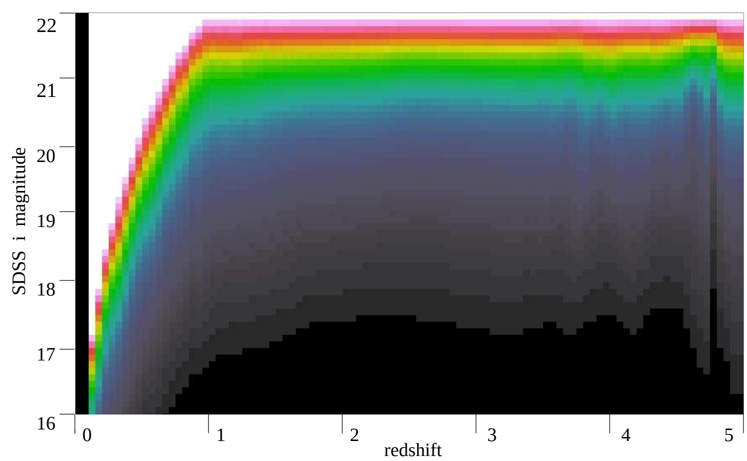

Figure 14: The cumulative distribution function $P_z(i)$ up to $i = 22$ as a function of redshift. Values increase from zero to one as the colour goes from black to white. The curved cut-off on the left-hand side of the graph is due to the minimum luminosity criterion.

any poor extrapolation for a peculiar redshift bin, the resulting 2-D cumulative distribution function $P_z(i)$ has been smoothed out along the redshift axis only with a triangle filter the size of which depends on the i magnitude. The final result is displayed in Fig.14 for $16.0 < i < 22.0$. It looks quite similar to the function obtained from the brighter data only (see Fig. 8).

This smoothed cumulative distribution $P_z(i)$ has been used to perform random trials and generate an all-sky list of objects with redshifts between 0.1 nd 5.0, i-band magnitudes in the range $15.0 < i < 22.0$, and a minimum luminosity of $M_i = -22.0$ (in a cosmology with $H_\circ = 70$ km s^{-1} Mpc^{-1}, $\Omega_M = 0.3$ and $\Omega_\Lambda = 0.7$). This simulated distribution is displayed in Figs. 15 & 16. Taking into account the 7-fold increase in area between the SDSS-DR5 sample and the celestial sphere, the statistical behaviour of this catalogue in the (z, i) plane seems to be satisfactory (but see next section). The overall shape is very similar to the one obtained when considering only objects brighter than $i = 19.1$ (Fig. 16) and the distribution of this subsample is nearly indistinguishable from that of the real data in the same magnitude range (cf. Fig. 4).

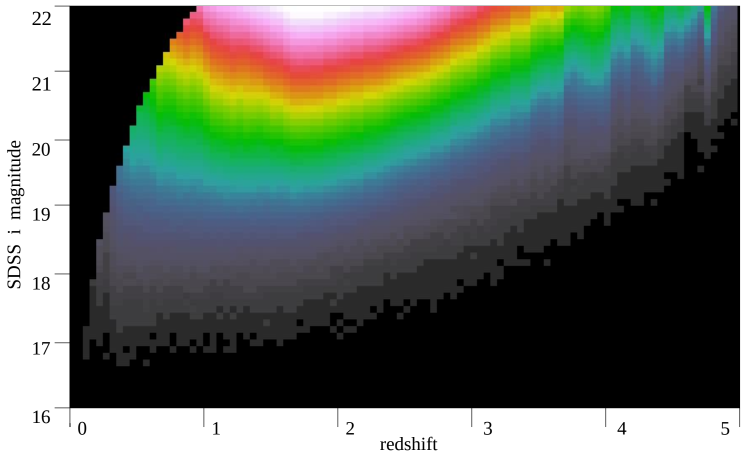

Figure 15: The i magnitude as a function of redshift for the simulated distribution of QSOs with $i < 22.0$ based on the statistical properties derived from the SDSS-DR5 homogeneous subsample with $i < 19.1$.

