THE GAIA REFERENCE FRAME : HOW TO SELECT THE SOURCES

F. MIGNARD
OCA/Cassiopée
Observatoire de la Côte d'Azur, BP. 4229, 06304 NICE CEDEX 4
francois.mignard@obs-nice.fr

ABSTRACT. The European space astrometry mission Gaia is scheduled for a launch in 2011 and aims to produce a complete sky survey down to $V = 20$ with an astrometric accuracy of 10 μas at $V = 15$. During its 5-year mission the satellite will also repeatedly measure the position of $\sim$500,000 quasars in a consistent way, leading to a direct realisation of the primary inertial frame in the visible in the framework the ICRS concepts. At $V = 20$ the sky density of the QSOs is about 1000 times smaller than that of the stars at mid galactic latitude, and given their number and their stellar-like images, this implies the construction an automatic recognition scheme of the non stellar sources with a sensitivity of the order of one part in a thousand. In this paper I present the expected performance for the realisation of the reference frame and discuss the procedure under development to select the primary sources.

1. PRESENTATION

Gaia will provide astrometric and photometric observations for the quasars (QSOs) at $G < 20$ mag over the whole sky, 5 times more than the number expected from the Sloan Digital Sky Survey. This will be the first all-sky, flux-limited survey to $V = 20$ of the extragalactic sources. Although there is no global survey of quasars available at the moment, local surveys indicate that the typical surface density of quasars is about 20 sources $V < 20$ per square degree, much smaller than that of the stars at any galactic latitude. So at the end one may reasonably expect a census of about 500,000 quasars at galactic latitudes $|b| > 20^\circ - 25^\circ$. Closer to galactic plane, Gaia faces two difficulties: (i) the galactic extinction and reddening that will block off the light of these distant and rather faint sources, (ii) the difficulty to discriminate between the stars as their relative density to that of the quasars increases drastically at low galactic latitudes (this ratio is about 10,000 at $b = 10^\circ$ and $G = 19$).

The extensive zero-proper motion survey will provide a direct realization of the quasi-inertial celestial reference frame with a residual rotation less than 0.5 $\mu\text{as}\,\text{yr}^{-1}$ (Fig. 1) and a space density at least 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. The random instability of the sources puts a serious limitation in the ultimate precision of the inertial frame and imposes a strict selection of the primary sources, among a population of quasars already free of stellar contaminants. I discuss in the following sections the basic properties of the methods that has been considered for Gaia.

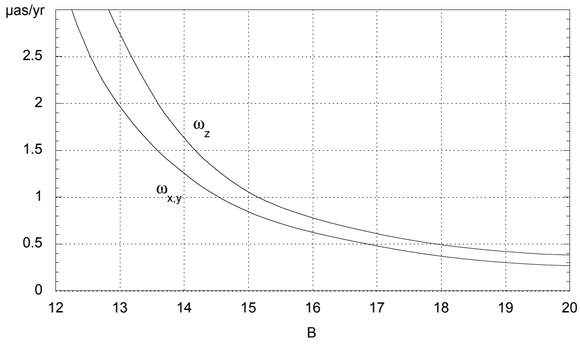

Figure 1: Precision of the spin rate of the inertial frame achievable with Gaia from the observations of the quasars. This is based on a simulation of 42,000 known sources (10,000 with $B < 18$) whose distribution with galactic latitude to that expected from Gaia. The precision read for a B magnitude is computed with only sources brighter than B selected. Here galactic coordinates have been used and the random instability has been taken equal to 20 μas yr^{-1}. Though there are many more faint objects than bright ones, the frame is primarily determined by the brightest sources, because of their astrometric precision. A similar situation should prevail in the real mission.

2. SELECTION OF KNOWN SOURCES

The way extragalactic sources are selected with Gaia depends on the science objective : to realise the survey one aims to generate a set of sources including most of the QSOs, accepting that the set may contain a significant number of stellar contaminants, to be removed later. Regarding the reference frame it is obvious that the selection must end with a very clean sample, probably much smaller, but without contaminant. In addition even among the well selected sources, many of them will not pass subsequent tests to be included in the list of primary sources to build the inertial frame, in case unaccounted proper motions are found or when a strong photometric instability will make the astrometric stability of the same sources questionable.

