1. Introduction

One of the key science objectives of the European Space Agency’s Gaia mission (Gaia Collaboration et al. 2016) is to build a rotation-free celestial reference frame in the visible wavelengths. This reference frame, which may be called the Gaia Celestial Reference Frame (Gaia-CRF), should meet the specifications of the International Celestial Reference System (ICRS; Arias et al. 1995) in that its axes are fixed with respect to distant extragalactic objects, that is, to quasars. For continuity with existing reference frames and consistency
across the electromagnetic spectrum, the orientation of the axes should moreover coincide with the International Celestial Reference Frame (ICRF; Ma et al. 1998) that is established in the radio domain by means of VLBI observations of selected quasars.

The second release of data from Gaia (Gaia DR2; Gaia Collaboration 2018) provides complete astrometric data (positions, parallaxes, and proper motions) for more than 550,000 quasars. In the astrometric solution for Gaia DR2 (Lindegren 2018), subsets of these objects were used to avoid the rotation and align the axes with a prototype version of the forthcoming third realisation of the ICRF\(^2\). The purpose of this paper is to characterise the resulting reference frame, Gaia-CRF2, by analysing the astrometric and photometric properties of quasars that are identified in Gaia DR2 from a positional cross-match with existing catalogues, including the ICRF3-prototype.

Gaia-CRF2 is the first optical realisation of a reference frame at sub-milliarcsecond (mas) precision, using a large number of extragalactic objects. With a mean density of more than ten quasars per square degree, it represents a more than 100-fold increase in the number of objects from the current realisation at radio wavelengths, the ICRF2 (Fey et al. 2015). The Gaia-CRF2 is bound to replace the HIPPARCOS Celestial Reference Frame (HCRF) as the most accurate representation of the ICRS at optical wavelengths until the next release of Gaia data. While the positions of the generally faint quasars constitute the primary realisation of Gaia-CRF2, the positions and proper motions of the \(\sim 1.3\) billion stars in Gaia DR2 are nominally in the same reference frame and thus provide a secondary realisation that covers the magnitude range \(G \approx 6\) to 21 mag at similar precisions, which degrades with increasing distance from the reference epoch J2015.5. The properties of the stellar reference frame of Gaia DR2 are not discussed here.

This paper explains in Sect. 2 the selection of the Gaia sources from which we built the reference frame. Section 3 presents statistics summarising the overall properties of the reference frame in terms of the spatial distribution, accuracy, and magnitude distribution of the sources. The parallax and proper motion distributions are used as additional quality indicators and strengthen the confidence in the overall quality of the product. In Sect. 4 the optical positions in Gaia DR2 are compared with the VLBI frame realised in the ICRF3-prototype. A brief discussion of other quasars in the data release (Sect. 5) is followed by the conclusions in Sect. 6.

2. Construction of Gaia-CRF2

2.1. Principles

Starting with Gaia DR2, the astrometric processing of the Gaia data provides the parallax and the two proper motion components for most of the sources, in addition to the positions (see Lindegren 2018). As a consequence of the Gaia observing principle, the spin of the global reference frame must be constrained in some way in order to deliver stellar proper motions in a non-rotating frame. Less relevant for the underlying physics, but of great practical importance, is that the orientation of the resulting Gaia frame should coincide with the current best realisation of the ICRS in the radio domain as well as possible, as implemented by ICRF2 and soon by ICRF3.

These two objectives were achieved in the course of the iterated astrometric solution by analysing the provisional positions and proper motions of a pre-defined set of sources, and by adjusting the source and attitude parameters accordingly by means of the so-called frame rotator (Lindegren et al. 2012). Two types of sources were used for this purpose: a few thousand sources identified as the optical counterparts of ICRF sources were used to align the positions with the radio frame, and a much larger set of probable quasars found by a cross-match with existing quasar catalogues were used, together with the ICRF sources, to ensure that the set of quasar proper motions was globally non-rotating. The resulting solution is then a physical realisation of the Gaia frame that is rotationally stabilised on the quasars. The detailed procedure used for Gaia DR2 is described in Lindegren (2018).

2.2. Selection of quasars

Although Gaia is meant to be autonomous in terms of the recognition of quasars from their photometric properties (colours, variability), this functionality was not yet implemented for the first few releases. Therefore the sources that are currently identified as quasars are known objects drawn from available catalogues and cross-matched to Gaia sources by retaining the nearest positional match. In Gaia DR1, quasars where flagged from a compilation made before the mission (Andrei et al. 2014), and a subset of ICRF2 was used for the alignment. The heterogeneous spatial distribution of this compilation did not greatly affect the reference frame of Gaia DR1 because of the special procedures that were used to link it to the HCRF (Lindegren et al. 2016; Mignard et al. 2016).

For Gaia DR2, which is the first release that is completely independent of the earlier HIPPARCOS and Tycho catalogues, it was desirable to use the most recent VLBI positions for the orientation of the reference frame, and a large, homogeneous set of quasars for the rotation. The Gaia data were therefore cross-matched with two different sets of known quasars:

- A prototype of the upcoming ICRF3, based on the VLBI solution of the Goddard Space Flight Center (GSFC) that comprises 4262 radio-loud quasars that are observed in the X (8.5 GHz) and S (2.3 GHz) bands. This catalogue, referred to here as the ICRF3-prototype, was kindly provided to the Gaia team by the IAU Working Group on ICRF3 (see Sect. 4) more than a year in advance of the scheduled release of the ICRF3. The positional accuracy is comparable to that of Gaia, and this set was used to align the reference frame of Gaia DR2 to the radio frame.