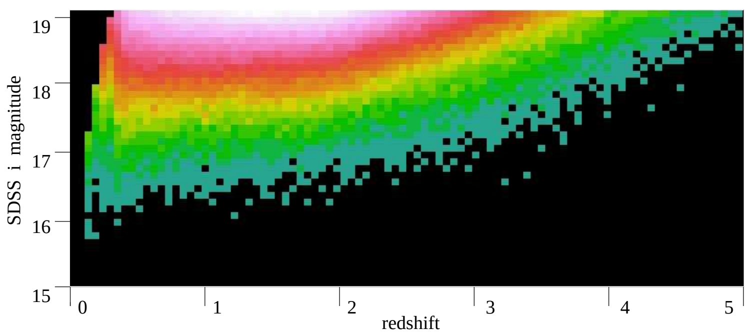

Figure 16: Zoom of Fig. 15 on $15.0 < i < 19.1$ with an ad-hoc look-up table for comparison purpose with Fig. 4 and Fig. 9.

6 QSO number counts and faint-end correction

Deep quasars surveys generally find a flatter slope for the quasar luminosity function fainter that some characteristic luminosity. One recent example of such a behaviour is given in Croom et al. (2004) for the 2QZ/6QZ catalogue based on sources selected from a photographic input catalogue. However, the SDSS quasar survey is pretty shallow, so that no object fainter than the break characteristic magnitude is observed at most redshifts. Even if some curvature can be found in the shape of the cumulative number counts for $i < 19.1$ (see Fig. 10), the magnitude range of our statistical sample derived from the SDSS-DR5 catalogue is not broad enough to get accurate information about this change in slope. A simulated data set simply extrapolating down to $i = 22$ mag the space density of the bright QSOs leads therefore to counts in excess with respect to real data for $i > 19.5$.

The standard parameterized form of the quasar luminosity function is therefore a double power-law. Choosing a functional form for a sample spanning a large redshift range like $0 < z < 5$ is however complex as stressed by Richards et al. (2006). One may for instance decide to adopt a single power-law but with an evolution of the characteristic luminosity in redshift (quadratic, exponential, ...). For this first simulation, we decided to adopt the standard form and to correct our list of objects so that the number count we derive at the faint-end does follow a flatter linear trend.

Taking benefit of the 2QZ/6QZ extra data points in the cumulative i-band number counts given by Richards et al. (2006) in their Fig. 13, we computed decimation factors for our first list of objects with $19.1 < i < 22.0$ in order to mimic the behaviour of the 2QZ/6QZ data at the faint end. The result of this correction procedure is displayed in Fig. 17. One can notice that the expected change in slope does occur at $i \sim 19.5$, which leads therefore to correct surface density values for objects fainter than this characteristic magnitude. At the limiting magnitude of $i = 20.5$, the correction factor appears to be larger than 3. The surface density for objects brighter than $G = 20$ is now found to be equal to 28.6 objects per square degree. At the bright end ($16 < i < 19$) we found a slope of 0.99 ± 0.01.

Considering from now on this improved decimated simulation, the distribution of the data points in the G-redshift plane is displayed in Fig. 18. When compared to Fig.15, one can note the effect of the colour transformation on the data, especially beyond $z = 3$ where colour values are found to be large. One can also see the impact of the decimation procedure for objects fainter than $G = 19.1$. Clearly seen is the fact that most objects lie at $1 < z < 2$. Part of this result may be due to the fact that our decimation factors does not change with redshift, which might be required to account for a space density evolution if any. The marginal distributions in redshift for different magnitude limits are shown in Fig. 19, both for the i-band (solid histograms) and for the G-band (dashed histograms). The fraction of objects with $1 < z < 2$ does indeed increase with the magnitude limit. The result of this trend can be compared with the redshift distribution of the SDSS-DR5 full quasar catalogue given in Fig. 3 where two peaks with related discontinuities are found at $z \sim 1.2$ and $z \sim 1.7$. Even if redshift values of our sample come