There are several possibilities available to create this clean sample where “clean” means that the sources are all extragalactic without contaminant. Hereafter the final selection of the most suitable sources for the reference frame is not considered and deferred to a further investigation. The number of sources is an important criteria, but not critical as long as $\sim$ 5000–10000 are available. However the space distribution is a key factor : we have many new sources with the SDSS, but primarily located around the galactic cap, definitely an undesirable feature to build a rigid reference frame. Thus one can :

- use the ICRF sources brighter than $G = 20$;
- take the sources found in ground-based surveys of QSOs;
- rely on the Gaia internal recognition scheme.

These three means will correspond to different size of the sample, and also to different risks of introducing stellar contaminants. Using the ICRF sources (primary sources and the others) is an obvious starting solution and in practice a requirement. Out of the $\sim$ 700 sources, there are $\sim$ 400 (resp. 150 primaries) off from the galactic plane by more than 20° and with V magnitude < 20 (Fig. 2) that will be well detected by Gaia.

Most of these sources are quasars, and a good 50% are brighter than $V = 18$ and will be observed with an astrometric accuracy of about 50 μas. The ICRS paradigm will be applied to eliminate any global rotation from the preliminary frame and then correct all the observations

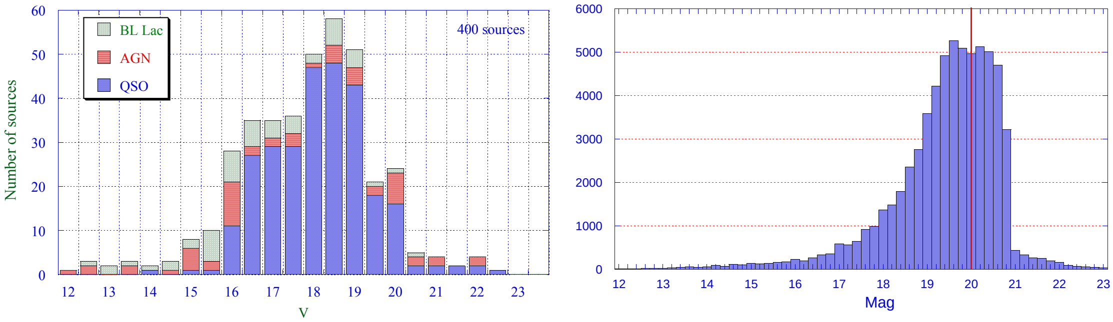

Figure 2: Magnitude distribution of the ICRF sources (left) and of the extragalactic sources in the Véron-Cetty & Véron Catalogue. Only the sources outside the galactic plane and brighter than $V = 20$ have been considered.

for this spin. It must be clear that the Gaia solution will be autonomous and not linked in any way to the ICRF, unlike what has been done for Hipparcos. The primary sources of the ICRF will be used as control : the residual rotation exhibited by these sources should be compatible with the ICRF uncertainty, that is to say less than 10 μas yr^{-1}. The origin of the right ascension is totally arbitrary and will be ascertained by constraining the orientation of the frame to be similar to that of the current ICRF, ensuring in this way the continuity between old and new standards. A global rotation will be fitted to the positional differences for the ~ 150 primary sources observable with Gaia to determine the three angles $\epsilon_x, \epsilon_y, \epsilon_z$. The final uncertainty will come from the $\sim 0.3 - 0.4$ mas positional error in the radio ICRF, much larger than Gaia's. Eventually the orientation should be the same as the ICRF axes within ~ 50 μas.