- The all-sky sample of 1.4 million active galactic nuclei (AGNs) of Secrest et al. (2015), referred to below as the AllWISE AGN catalogue (AW in labels and captions). This catalogue resulted from observations by the Wide-field Infrared Survey Explorer (WISE; Wright et al. 2010) that operates in the mid-IR at 3.4, 4.6, 12, and 22 \(\mu\)m wavelength. The AllWISE AGN catalogue has a relatively homogenous sky coverage, except for the Galactic plane, where the coverage is less extensive because of Galactic extinction and confusion by stars, and at the ecliptic poles, which have a higher density because of the WISE scanning law. The sources are classified as AGNs from a two-colour infrared photometric criterion, and Secrest et al. (2015) estimated that the probability of stellar contamination is \(\lesssim 4.0 \times 10^{-7}\) per

---

\(^1\) “Rotation” here refers exclusively to the kinematical rotation of the spatial axes of the barycentric celestial reference system (BCRS), as used in the Gaia catalogue, with respect to distant extragalactic objects (see e.g. Klioner & Soffel 1998). Similarly, “orientation” refers to the (non-) alignment of the axes at the reference epoch J2015.5.
About half of the AllWISE AGN sources have an optical counterpart that is detected at least once by Gaia in its first two years.

Cross-matching the two catalogues with a provisional Gaia solution and applying some filters based on the Gaia astrometry (see Sect. 5.1, Eq.(13) in Lindegren 2018) resulted in a list of 492,007 putative quasars, including 2844 ICRF3-prototype objects. The filters select sources with good observation records, a parallax formal uncertainty \( \sigma_{\pi} < 1 \) mas, a reliable level of significance in parallax and proper motion, and they avoid the Galactic plane by imposing \( |\sin b| > 0.1 \). These sources were used by the frame rotator, as explained above, when calculating the final solution; in Gaia DR2, they are identified by means of the flag frame_rotator_object_type. This subset of (presumed) quasars cannot, however, be regarded as a proper representation of Gaia-CRF2 because of the provisional nature of the solution used for the cross-matching and the relatively coarse selection criteria. Several of these sources were indeed later found to be Galactic stars.

A new selection of quasars was therefore made after Gaia DR2 was completed. This selection took advantage of the higher astrometric accuracy of Gaia DR2 and applied better selection criteria that are detailed in Sect. 5.2, Eq. (14), of Lindegren (2018). In particular, this updated selection takes the parallax zeropoint into account. This resulted in a set of 555,934 Gaia DR2 sources that are matched to the AllWISE AGN catalogue and 2820 sources that are matched to the ICRF3-prototype. The union of the two sets contains 556,869 Gaia DR2 sources. These sources and their positions in Gaia DR2 are a version of the Gaia-CRF that we call Gaia-CRF2.

The entire subsequent analysis in this paper (except in Sect. 5) is based on this sample or on subsets of it. For simplicity, we use the term quasar (QSO) for these objects, although other classifications (BL Lac object, Seyfert 1, etc.) may be more appropriate in many cases, and a very small number of them may be distant (>1 kpc) Galactic stars.

### 3. Properties of Gaia-CRF2

This section describes the overall astrometric and photometric properties in Gaia DR2 of the sources of the Gaia-CRF2, that is, the 556,869 quasars we obtained from the match to the AllWISE AGN catalogue and the ICRF3-prototype. Their sky density is displayed in Fig. 1. The Galactic plane area is filtered out by the AllWISE AGN selection criteria, while areas around the ecliptic poles are higher than the average density, as noted above. Lower density arcs from the WISE survey are also visible, but as a rule, the whole-sky coverage outside the Galactic plane has an average density of about 14 sources per deg\(^2\). The few sources in the Galactic plane area come from the ICRF3-prototype.

#### 3.1. Magnitude and colour

Figure 2 shows the magnitude distribution of the Gaia-CRF2 quasars. Percentage per bin of 0.1 mag (top) and cumulative distribution (bottom).

![Fig. 2. G-magnitude distribution of the Gaia-CRF2 quasars. Percentage per bin of 0.1 mag (top) and cumulative distribution (bottom).](image)

#### 3.2. Astrometric quality

In this section we describe the astrometric quality of the Gaia-CRF2 quasars based on the formal positional uncertainties and on the distribution of observed parallaxes and proper motions, which are not expected to be measurable by Gaia at the level of individual sources. We defer a direct comparison of the Gaia positions with VLBI astrometry to Sect. 4.

#### 3.2.1. Formal uncertainty in position

As a single number characterising the positional uncertainty of a source, \( \sigma_{\text{pos, max}} \), we take the semi-major axis of the dispersion...
The Galactic centre is at the origin of coordinates (centre $G$ shows the median values of the number of sources per bin of 0.1 mag).

Parallaxes and proper motions

Parallaxes and proper motions are nominally zero for the quasars that were selected for the reference frame (we neglect here the expected global pattern from the Galactic acceleration, which is expected to have an amplitude of $4.5\,\mu\text{as yr}^{-1}$, see Sect. 3.3). Their statistics are useful as complementary indicators of the global quality of the frame and support the accuracy claim. Here we consider the global statistics before investigating possible systematics in Sect. 3.3. Figure 8 shows the distribution of the parallaxes for different magnitude-limited subsets. As explained in Lindegren (2018), the Gaia DR2 parallaxes have a global zero-point error of $-0.029\,\mu\text{as}$, which is not corrected for in the data available in the Gaia archive. This feature is well visible for the quasar sample and is a real instrumental effect that is not yet eliminated by the calibration models. Fortunately, the offset is similar for the different subsets. The shape of the distributions (best visible in the full set) is clearly non-Gaussian because of the expected global pattern from the Galactic acceleration, which is

Because this is also the highest eigenvalue of the $2 \times 2$ covariance matrix, it is invariant to a change of coordinates. These are formal uncertainties (see Sect. 3.2.2 for a discussion of how real they are) for the reference epoch J2015.5 of Gaia DR2. The results for the whole sample of Gaia-CRF2 quasars and the subset with $G < 19$ are shown in Fig. 5. The median accuracy is 0.40 mas for the full set and 0.20 mas for the brighter subset. Additional statistics are given in Table 1.

The main factors governing the positional accuracy are the magnitude (Fig. 6) and location on the sky (Fig. 7). The larger-than-average uncertainty along the ecliptic in Fig. 7 is conspicuous; this is a signature of the Gaia scanning law. This feature will also be present in future releases of Gaia astrometry and will remain an important characteristic of the Gaia-CRF.
mixture of normal distributions with a large spread in standard deviation, which is primarily linked to the source magnitude. The typical half-widths of the distributions (0.4, 0.3, and 0.2 mas) are of a similar size as the median positional uncertainties in Table 1.