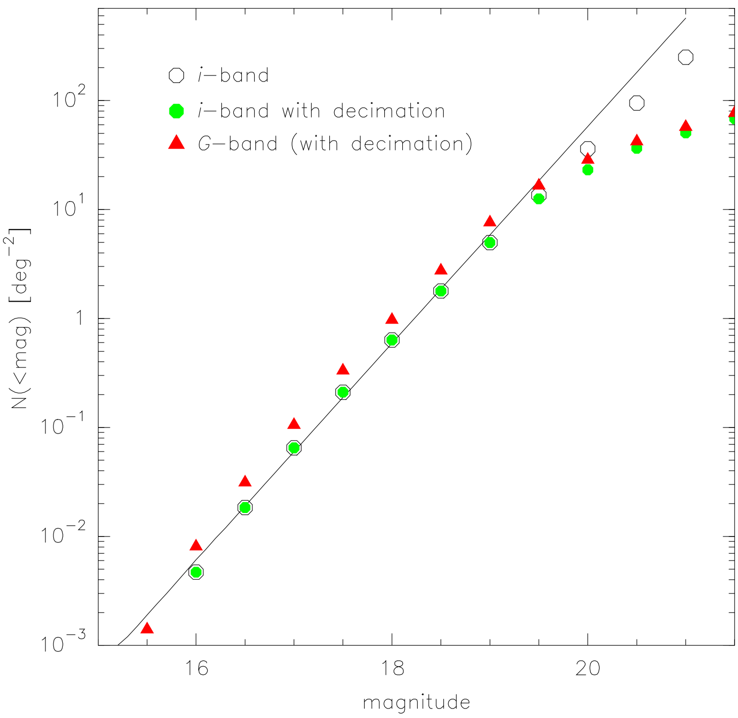

Figure 17: Cumulative number counts for simulated objects up to $i = 22$ with $0 < z < 5$ in the i-band with/without decimation (circles and disks, respectively) and G-band (triangles). A power-law fit over the range $16 < i < 19$ shown by the dashed line yields a slope of 0.99 ± 0.01.

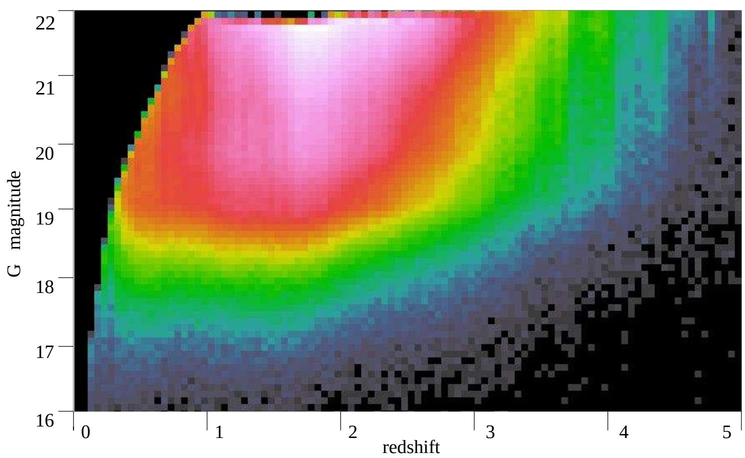

Figure 18: The Gaia G magnitude as a function of redshift for the simulated decimated distribution of QSOs with $i < 22.0$.

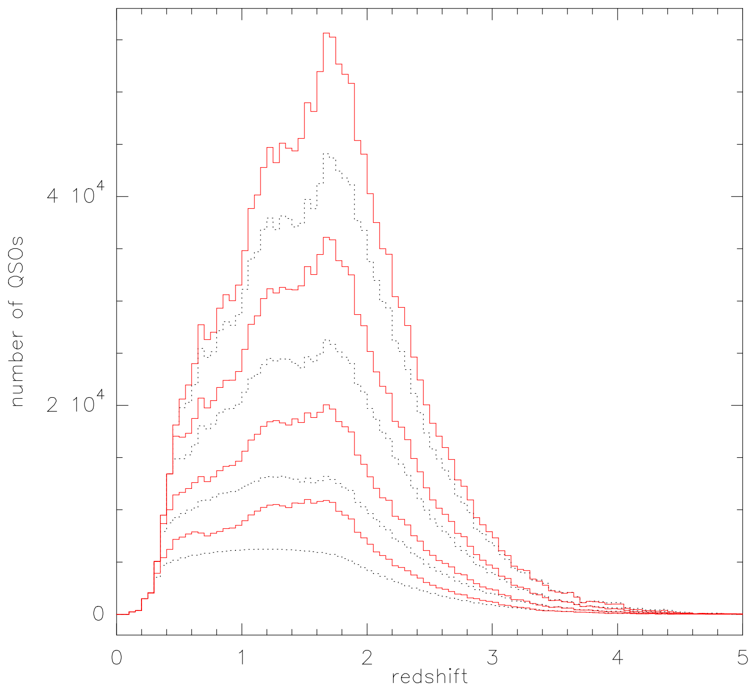

Figure 19: Number of QSOs per redshift bin of 0.05 in the decimated simulation up to limiting magnitudes of 19.1, 19.5, 20.0 and 20.5, for the i-band (dashed histograms) and G-band (solid histograms).