The second source of QSOs that Gaia could use, even if the internal recognition scheme fails, will come from the ground-based observations. At the moment the systematic compilation of QSO is available with the 11th edition of the Catalog of Quasars and Active Galactic Nuclei by Véron-Cetty and Véron [1]. It includes 49,000 quasars brighter than $M_B = -23$ and 15,000 AGNs fainter than $M_B = -23$, although this distinction is purely historical and no longer relevant. In total one finds 42,000 sources brighter than 20th magnitude (Fig. 2, right) and potentially measurable with Gaia.. However, given the nature of the compilation, a fraction of theses sources will be eliminated as not being really extragalactic, either from the astrometry or from the internal photometric detection. Once the Catalogue is cleaned of its contaminants, a more severe selection will be applied to the remaining sources to qualify them to become primary sources. By the time Gaia will fly, new versions of this Catalogue will be issued to incorporate the results of the SDSS, whose 3rd data release has produced more than 50,000 quasars identified with spectroscopy [2]. Virtually all these sources are brighter than $g = 20.2$ and will be seen by Gaia. This again will constitute a excellent subset to calibrate the Gaia classification procedure together with the photometric determination of the redshift.

3. RECOGNITION OF NEW SOURCES

The major problem in any QSO survey is the recognition among many starlike sources of the rare quasars. The ratio between the number of stars and of quasars per square degree is a function of the magnitude and of the galactic latitude. The quasars being very distant extragalactic objects are basically uniformly distributed on the sky in contrast with the stars primarily concentrated in the galactic plane and in the direction of the galactic center. As can be seen in Table 1 the extent of the problem is not the same outside the galactic plane as very close to this plane. Fortunately the quasar population grows faster than that of the stars toward faint magnitudes, hence the proportion of quasars increases with fainter sources. Find a quasar out of 100 stellar contaminants at high galactic latitude does not imply too powerful tests, while doing the same for 100 times more stars at low galactic latitude is very challenging and not yet fully solved.

Table 1: Relative surface density of stars and quasars for different magnitude and galactic latitude. The numbers are $N_\star/N_{\rm qso}$

b(°)	G=18	G=19	G=20
90	725	125	75
60	1150	280	120
30	3075	765	330
10	33,000	10,500	5600

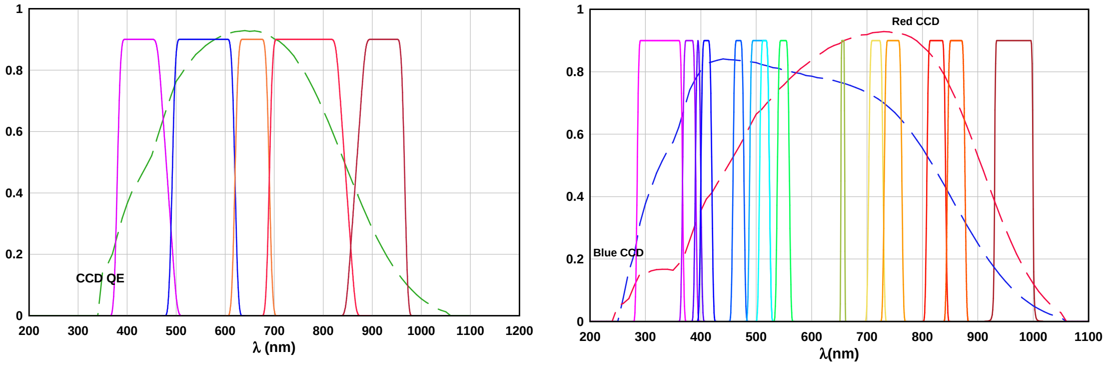

Figure 3: The two photometric systems to be used by Gaia. The left panel shows the five filters of the broad band photometre placed on fours strips of CCD at the edge of the astrometric fields. The right panel corresponds to the filters of the medium band photometre working with an independent telescope and different detector.