The distribution of the normalised debiased parallaxes, computed as \((\sigma + 0.029 \text{ mas})/\sigma_\star\), should follow a standard normal distribution (zero mean and unit variance) if the errors are Gaussian and the formal uncertainties \(\sigma_\star\) are correctly estimated. The actual distribution for the full set of 556,869 quasars is plotted in Fig. 9. The red continuous curve is a normal distribution with zero mean and standard deviation 1.08; that this very closely follows the distribution up to normalised values of ±3.5 is an amazing feature for real data. The magnitude effect is then fully absorbed by the normalisation, indicating that the Gaia accuracy in this brightness range is dominated by the photon noise. The factor 1.08 means that the formal uncertainties of the parallaxes are too small by 8%.

Similarly, the distributions in Fig. 10 for the normalised components of proper motions are very close to a normal distribution, with zero mean and standard deviations of 1.09 (\(\mu_\alpha\)) and 1.11 (\(\mu_\delta\)). The extended distributions in log scale are very similar to the parallax and are not plotted.

3.3. Systematic effects

3.3.1. Spatial distributions

In an ideal world, the errors in position, parallax, and proper motion should be purely random and not display any systematic patterns as function of position on the celestial sphere. While the non-uniform sampling of the sky produced by the Gaia scanning law is reflected in the formal uncertainties of the quasar astrometry, as shown in Fig. 7 for the positions, this does not imply that the errors (i.e. the deviations from the true values) show patterns of a similar nature. In the absence of a reliable external reference for the positions (except for the VLBI subset), the possibility of investigating the true errors in position is limited. However, the positions are derived from the same set of observations as the other astrometric parameters, using the same solution. Since the errors in parallax and proper motion are found

Fig. 7. Spatial distribution of the formal position uncertainty in Eq. (1) for the 407,959 sources of the Gaia-CRF2 with \(G < 20\). The map shows the median value in each cell of approximately 0.84 deg², using a Hammer–Aitoff projection in Galactic coordinates with zero longitude at the centre and increasing longitude from right to left. The solid black line shows the ecliptic.

Fig. 8. Distribution of parallaxes in the Gaia archive for the Gaia-CRF2 quasars, subdivided by the maximum magnitude. The line at \(\sigma = -0.029\) mas shows the global zeropoint offset.

Fig. 9. Distribution of the normalised debiased parallaxes, \((\sigma + 0.029 \text{ mas})/\sigma_\star\), for the Gaia-CRF2 quasars in linear scale (top) and logarithmic (bottom). The red curve is a normal distribution with zero mean and standard deviation 1.08.

Fig. 10. Distributions of the normalised components of proper motions of the QSOs found with Gaia data, with \(\mu_\alpha\) (top) and \(\mu_\delta\) (bottom) A normal distribution with zero mean and standard deviation 1.09 for \(\mu_\alpha\) (1.11 for \(\mu_\delta\)) is drawn in red.
to be in good agreement with the formal uncertainties calculated from the solution, we expect this to be the case for the positional errors as well.

Figure 11 shows maps in Galactic coordinates of the median parallax and proper motion components of the Gaia-CRF2 sources, calculated over cells of 4.669 deg². For cells of this size, the median number of sources per cell is 70, with the exception of low Galactic latitude, where the density is lower (see Fig. 1), resulting in a larger scatter of the median from cell to cell than in other parts of the map. In Fig. 11 this is visible as an increased number of cells with red and blue colours, instead of green and yellow, in the less populated areas.

The median parallaxes shown in the top panel of Fig. 11 were corrected for the global zeropoint of $-29 \mu\text{as}$ (Sect. 3.2.2). In all three maps, various large-scale patterns are seen for Galactic latitudes $|b| > 10$–15 deg, while at small angles (cell size), only a mixture of positive or negative offsets is visible that results from normal statistical scatter. The visual interpretation is complicated by large-scale patterns in the amplitude of the statistical scatter, in particular the smaller scatter in the second and fourth quadrants, that is, around the ecliptic poles. This is the result of a combination of the sky distribution of the sources (Fig. 1), their magnitudes (Fig. 4), and the Gaia scanning law (Fig. 7), which all exhibit similar patterns. Quantifying the large-scale systematics therefore requires a more detailed numerical analysis.

3.3.2. Spectral analysis

The vector field of the proper motions of the Gaia-CRF2 quasars was analysed using expansions on a set of vector spherical harmonics (VSH), as explained in Mignard & Klioner (2012) or Vityazev & Tsvetkov (2014).

In this approach the components of proper motion are projected onto a set of orthogonal functions up to a certain degree $l_{\text{max}}$. The terms of lower degrees provide global signatures such as the rotation and other important physical effects (secular acceleration, gravitational wave signatures), while harmonics of higher degree hold information on local distortions at different scales. Given the patterns seen in Fig. 11, we expect to see a slow decrease in the power of harmonics with $l > 1$. The harmonics of degree $l = 1$ play a special role, since any global rotation of the system of proper motions will be observed in the form of a rotation vector directly extracted from the three components with $(l,m) = (1,0)$, $(1,1)$, and $(1,-1)$, where $m$ is the order of the harmonic $|m| \leq l$.

Mignard & Klioner (2012) derived a second global term from $l = 1$ that they called glide. This physically corresponds to a dipolar displacement originating at one point on a sphere and ending at the diametrically opposite point. For the quasar proper motions, this vector field is precisely the expected signature of the galactocentric acceleration (Fanselow 1983; Bastian 1995; Sovers et al. 1998; Kovalevsky 2003; Titov & Lambert 2013). As summarised in Table 2, several VSH fits were made using different selections of quasars or other configuration parameters. Fit 1 uses all the quasars and fits only the rotation, without glide or harmonics with $l > 1$. This is very close to the conditions used to achieve the non-rotating frame in the astrometric solution for Gaia DR2. It is therefore not surprising that the rotation we find is much smaller than in the other experiments. The remaining rotation can be explained by differences in the set of sources used, treatment of outliers, and so on. This also illustrates the difficulty of producing a non-rotating frame that is non-rotating for every reasonable subset that a user may wish to select: This is not possible, at least at the level of formal uncertainties. Experiment 2 fits both the rotation and glide to all the data. The very small change in rotation compared with fit 1 shows the stability of the rotation resulting from the regular spatial distribution of the sources and the consequent near-orthogonality of the rotation and glide on this set. Fit 3 includes all harmonics of degree $l \leq 5$, that is, 70 fitted parameters. Again the results do not change very much because of the good spatial distribution. The next five fits show the influence of the selection in magnitude and modulus of proper motion, and of not weighting the data by the inverse formal variance. In the next two fits (9a and 9b), the data are divided into two independent subsets, illustrating the statistical uncertainties. Most of these fits use fewer sources with a less regular distribution on the sky.