from random trials drawn from the featureless curve displayed in Fig. 5, such structures may still be embedded in the cumulative distribution function $P_z(i)$ we used to model the magnitude distribution as a function of redshift (see Fig. 14).

7 Nearby objects with $z < 0.1$

Target selection for spectroscopic follow-up of the SDSS quasar candidates implies a maximum brightness limit of $i = 15.0$ on the selected objects. One can also easily see for instance in Fig. 4 that very few objects with $z < 0.1$ are included in the SDSS-DR5 homogeneous subsample. Both kind of objects must be correctly taken into account in our simulation in order to obtain a complete list, even if their total number all-over the sky is quite small.

To do so we simply decided to add to our simulated data all the bright and nearby enough QSOs listed in a catalogue of genuine objects. We used the 12th edition of the catalogue of QSOs and AGNs by Véron-Cetty & Véron (2006) to identify real QSOs with $z < 0.1$ and an absolute luminosity brighter than $M = -23$. Some 182 objects were retrieved this way, but more than half of them without any optical flux measurement and only 19 of them with a colour indication.

To get a magnitude for all the objects, we first calibrated the V vs. z relationship on the available data with $z < 0.14$. Taking into account the shape of this 2-D distribution displayed in Fig. 20, we binned these data in redshift with a bin width of 0.05 and we fitted a 2-nd order polynomial to the average value per bin as shown. A V magnitude was then calculated for each QSO according to its redshift and to the rms dispersion for its redshift bin.

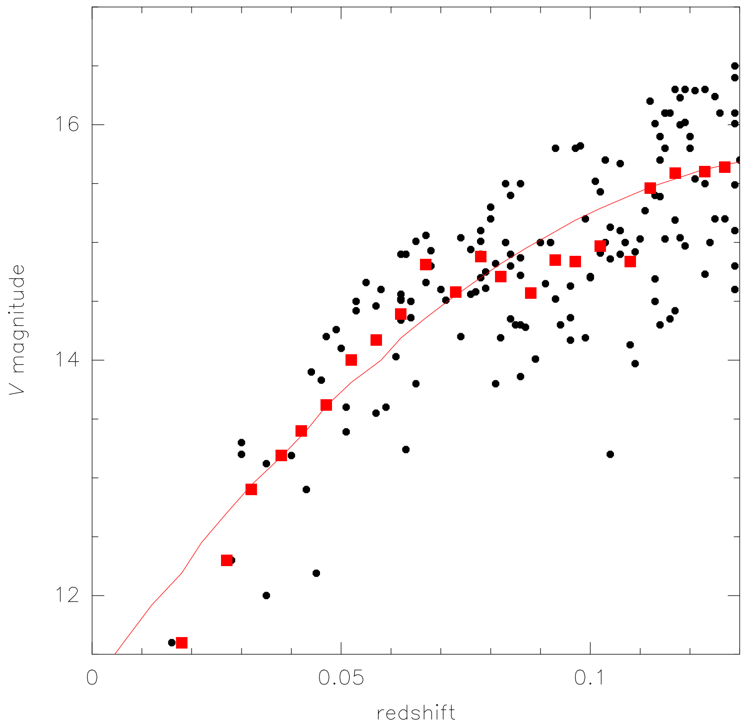

Figure 20: The V magnitude as a function of redshift for the QSOs with $z < 0.13$ and an available V measure as listed in the Véron-Cetty & Véron (2006) catalogue of AGNs. The average luminosity per 0.05 redshift bin is displayed as squares. The solid line shows a maximum likelihood fit with a 2nd-order polynomial.