There are several indicators that will help make the decision, like the absence of parallax, the negligible proper motion (in an inertial frame), the short term variability and, before all, the spectral signature summarised in a set of color indices. Given the astrometric capabilities of Gaia it is highly advisable that the parallaxes and proper motions should be taken as good indicators only to exclude an object when its distance or motion are definitely not compatible with that expected for an extragalactic source : no quasar will have a significant parallax or a large proper motion. However some stars will have a negligible parallax or proper motion making this criteria not powerful to discriminate between the two populations. In addition with the astrometric accuracy of Gaia there are important science issues in trying to detect quasars with non-zero proper motion, for example to determine the acceleration induced by the galactic rotation or the accelerated motion of the Local group with respect to the distant Universe. Variable macro lensing on quasars may also be a source of spurious proper motions, that will deserve specific treatment. So these tests should only be used as confirmation tool, when the photometric testing has concluded positively that a particular source is likely to be a QSO. Objects selected as quasars from their photometric signature and not passing the final astrometric tests, will be looked at carefully for possible multi-images quasars, with relevant cosmological implications.

The photometric testing with Gaia rests upon the measurements to be carried out with the BBP (broad-band photometre) and MBP (medium-band photometre). The current photometric system is shown in Fig. 3. The choice is the result of many optimisations, primarily based on stellar physics to retrieve the fundamental stellar parameters for every kind of stars. For the QSO photometric testing, the most important point was to avoid large gaps between filters. The standard energy distribution of a quasar is markedly different for that of a star, having a large UV flux and a regular decrease in the continuum (in the rest frame). There are strong and broad emission lines like Lyα, SI IV, O IV, CIV, not seen in normal stars. Due to the redshift the spectrum will be quite often displaced in the visible and strongly squeezed. Gaia will observe repeatedly each quasar and determine its location in the multi-dimensional color space. In this space they should occupy a different location from that of the stars, including white dwarfs. The difficult cases appear with very reddened stars : but this will mean strong absorption in the vicinity of the galactic plane, where QSOs are also reddened and difficult to observe and distinguish among the star filled regions.

Claeskens and collaborators [3] have generated a large library of synthetic spectra for the quasars, covering a wide range of redshifts, reddening and line strengths. Using in parallel spectral libraries for stars of every spectral class and type, binaries stars and white dwarfs, they have correlated all these energy distributions with the different filters proposed for Gaia. Tests were conducted with a set of templates to simulate the observations with realistic noises. From the location of each source in the colour space they have shown conclusively that the combination of the broad band and medium band filters permits a photometric identification of quasars from virtually any type of stars. The photometric redshift ($\sim$ redshift seen in the continuum) can be retrieved as well with an accuracy of few percents, even if the actual spectral index differs significantly from that of the template used for the analysis.

The Gaia internal autonomous multi-color detection will be very efficient to get rid of the traditional contaminants like the white dwarfs and the very reddened stars and will permit the selection of a 99.6% star-free sample of QSOs. At $G = 19$, 85% of the QSOs are correctly identified with a level of stellar contaminants $< 0.01\%$ (0.4% at $G = 20$) [3]. These contaminants are not uniformly distributed in the colour space, but are associated to extreme reddening or low metallicity. By reducing the level of completeness of the QSO sample it is possible to virtually eliminate any contaminants. Such a clean sample should contain about 15% of the QSO population, nonetheless amounting to more than 50,000 sources, largely enough to tie the Gaia astrometric solution to the non-rotating Universe.

4. TESTING

As I have mentioned earlier in each range of magnitude the field stars outnumber the QSOs making the recognition of the latter difficult. Any recognition criterion will be probabilistic with a risk of ending with a wrong conclusion as to the true nature of the object examined. In this kind of testing, like in medical testing of the efficiency of a treatment, one has two different kinds of errors, each quantified by a probability :

- A QSO may be classified as a star, that is to say be unrecognized by the test. As a result, if the test is not sensitive enough, the survey will be incomplete and many QSOs will be left out. Quantitatively this is determined by the risk of the test α or equivalently by its significance level $1 - \alpha$. The risk of rejecting genuine QSOs from the sample can be made very small by decreasing α, but this will increase the fraction of stars in the samples.

- On the opposite a star may be wrongly taken as a QSO and flagged as such. In this case the sample of supposed QSOs will be contaminated by genuine stars. This is represented by the risk of contamination β and its complement $1-\beta$ is called the power of the test. Since the stars are much more frequent than the QSO, even if this risk has a low probability, this may end up with a large number of contaminants in the selected sample.