The last fit, fit 10, uses only the faint sources and has a similar glide but a very different rotation ($x$ and $y$ components, primarily), although it comprises the majority (73%) of the Gaia-CRF2 sources. This agrees with Figs. 3 and 4 in Lindegren (2018), which show a slight dependency on colour and magnitude of the Gaia spin relative to quasars. Again, this illustrates the...
sensitivity of the determination of the residual spin to the source selection, and at this stage, we cannot offer a better explanation than that a single solid rotation is too simple a model to fit the entire range of magnitudes. No attempt was made to introduce a magnitude equation in the fits.

The formal uncertainty of all the fits using at least a few hundred thousand quasars is of the order of \(1 \mu \text{mas yr}^{-1}\). It is tempting to conclude from this that the residual rotation of the frame with respect to the distant universe is of a similar magnitude. However, the scatter from one fit to the next is considerably larger, with some values exceeding \(10 \mu \text{mas yr}^{-1}\). Clearly, an overall solid rotation does not easily fit all the Gaia data, but gives results that vary with source selection well above the statistical noise. However, the degree of consistency between the various selections allows us to state that the residual rotation rate of the Gaia-CRF is probably not much higher than \(\pm 10 \mu \text{mas yr}^{-1}\) in each axis for any subset of sources.

The typical glide vector is about \((-8, +5, +12) \pm 1 \mu \text{mas yr}^{-1}\) for the components in the ICRS. The expected signature for the galactocentric acceleration is a vector directed towards the Galactic centre with a magnitude of \(\approx 4.50 \mu \text{mas yr}^{-1}\), or \((-0.25, -3.93, -2.18) \mu \text{mas yr}^{-1}\) in the ICRS components. Clearly, the large-scale systematic effects in the Gaia proper motions, being of the order of \(10 \mu \text{mas yr}^{-1}\) at this stage of the data analysis, prevent a fruitful analysis of the quasar proper motion field in terms of the Galactic acceleration. For this purpose, an order-of-magnitude improvement is needed in the level of systematic errors, which may be achieved in future releases of Gaia data based on better instrument calibrations and a longer observation time-span. A similar improvement is needed to achieve the expected estimate of the energy flux of the primordial gravitational waves (Gwinn et al. 1997; Mignard & Klioner 2012; Klioner 2018).

The overall stability of the fits in Table 2 is partly due to the fairly uniform distribution of the Gaia-CRF2 sources over the celestial sphere, and it does not preclude the existence of significant large-scale distortions of the system of proper motions. Such systematics may be quantified by means of the fitted VSH, however, and a convenient synthetic indicator of how much signal is found at different angular scales is given by the total power in each degree \(l\) of the VSH expansion. This power \(P_l\) is invariant under orthogonal transformation (change of coordinate system) and therefore describes a more intrinsic, geometric feature than the individual components of the VSH expansion. The degree \(l\) corresponds to an angular scale of \(\sim 180^\circ/l\).

In Fig. 12 (top panel) we plot \((P_l/4\pi)^{1/2}\) in \(\mu \text{mas yr}^{-1}\), representing the RMS value of the vector field for the corresponding degree \(l\). The lower panel in Fig. 12 shows the significance level of the power given as the equivalent standard normal variate derived from the asymptotic \(\chi^2\) distribution; see Mignard & Klioner (2012) for details. The points labelled \(S\) and \(T\) correspond to the spheroidal and toroidal harmonics, with \(T&S\) for their quadratic combination. To illustrate the interpretation of the diagrams, for \(T_1\) the RMS value is \((P_1/4\pi)^{1/2} \approx 10 \mu \text{mas yr}^{-1}\), which should be similar to the magnitude of the rotation vector for fit 3 in Table 2. The significance of this value is \(Z_{2T} \approx 7\), corresponding to \(7\sigma\) of a normal distribution, or a probability below \(10^{-11}\).

For the low degrees plotted in Fig. 12, the power generally decreases with increasing \(l\) (smaller angular scales). This indicates that the systematics are generally dominated by the large angular scales. The total RMS for \(l \leq 10\) (angular scales \(\geq 18\) deg) is \(42 \mu \text{mas yr}^{-1}\). Lindgren (2018) analysed the large-scale systematics of the Gaia DR2 proper motions of exactly the same quasar sample, using a very different spatial correlation technique. A characteristic angular scale of 20 deg was found, with an RMS amplitude of \(28 \mu \text{mas yr}^{-1}\) per component of proper motion (their Eq. (18)). Since this corresponds to \(40 \mu \text{mas yr}^{-1}\) for the total proper motion, their result is in good agreement with ours. They also found higher-amplitude oscillations with a spatial period of \(\approx 1\) deg, which in the present context of Gaia-CRF2 are almost indistinguishable from random noise, however.

### Table 2. Large-scale structure of the proper motion field of the Gaia-CRF2 quasars analysed using vector spherical harmonics.