Then, each QSO has been affected with a C $= (B - V)$ colour index according to the mean $\bar{\mathrm{C}} = 0.40$ and rms $\sigma_{\mathrm{C}} = 0.08$ values computed on the 19 objects with a colour measurement. This allowed us to compute Gaia magnitudes using the colour transformation given in Table 1 of Jordi (2007):

$$\begin{aligned}
V - G &= +0.0124 \\
&\quad +0.2852 \times \bar{\mathrm{C}} + 0.0391 \times \bar{\mathrm{C}}^2 + 0.0906 \times \bar{\mathrm{C}}^3 \\
V - G_{rvs} &= +0.0976 \\
&\quad +1.5108 \times \bar{\mathrm{C}} - 0.1130 \times \bar{\mathrm{C}}^2 + 0.0832 \times \bar{\mathrm{C}}^3 \\
V - BP &= +0.0181 \\
&\quad -0.3625 \times \bar{\mathrm{C}} + 0.1679 \times \bar{\mathrm{C}}^2 - 0.0053 \times \bar{\mathrm{C}}^3 \\
V - RP &= +0.0431 \\
&\quad +1.2506 \times \bar{\mathrm{C}} - 0.1743 \times \bar{\mathrm{C}}^2 + 0.1120 \times \bar{\mathrm{C}}^3
\end{aligned} \tag{2}$$

8 Astrometric parameters

8.1 Positions

The equatorial coordinates α and δ have been generated from a uniform drawing on the sphere in each of the sub-populations defined by its redshift. No screening has been applied in the vicinity of the galactic plane since this will result directly from the application of the absorption and reddening model at a later stage, when the catalogue is used to generate the observations.

8.2 Distances

Distance indicators can be derived for each object from its redshift value by specifying a cosmological model. Each of these sources lying at cosmological distances, a zero parallax has been assigned to all of them. However to avoid possible overflow/underflow problems with a true zero parallax we have actually set all the parallax to 10^{-6} mas, equivalent to an Euclidean distance of about 1 Gpc.

8.3 Proper motions

This is the only critical issue for the astrometric parameters. In principle distant sources are assumed to be co-moving with the general expansion of the distant Universe and have no transverse motion. If the transverse displacement were comparable to that in the radial direction (a totally unrealistic assumption) this would yield transverse motion of the order of the Hubble constant or ≈ 10 μas/yr. However the observer is not at rest with respect to the distant Universe and the accelerated motion around the galactic center, or more generally that of the Local group toward the Virgo cluster is the source of a spurious proper motion with a systematic pattern. This has been discussed in many places Kovalevsky (2003), Mignard (2005) and the underlying physics is not repeated here.

Eventually the effect of the acceleration (centripetal acceleration of the solar system) will show up as a small proper motion of the quasars, or stated differently we will see the motion of the quasars on a tiny fraction of it aberration ellipse whose period is 250 million years. Mathematically this can be modeled as,

$$\delta \boldsymbol{u} = \frac{\boldsymbol{v}}{c} - \frac{\boldsymbol{v} \cdot \boldsymbol{u}}{c} \boldsymbol{u}, \tag{3}$$

where $\boldsymbol{u}$ is the unit vector in the direction of a quasar and $\delta \boldsymbol{u}$ the change in direction due to galactic aberration caused by the velocity $\boldsymbol{v}$ in the galactic frame. What is actually observable is the change in this aberration as,