The overall performance is easily modelled by introducing the frequency f of QSOs among the celestial sources detectable by Gaia in a given region of the sky. The meaning of the risks is summarized in Table 2 where the columns refer to the true nature of the source and the lines to the conclusions produced by the tests.

Table 2: Probabilities of wrongly labelling a star or a QSO with the recognition tests.

		The object is a :	
		QSO	star
Found as :	QSO	$1-\alpha$	β
	star	α	$1-\beta$

What matters now is the probability of making a correct decision : if a source has been found as being a QSO, what is the probability that this conclusion is correct. Application of the elementary laws of probability yields the a posteriori probability that a source being found as a QSO is in fact a QSO (right) or a star (wrong decision) (FQ : means found to be a QSO) :

$$P(Q/FQ) = \frac{f(1-\alpha)}{f(1-\alpha)+(1-f)\beta} \sim \frac{f}{f+\beta} \tag{1}$$

$$P(S/FQ) = \frac{(1-f)\beta}{f(1-\alpha)+(1-f)\beta} \sim \frac{\beta}{f+\beta} \tag{2}$$

Therefore, as expected, if $f \ll \beta$ then $P(S/FQ) \sim 1$ and most purported to be quasars are in fact ordinary stars. So the testing must be such that the risk β to misclassify a star should be much less than the frequency of QSO among the stars. The efficiency parameter (proportion of QSO in the selected sample) is given approximately by $f/(f+\beta)$. For the sake of illustration consider a sample of 10,000 sources with a 0.1 percent error selection ($\beta = 0.001$) down to the Gaia limiting magnitude $G = 20$. When these sources are classified as QSOs the actual number of quasars and stars in this set is shown in Table 3 as a function of the galactic latitude, where f has been taken from Table 1. Therefore, with the typical quasar density of 20 QSOs per square degree brighter than $B = 20$, the realisation of a clean sample of sources needed for the reference system, just based on the Gaia internal detection system, is very challenging at low galactic latitude. Combined with the absorbtion and the reddening of the sources this makes very likely that the Gaia inertial frame will show a deficiency of sources around the galactic plane.

5. PARALLAXES AND PROPER MOTIONS

So far I have shown that the Gaia photometric detection alone will bring a large number of QSOs without stellar contaminants. The astrometry will be used to reject obvious non stellar object from this sample, but also from the larger sample planned for the QSO survey. While the selected power of the photometric tests for the clean sample implies that very few sources will not be QSOs this is not the case for the survey sample, which will contain most of the

Table 3: Number of quasars and stars in a set of 10,000 sources categorised as quasars with a test having a risk $\beta = 0.001$. b is the galactic latitude in degrees.

b (°)	QSOs	Stars
90	9300	700
60	8900	1100
30	7500	2500
10	1500	8500

QSOs, but also many contaminants. So additional testing based on the proper motions and parallaxes will be useful to eliminate the most conspicuous contaminants. For example late type red stars or white dwarfs will show up with large parallaxes and proper motions ruling out the QSO identification.

The testing is based on the null hypothesis H_0 that extragalactic sources have zero parallax and zero proper motion, or at least $\mu \ll 1\mu\mathrm{as\,yr}^{-1}$. If the measurement error is σ_μ on each component $\mu_\alpha^\star$ and μ_δ, with a gaussian distribution, one has under H_0 the probability distribution of the magnitude μ of the proper motion,

$$P(\mu/\sigma_\mu) < x = \exp(-x^2/2) \tag{3}$$

and then with the risk α one can reject the null hypothesis whenever

$$\mu > (-2\log\alpha)^{1/2}\,\sigma_\mu \tag{4}$$

and similarly for the parallax, one would have a rejection if,

$$\pi > \Phi^{-1}(1-\alpha)\,\sigma_\pi \tag{5}$$

where $\Phi(x)$ is the cumulative probability function of the normal law. The test is one-sided because all the alternative hypotheses have positive parallaxes or proper motions. For faint objects the threshold of rejection is given in Table 4 as a function of magnitude and risk. For example the test will reject with a risk 0.1% photometrically selected quasars of magnitude 19 if the proper motion measured by Gaia is larger than 223 $\mu\mathrm{as\,yr}^{-1}$.