<table>
<thead>
<tr>
<th>Fit</th>
<th>Source selection</th>
<th>(W)</th>
<th>(l_{\text{max}})</th>
<th>(N)</th>
<th>Rotation (\mu)</th>
<th>Glide (\mu)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(\text{mas yr}^{-1})</td>
<td>(\text{mas yr}^{-1})</td>
</tr>
<tr>
<td>1</td>
<td>all</td>
<td>y</td>
<td>1</td>
<td>556869</td>
<td>(-3.1 \pm 0.8)</td>
<td>(-1.9 \pm 0.7)</td>
</tr>
<tr>
<td>2</td>
<td>all</td>
<td>y</td>
<td>1</td>
<td>556869</td>
<td>(-3.6 \pm 0.8)</td>
<td>(-2.2 \pm 0.7)</td>
</tr>
<tr>
<td>3</td>
<td>all</td>
<td>y</td>
<td>5</td>
<td>556869</td>
<td>(-5.5 \pm 1.1)</td>
<td>(-7.9 \pm 0.9)</td>
</tr>
<tr>
<td>4</td>
<td>(\mu &lt; 2) (\text{mas yr}^{-1}), (G &lt; 18)</td>
<td>y</td>
<td>5</td>
<td>27189</td>
<td>(-13.8 \pm 2.0)</td>
<td>(-13.2 \pm 1.7)</td>
</tr>
<tr>
<td>5</td>
<td>(\mu &lt; 2) (\text{mas yr}^{-1}), (G &lt; 18) n</td>
<td>y</td>
<td>5</td>
<td>27189</td>
<td>(-8.9 \pm 2.9)</td>
<td>(-12.1 \pm 2.4)</td>
</tr>
<tr>
<td>6</td>
<td>(\mu &lt; 2) (\text{mas yr}^{-1}), (G &lt; 19)</td>
<td>y</td>
<td>5</td>
<td>149146</td>
<td>(-11.2 \pm 1.3)</td>
<td>(-12.0 \pm 1.1)</td>
</tr>
<tr>
<td>7</td>
<td>(\mu &lt; 3) (\text{mas yr}^{-1}), (G &lt; 20)</td>
<td>y</td>
<td>5</td>
<td>400472</td>
<td>(-5.9 \pm 1.1)</td>
<td>(-8.6 \pm 0.9)</td>
</tr>
<tr>
<td>8</td>
<td>(\mu &lt; 3) (\text{mas yr}^{-1})</td>
<td>y</td>
<td>5</td>
<td>513270</td>
<td>(-5.7 \pm 1.1)</td>
<td>(-7.9 \pm 0.9)</td>
</tr>
<tr>
<td>9a</td>
<td>([10^5a] \mod 2 = 0)</td>
<td>y</td>
<td>5</td>
<td>278170</td>
<td>(-5.8 \pm 1.6)</td>
<td>(-8.9 \pm 1.3)</td>
</tr>
<tr>
<td>9b</td>
<td>([10^5a] \mod 2 = 1)</td>
<td>y</td>
<td>5</td>
<td>278699</td>
<td>(-5.1 \pm 1.6)</td>
<td>(-5.8 \pm 1.3)</td>
</tr>
<tr>
<td>10</td>
<td>(G &gt; 19)</td>
<td>y</td>
<td>5</td>
<td>406356</td>
<td>9.8 \pm 2.1</td>
<td>6.2 \pm 1.8</td>
</tr>
</tbody>
</table>

Notes. \(\mu = (\mu_x^2 + \mu_y^2 + \mu_z^2)^{1/2}\) is the modulus of the proper motion. \(N\) is the number of sources used in the solution. \(W = “y” \text{ or “n”}\) for weighted or unweighted solution. The weighted solutions use a block-diagonal weight matrix obtained from the \(2 \times 2\) covariance matrix of each source. \(l_{\text{max}}\) is the highest degree of the fitted VSH from which rotation and glide are extracted for \(l = 1\). The columns headed \(x, y, z\) give the components of the rotation and glide along the principal axes of the ICRS. In rows 9a and 9b, two independent halves of the sample are selected according to whether \([10^5a] \text{ is even (9a) or odd (9b)}, \text{with } a \text{ in degrees.}\)
4. ICRF3-prototype subset of Gaia-CRF2

This section describes the subset of 2820 Gaia-CRF2 quasars matched to the ICRF3-prototype (Sect. 2.2), that is, the optical counterparts of compact radio sources with accurate VLBI positions. A comparison between the optical and VLBI positions is in fact a two-way exercise, as useful for understanding positions. A comparison between the optical and VLBI positions can be assessed and individual cases of truly discrepant positions between the radio and optical domains can be identified. The VLBI positions are less homogeneous in accuracy than the corresponding Gaia DR1 data, but the ≳1650 ICRF3-prototype sources with a (formal) position uncertainty <0.2 mas match the Gaia positions of the brighter (G < 18 mag) sources well in quality.

4.1. Properties of the Gaia sources in the ICRF3-prototype

Figure 13 shows the spatial distribution of the 2820 optical counterparts of ICRF3-prototype sources on the sky. The plot is in Galactic coordinates to facilitate comparison with Fig. 1, showing the full Gaia-CRF2 sample. The area in the lower right quadrant with low density corresponds to the region of the sky at δ < −40 deg with less VLBI coverage. Otherwise the distribution is relatively uniform, but with a slight depletion along the Galactic plane, as expected for an instrument operating at optical wavelengths.

The magnitude distribution of the ICRF3-prototype sources is shown in Fig. 14. The median is 18.8 mag, compared with 19.5 mag for the full Gaia-CRF2 sample shown in Fig. 2. The colour distribution (not shown) is similar to that of the full sample, shown in Fig. 2, only slightly redder: the median GRP − GBP is ≈0.8 mag for the ICRF3-prototype subset, compared with 0.7 mag for the full sample.

In terms of astrometric quality, the Gaia DR2 sources in the ICRF3-prototype subset do not differ significantly from other quasars in Gaia-CRF2 at the same magnitude. Figure 15 displays the formal uncertainty in position, computed with Eq. (1), as function of the G magnitude. Both the median relation and the scatter about the median are virtually the same as for the general population of quasars in Gaia-CRF2 shown in Fig. 6. For G ≥ 16.2, only few points in Fig. 15 lie clearly above the main relation. This may be linked to the change in the onboard CCD observation window allocation that occurs at G ≈ 16 (Gaia Collaboration et al. 2016). Four hundred and nine sources are brighter than G = 17.4, where the median position uncertainty as shown on Fig. 15 reaches 100 μas.
4.2. Angular separations

We now compare the positions in Gaia DR2 and ICRF3-prototype directly for the 2820 quasars in common. Figure 16 gives in log-scale the distribution of the angular distances computed as

\[ \rho = (\Delta \alpha^2 + \Delta \delta^2)^{1/2}, \]

where \( \Delta \alpha = (\alpha_{\text{Gaia}} - \alpha_{\text{VLRB}}) \cos \delta \). While for most of the sources, \( \rho \) is lower than 1 mas and very often much below this level, the number of discrepant sources is significant, and a few even have a position difference higher than 10 mas that would require individual examination.