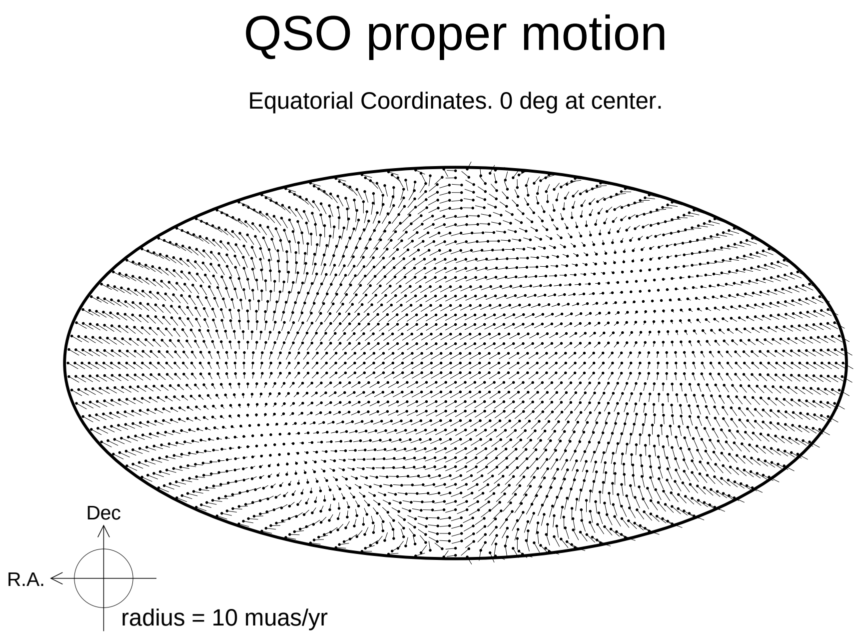

Figure 21: The regular QSO proper motion field resulting from the accelerated motion of the solar system around the galactic center. The diagram is in equatorial coordinates.

$$\frac{d(\delta \boldsymbol{u})}{dt} = \frac{\boldsymbol{\Gamma}}{c} - \left(\frac{\boldsymbol{\Gamma}}{c} \cdot \boldsymbol{u}\right) \boldsymbol{u}, \tag{4}$$

where $\boldsymbol{\Gamma} = d\boldsymbol{v}/dt$ is the acceleration of the solar system barycentre with respect to the extra-galactic sources.

Eq. (4) leads to the two scalar equations (per star) relating the components of the proper motion to the cartesian components of the acceleration,

$$\mu_\alpha \cos\delta = -\frac{\Gamma_x}{c}\sin\alpha + \frac{\Gamma_y}{c}\cos\alpha \tag{5}$$

$$\mu_\delta = -\frac{\Gamma_x}{c}\sin\delta\cos\alpha - \frac{\Gamma_y}{c}\sin\delta\sin\alpha + \frac{\Gamma_z}{c}\cos\delta \tag{6}$$

To introduce this effect in the simulated QSO data one has to choose the vector $\boldsymbol{\Gamma}$ both in magnitude and direction. Taking as order of magnitude the centripetal acceleration of the motion of the solar system around the galactic center, this gives $\Gamma \simeq v_g^2/R$ and then $\Gamma/c \simeq 4\mu$as per year. Regarding the orientation we took the acceleration roughly directed toward the galactic center, using at the end round numbers in the angles. This gives for the direction of $\boldsymbol{\Gamma}$: $\alpha_\Gamma = 270\,$deg, $\delta_\Gamma = -30\,$deg. Taking $\Gamma/c = 4\mu$as/yr this leads to the numerical expressions that have been used to generate the proper motions components in mas/yr as,

$$\mu_\alpha^\star = -3.464 \times 10^{-3} \cos\alpha \quad (7)$$
$$\mu_\delta = 3.464 \times 10^{-3} \sin\delta \sin\alpha - 2.0 \times 10^{-3} \cos\delta \quad (8)$$

The regular streaming motion expressed in Eqs. 7-8 is illustrated in equatorial and galactic coordinates [2] respectively in Fig. 21 and 22, with one average value on every square of 4×4 deg^2. The regular pattern is obvious in galactic coordinates.