The power of the test $(1-\beta)$ can be established only if we know something about the alternative possibilities, that is to say the distribution of proper motions and parallaxes for every other sources which are not quasars. From the values in Table 4 the prospect of testing does not appear very good : if stars with parallax $< 600\,\mu$as or proper motion $< 400\,\mu\mathrm{as\,yr}^{-1}$ are common among the field stars, the test will be very poor to select QSOs, but at the level of significance $1-\alpha$ good to reject nearby white dwarfs or very red stars. Consider in more detail separately the case of a test built on the parallaxes from that on the proper motions.

- Gaia will observe stars in the Milky way all the way through the disk and the halo and in few external galaxies like the Magellanic clouds and Andromeda. Unless a quasar is precisely observed in the same direction as the LMC or M31 the largest distance of the field stars will be less than 20 kpc, yielding a parallax of 50 μas. Most of the stars will have a parallax larger than this level, but the important point is that a non negligible population will have a parallax of this order of magnitude. It is clear that the probability

Table 4: Threshold of rejection for the proper motion or parallax of photometrically selected quasars as a function of their magnitude and of the risk α.

	Proper Motions				Parallaxes			
V	σ_μ		α		σ_π		α	
mag	μas yr^{-1}	0.01	0.005	0.001	μas	0.01	0.005	0.001
			μas yr^{-1}				μas	
18	35	106	115	130	65	151	168	201
19	60	182	195	223	125	292	323	386
20	115	350	375	430	230	536	592	711

distributions of the measured parallaxes under H_0 and that of the possible alternative hypotheses overlap significantly, meaning that the parallax test is virtually powerless to reject non-QSO sources from the sample. It is just efficient to reject non-QSO sources with totally anomalous parallaxes, as it can be seen in the last three columns of Table 4, corresponding to stellar sources at distances of 2 to 5 kpc.

- For the proper motions the situation is much more promising. Excluding again isolate stars in external galaxies which will be limited in number, one restricts to the field stars of the Milky Way. The proper motions follow from the combination of the relative motion due to the galactic rotation and that of the proper velocity–with magnitude equal to the velocity dispersion– of each star in its own local standard of rest. There are essentially three categories : (i) very nearby stars at less than 1 kpc, with typical velocity of 20 km s^{-1}, giving $\mu \sim 5\,\mathrm{mas\,yr}^{-1}$; (ii) then the stars up to 3 to 5 kpc, where the shear motion of the Galaxy is the dominant factor, with the kinematics well described by the Oort constants. With $A = 15$ km s^{-1} kpc^{-1}, this give $\mu \sim 3\,\mathrm{mas\,yr}^{-1}$. (iii) The very distant stars are in the last group with $d \geq 10$ kpc and a relative velocity comparable to that of the Sun around the galactic center, $V \sim 200$ km s^{-1}, yielding $\mu \sim V/d \sim 4$ mas yr^{-1}. Therefore the range of proper motions for stars not rejected by the parallax test is rather limited and in fact fairly constant for the field stars. No need to use sophisticated probabilistic argument to notice that the two probability distributions have virtually no overlapping, or equivalently that the power of the test will be very high (provided the assumption about the stellar proper motions is correct).

A the time of writing the principles laid out above for the astrometric tests have not yet been implemented in the simulation. This should be done in 2005 in complement to the photometric selection, in order to assess the efficiency and to identify potential problems.

6. REFERENCES

[1] Véron-Cetty, M.P., Véron, P., 2003, A&A 412, 399-403.

[2] Abazajian, K. et al., 2004, AJ in press.

[3] J.-F. Claeskens, A. Smette, J.Surdej, 2005, “Quasars with Gaia: Identification and Astrophysical Parameters” in: Proceedings of the Symposium “The Three-Dimensional Universe with Gaia”, ESA SP-576, in press.