To illustrate the dependence on the solution accuracies, Fig. 17 shows scatter plots of \( \rho \) versus the formal uncertainty in the ICRF3-prototype (top) and Gaia-CRF2 (bottom). Several of the most extreme distances in the top diagram are for sources with a large uncertainty in the ICRF3-prototype. However, some sources with nominally good solutions in both datasets exhibit large positional differences. These deserve more attention as the differences could represent real offsets between the centres of emission at optical and radio wavelengths. This is not further investigated in this paper, which is devoted to present the main properties of the Gaia-CRF2. Other explanations for the large differences can be put forward, such as a mismatch on the Gaia side when the optical counterpart is too faint and a distant star happens to be matched instead (unlikely at <10 mas distance); an extended galaxy around the quasar that is misinterpreted by

the Gaia detector (should in general produce a poor solution); double or lensed quasars; or simply statistical outliers from the possibly extended tails of random errors. Although the ICRF3-prototype data in Fig. 17 cover a wider range in \( \sigma_{\text{pos,max}} \) than the Gaia data, the cores of both distributions extend from \( \approx 0.1 \) to 0.5 mas.

4.3. Normalised separations

The angular separations \( \rho \) become statistically more meaningful when scaled with the combined standard uncertainties. In the case of correlated variables, Mignard et al. (2016) have shown how to compute a dimensionless statistic \( X \), called the normalised separation (their Eq. (4)). If the positional errors in both catalogues are Gaussian with the given covariances, then \( X \) is expected to follow a standard Rayleigh distribution, and values
4.4. Large-scale systematics

In this section we analyse the positional difference between Gaia DR2 and the ICRF3-prototype in terms of large-scale spatial patterns. As in Sect. 3.3, the vector field of position differences is decomposed using VSH, where in particular the coefficients for degree $l = 1$ give the orientation difference of the two frames and a glide in position. Several fits were made to assess the stability of the orientation rotation against various selections of sources. Nominally, Gaia DR2 has been aligned to the ICRF3-prototype and no significant orientation difference should remain. However, stating that the two frames have been aligned is not the complete story, since the final alignment depends on many details of the fit: weighting scheme, outlier filtering, magnitude selection, and the model used for the fit. Furthermore, as explained in Sect. 2.2, the alignment was made using a slightly different set of ICRF3-prototype sources than currently considered. As a consequence of these differences, we often find statistically significant non-zero orientation errors in our fits. The amplitude of these errors provides the best answer to the question of how precisely the two frames share the same axes.

The results of the various fits are summarised in Table 3. The first fit is similar to the alignment procedure in the astrometric solution for Gaia DR2 in that only the three orientation parameters (otherwise denoted $e_x$, $e_y$, $e_z$) are fitted without a glide component. Of all the fits in the table, this has the overall smallest, statistically most insignificant orientation parameters. It gives a formal uncertainty in the alignment of about $30 \mu$as per axis. Fit 2, using the same data set, but fitting the glide as well, reveals a different picture. The orientation parameters remain negligible, but not as close to zero as in fit 1, and the glide components have a significant amplitude. The uncertainty is unchanged at about $30 \mu$as. This is a good illustration of the ambiguity in the alignment when the procedure is not fully implemented.

In fits 3 to 5, only the orientation parameters are estimated, but with different filtering of the data, with or without statistical weighting of the differences. We showed in Sect. 4.1 that a subset of sources has good astrometric quality in both catalogues, but the position differences are not compatible with the formal uncertainties. Removing these sources from the fit greatly improves the formal precision of the fit, while the orientation parameters are changed by a few tens of $\mu$as, which is still only marginally significant. More significant changes result from including the glide and higher degrees of VSH (fits 6 to 8), or restricting the sample to the brighter subset (fits 9 and 10) with or without weighting. In these fits particularly the orientation error in $x$ and the glide in $y$ become significant. Finally, cases 11a and b are run on two independent halves of the data to ascertain the sensitivity of the solution to the selection.

Based on these experiments, we state that the axes of the Gaia-CRF2 and the ICRF3-prototype are aligned with an uncertainty of 20 to 30 $\mu$as, but no precise value can be provided without agreeing on the detailed model and numerical procedures for determining the orientation errors.

5. Other quasars in Gaia DR2

The cross-match of Gaia DR2 with the AllWISE AGN catalogue provided a very clean and homogeneous sample of quasars that is suitable for the definition of the Gaia-CRF2 and systematic investigation of its properties. However, other catalogues exist that will enlarge the sample of known or probable quasars in Gaia DR2 for other purposes. The Million
Table 3. Global differences between the Gaia-CRF2 positions of ICRF sources and their positions in the ICRF3-prototype, expressed by the orientation and glide parameters.