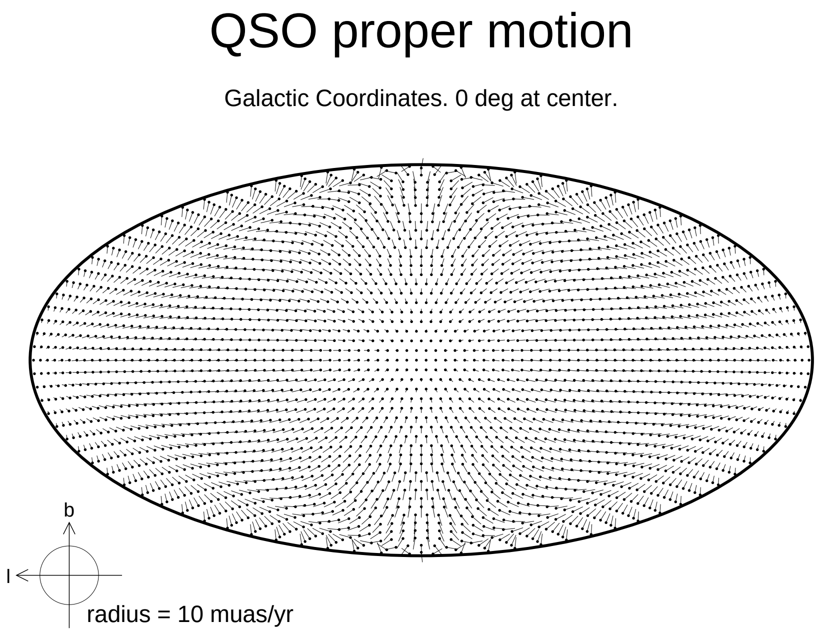

Figure 22: The regular QSO proper motion field resulting from the accelerated motion of the solar system around the galactic center. The diagram in galactic coordinates shows the systematic stream along the direction of the galactic center.

8.4 Radial velocity

There is no radial velocity included in the simulated data but a value for the redshift z. The computation of the astrometric position must be done by using $v_r = 0$.

[2]Plots based on a program initially written by L.Lindegren and maintained by F. Mignard

9 File properties

Two catalogues of QSOs are provided for two different limiting magnitudes in the G-band. The main catalogue includes 1 741 125 entries with $0 < z < 5$ down to $G = 20.5$. A subsample with only QSOs brighter than $G = 19.5$ lists 683 049 objects. The surface densities are 42.2 obj/sq.deg and 16.6 obj/sq.deg, respectively. Another catalogue of $\sim$ 68 000 objects including only one object every ten with $G < 19.5$ is also available for test purposes on a smaller file.

The catalogue format is given in Table 1.

Table 2: Field description and format

Column	C1	C2	Format	Description
1	2	10	I9	identification number ([1])
2	13	23	F11.7	R.A.(ICRF) in degrees
3	26	36	F11.7	Dec.(ICRF) in degrees
4	40	44	F5.3	Redshift
5	48	53	F6.3	G magnitude
6	56	61	F6.4	BP magnitude
7	64	69	F6.4	RP magnitude
8	72	77	F6.4	Grvs magnitude
9	80	86	E7.1	Parallax : set to 10^{-6} mas
10	89	99	E11.4	Proper motion : RA component in mas/yr ([2])
11	102	112	E11.4	proper motion : Dec component in mas/yr ([2])

[1] Negative values are for objects taken from the Veron-Cetty Catalogue (12th Ed.)

[2] see Sec. 8.3

Sample records

```
 1   16.0712370  67.3319848 0.423 18.275  18.538  17.809  17.480  0.1E-05 -0.3329E-02  0.1141E-03
 3  160.0863483 -16.9640052 0.669 19.171  19.314  18.875  18.686  0.1E-05  0.3257E-02 -0.2257E-02
 4  306.9330511 -24.3646077 0.307 17.749  17.939  17.378  17.169  0.1E-05 -0.2081E-02 -0.6796E-03
 5  175.5820110 -33.0187266 0.595 18.231  18.416  17.870  17.628  0.1E-05  0.3454E-02 -0.1822E-02
 7  293.0979138 -34.5182064 0.388 19.425  19.482  19.305  19.310  0.1E-05 -0.1359E-02  0.1577E-03
12  114.6194549  55.8740331 0.512 18.513  18.719  18.128  17.842  0.1E-05  0.1443E-02  0.1485E-02
13   47.4412421 -61.8972724 0.386 19.172  19.268  18.985  18.993  0.1E-05 -0.2343E-02 -0.3193E-02
17   46.8853712 -16.4128615 0.614 18.451  18.600  18.145  17.940  0.1E-05 -0.2368E-02 -0.2633E-02
```

References

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

Jordi C., 2007, Photometric relationships between Gaia photometry and existing photometric systems, Tech. rep., UB, gAIA-C5-TN-UB-CJ-041-1

Kovalevsky J., 2003, A&A, 404, 743

Mignard F., 2005, In: Turon C., O'Flaherty K., Perryman M. (eds.) Proc. of The Three Dimensional Universe with Gaia, 5–14

Richards G.T., Strauss M.A., Fan X., et al., 2006, AJ, 131, 2766

Schneider D.P., Hall P.B., Richards G.T., et al., 2007, AJ, 134, 102

Vanden Berk D.E., Richards G.T., Bauer A., et al., 2001, AJ, 122, 549

Véron-Cetty M.P., Véron P., 2006, A&A, 455, 773