<table>
<thead>
<tr>
<th>Fit</th>
<th>Source selection</th>
<th>W</th>
<th>$l_{\text{max}}$</th>
<th>$N$</th>
<th>Orientation ($\mu$as)</th>
<th>Glide ($\mu$as)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$x$</td>
<td>$y$</td>
</tr>
<tr>
<td>1</td>
<td>all</td>
<td>y</td>
<td>1</td>
<td>2820</td>
<td>$-9 \pm 29$</td>
<td>$4 \pm 27$</td>
</tr>
<tr>
<td>2</td>
<td>all</td>
<td>y</td>
<td>1</td>
<td>2820</td>
<td>$-28 \pm 31$</td>
<td>$-8 \pm 29$</td>
</tr>
<tr>
<td>3</td>
<td>$\rho &lt; 10$ mas</td>
<td>y</td>
<td>1</td>
<td>2773</td>
<td>$-17 \pm 16$</td>
<td>$22 \pm 15$</td>
</tr>
<tr>
<td>4</td>
<td>$\rho &lt; 2$ mas</td>
<td>y</td>
<td>1</td>
<td>2423</td>
<td>$-35 \pm 9$</td>
<td>$21 \pm 8$</td>
</tr>
<tr>
<td>5</td>
<td>$\rho &lt; 2$ mas</td>
<td>n</td>
<td>1</td>
<td>2423</td>
<td>$-13 \pm 14$</td>
<td>$5 \pm 14$</td>
</tr>
<tr>
<td>6</td>
<td>$\rho &lt; 2$ mas</td>
<td>y</td>
<td>5</td>
<td>2423</td>
<td>$-47 \pm 12$</td>
<td>$30 \pm 10$</td>
</tr>
<tr>
<td>7</td>
<td>$\rho &lt; 1$ mas</td>
<td>y</td>
<td>5</td>
<td>1932</td>
<td>$-47 \pm 10$</td>
<td>$12 \pm 9$</td>
</tr>
<tr>
<td>8</td>
<td>$\rho &lt; 1$ mas</td>
<td>n</td>
<td>5</td>
<td>1932</td>
<td>$-15 \pm 12$</td>
<td>$2 \pm 12$</td>
</tr>
<tr>
<td>9</td>
<td>$\rho &lt; 2$ mas, $G &lt; 19$</td>
<td>y</td>
<td>5</td>
<td>1382</td>
<td>$-57 \pm 16$</td>
<td>$33 \pm 13$</td>
</tr>
<tr>
<td>10</td>
<td>$\rho &lt; 2$ mas, $G &lt; 19$</td>
<td>n</td>
<td>5</td>
<td>1382</td>
<td>$-65 \pm 20$</td>
<td>$0 \pm 18$</td>
</tr>
<tr>
<td>11a</td>
<td>$\rho &lt; 2$ mas, $[10^5 \mu \alpha]$ mod 2 = 0</td>
<td>y</td>
<td>5</td>
<td>1255</td>
<td>$-19 \pm 18$</td>
<td>$34 \pm 15$</td>
</tr>
<tr>
<td>11b</td>
<td>$\rho &lt; 2$ mas, $[10^5 \mu \alpha]$ mod 2 = 1</td>
<td>y</td>
<td>5</td>
<td>1168</td>
<td>$-61 \pm 17$</td>
<td>$33 \pm 15$</td>
</tr>
</tbody>
</table>

Notes. $\rho$ is the angular separation between the optical and radio positions. $N$ is the number of sources used in the fit. $W = "y"$ or "n" for weighted or unweighted solution. The weighted solutions use a non-diagonal weight matrix resulting from the combination of Gaia covariances and the covariances from the ICRF3-prototype. $l_{\text{max}}$ is the highest degree of the fit from which orientation and glide are extracted for $l = 1$. The columns headed $x$, $y$, $z$ give the components of the orientation and glide along the principal axes of the ICRS.

6. Conclusions

With Gaia DR2, a long-awaited promise of Gaia has come to fruition: the publication of the first full-fledged optical realisation of the ICRS, that is to say, an optical reference frame built only on extragalactic sources. Comprising more than half a million extragalactic sources that are globally positioned on the sky with a median uncertainty of 0.4 mas on average, this represents a major step in the history of non-rotating celestial reference frames built over the centuries by generations of astronomers. The brighter subset with $G < 18$ mag, comprising nearly 30 000 quasars with $\pm 0.12$ mas astrometric accuracy, is the best reference frame available today and within relatively easy reach for telescopes of moderate size.

We have summarised the detailed content and mapped the main properties of Gaia-CRF2 as functions of magnitude and position. The quality claims regarding positional accuracy are supported by independent indicators such as the distribution of parallaxes or proper motions. Large-scale systematics are characterised by means of expansions in vector spherical harmonics. Comparison with VLBI positions in a prototype version of the forthcoming ICRF3 shows a globally satisfactory agreement at the level of 20 to 30 $\mu$as. Several sources with significant radio-optical differences of several mas require further investigation on a case-by-case basis.

Acknowledgements. This work has made use of data from the ESA space mission Gaia, processed by the Gaia Data Processing and Analysis Consortium (DPAC). We are grateful to the developers of TOPCAT (Taylor 2005) for their software. Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement. The Gaia mission website is http://www.cosmos.esa.int/gaia. The authors are members of the Gaia DPAC. This work has been supported by the European Space Agency in the framework of the Gaia project; the Centre National d‘Etudes Spatiales (CNES), the French Centre National de la Recherche Scientifique (CNRS) and the Programme National GRAM of CNRS/INSU with INP and IN2P3 co-funded by CNES; the German Aerospace Agency DLR under grants 50QG0501, 50QG1401 50QG0601, 50QG0901, and 50QG1402; and the Swedish National Space Board. We gratefully acknowledge the IAU Working Group on ICRF3 for their cooperation during the preparation of this work and

Fig. 20. Magnitude distribution for $\sim 507,000$ Gaia sources that are not included in Gaia-CRF2, but are tentatively identified as quasars through a cross-match with the MILLIQUAS catalogue.

Quasars Catalogue (MILLIQUAS; Flesch 2015) is a compilation of quasars and AGNs from the literature, including the release of SDSS-DR14 and AllWISE. We have cross-matched MILLIQUAS5 to Gaia DR2 using a matching radius of 5 arcsec, but otherwise applying the same selection criteria as for Gaia-CRF2. This yielded 1 007 920 sources with good five-parameter solutions in Gaia DR2, of which 501 204 are in common with the AllWISE selection in Gaia-CRF2. The magnitude distribution of the 506 716 additional sources is shown in Fig. 20. With a median $G = 20.2$ mag, these sources are typically one magnitude fainter than the AllWISE AGNs in Gaia-CRF2, with positional uncertainties of about 1 mas.

Obviously, the Gaia DR2 release contains even more quasars. They can be found by cross-matching with other QSO catalogues such as the LQAC (Souchay et al. 2015) and various VLBI catalogues. Ultimately, a self-consistent identification of quasars from photometric and astrometric data of Gaia will be possible in a future release.

for their willingness to let us use an unpublished working version of the ICRF3
(solution from the GSFC).

References

référence spatia-temporels”, ed. N. Capitanne, 84
Bastian, U. 1995, in Future Possibilities for astrometry in Space, eds. M. A. C.
Perryman, & F. Van Leeuwen, ESA SP, 379, 99
Fenselow, J. L. 1983, Observation Model and parameter partial for the JPL VLBI
Gwinn, C. R., Eubanks, T. M., Pyne, T., Birkinshaw, M., & Matsakis, D. N. 1997,
Petrov, L., & Kovalev, Y. Y. 2017, MNras, 467, L71
Taylor, M. B. 2005, in Astronomical Data Analysis Software and Systems XIV,
eds. P. Shopbell, M. Britton, & R. Ebert, ASP Conf. Ser., 347, 29

1 Université Côte d’Azur, Observatoire de la Côte d’Azur, CNRS, Laboratoire Lagrange, Bd de l’Observatoire, CS 34229, 06304 Nice Cedex 4, France
2 Lohrmann Observatory, Technische Universität Dresden, Momm- senstraße 13, 01062 Dresden, Germany
3 Lund Observatory, Department of Astronomy and Theoretical Physics, Lund University, Box 43, 22100 Lund, Sweden
4 European Space Astronomy Centre (ESAC), Camino bajo del Castillo, s/n, Urbanizacion Villarfranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
5 Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Mönchhofstr. 12-14, 69120 Heidelberg, Germany
6 HE Space Operations BV for ESA/ESAC, Camino bajo del Castillo, s/n, Urbanizacion Villarfranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
7 Vitrociset Belgium for ESA/ESAC, Camino bajo del Castillo, s/n, Urbanizacion Villarfranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
8 ATG Europe for ESA/ESAC, Camino bajo del Castillo, s/n, Urbanizacion Villarfranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
9 Telespazio Vega UK Ltd for ESA/ESAC, Camino bajo del Castillo, s/n, Urbanizacion Villarfranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
10 GEPI, Observatoire de Paris, Université PSL, CNRS, 5 Place Jules Janssen, 92190 Meudon, France
11 Univ. Grenoble Alpes, CNRS, IPAG, 38000 Grenoble, France
12 SYRTE, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, LNE, 61 avenue de l’Observatoire 75014 Paris, France
13 ON/MCTI-BR, Rua Gal. José Cristino 77, Rio de Janeiro, CEP 20921-400, RJ, Brazil
14 OV/URFJ-BR, Ladeira Pedro Antônio 43, Rio de Janeiro, CEP 20080-090, RJ, Brazil
15 Laboratoire d’astrophysique de Bordeaux, Univ. Bordeaux, CNRS, B18N, allée Geoffroy Saint-Hilaire, 33615 Pessac, France
16 Leiden Observatory, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands
17 INAF - Osservatorio Astronomico di Padova, Vicolo Osservatorio 5, 35122 Padova, Italy
18 Science Support Office, Directorate of Science, European Space Research and Technology Centre (ESA/ESTEC), Keplerlaan 1, 2201 AZ, Noordwijk, The Netherlands
19 Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany
20 Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
21 Department of Astronomy, University of Geneva, Chemin des Maillettes 51, 1290 Versoix, Switzerland
22 Mission Operations Division, Operations Department, Directorate of Science, European Space Research and Technology Centre (ESA/ESTEC), Keplerlaan 1, 2201 AZ, Noordwijk, The Netherlands
23 Institut de Ciències del Cosmos, Universitat de Barcelona (IEEC-UB), Martí i Franquès 1, 08028 Barcelona, Spain
24 CNES Centre Spatial de Toulouse, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France
25 Institut d’Astronomie et d’Astrophysique, Université Libre de Brux- elles CP 226, Boulevard du Triomphe, 1050 Brussels, Belgium
26 F.R.S.-FNRS, Rue d’Egmont 5, 1000 Brussels, Belgium
27 INAF - Osservatorio Astrofisico di Arcetri, Largo Enrico Fermi 5, 50125 Firenze, Italy
28 Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK
29 INAF - Osservatorio Astrofisico di Torino, via Osservatorio 20, 10025 Pino Torinese (TO), Italy
30 INAF - Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, via Piero Gobetti 93/3, 40129 Bologna, Italy
31 SERCO Gestión de Negocios for ESA/ESAC, Camino bajo del Castillo, s/n, Urbanizacion Villarfranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
32 ALTEC S.p.A., Corso Marche, 79, 10146 Torino, Italy
33 Department of Astronomy, University of Geneva, Chemin d’Ecogia 16, 1290 Versoix, Switzerland
34 Gaia DPAC Project Office, ESAC, Camino bajo del Castillo, s/n, Urbanizacion Villarfranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
35 National Observatory of Athens, I. Metaxa and Vas. Pavlou, Palaia Penteli, 15236 Athens, Greece
36 IMCCE, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, Univ. Lille, 77 av. Denfert-Rochereau, 75014 Paris, France
37 Royal Observatory of Belgium, Ringlaan 3, 1180 Brussels, Belgium
38 Institute for Astronomy, University of Edinburgh, Royal Observa- tory, Blackford Hill, Edinburgh EH9 3HJ, UK
39 Institut voor Sterrenkunde, KU Leuven, Celestijnenlaan 200D, 3001 Leuven, Belgium
40 Institut d’ Astrophysique et de Géophysique, Université de Liège, 19c, Allée du 6 Août, 4000 Liège, Belgium
41 Área de Lenguajes y Sistemas Informáticos, Universidad Pablo de Olavide, Ctra. de Utrera, km 1. 41013 Sevilla, Spain
42 ETSE Telecomunicación, Universidade de Vigo, Campus Lagoas- Olavide, Ctra. de Utrera, km 1. 41013 Sevilla, Spain
43 Large Synoptic Survey Telescope, 950 N. Cherry Avenue, Tucson, AZ 85719, USA
44 Observatorio Astronomique de Strasbourg, Université de Strasbourg, CNRS, UMR 7550, 11 rue de l’Université, 67000 Strasbourg, France

A14, page 13 of 15