Gaia Data Release 2: The astrometric solution

Astronomy & Astrophysics manuscript no. ms
April 26, 2018

Gaia Data Release 2 – The astrometric solution

L. Lindegren[1,⋆], J. Hernández[2], A. Bombrun[3], S. Klioner[4], U. Bastian[5], M. Ramos-Lerate[6], A. de Torres[3], H. Steidelmüller[4], C. Stephenson[7], D. Hobbs[1], U. Lammers[2], M. Biermann[5], R. Geyer[4], T. Hilger[4], D. Michalik[1], U. Stampa[5], P.J. McMillan[1], J. Castañeda[8], M. Clotet[8], G. Comoretto[7], M. Davidson[9], C. Fabricius[8], G. Gracia[10], N.C. Hambly[9], A. Hutton[11], A. Mora[11], J. Portell[8], F. van Leeuwen[12], U. Abbas[13], A. Abreu[14], M. Altmann[5,15], A. Andrei[16], E. Anglada[17], L. Balaguer-Núñez[8], C. Barache[15], U. Becciani[18], S. Bertone[13,15,19], L. Bianchi[20], S. Bouquillon[15], G. Bourda[21], T. Brüsemeister[5], B. Bucciarelli[13], D. Busonero[13], R. Buzzi[13], R. Cancelliere[22], T. Carlucci[15], P. Charlot[21], N. Cheek[17], M. Crosta[13], C. Crowley[3], J. de Bruijne[23], F. de Felice[24], R. Drimmel[13], P. Esquej[25], A. Fienga[26], E. Fraile[25], M. Gai[13], N. Garralda[8], J.J. González-Vidal[8], R. Guerra[2], M. Hauser[5,27], W. Hofmann[5], B. Holl[28], S. Jordan[5], M.G. Lattanzi[13], H. Lenhardt[5], S. Liao[13,29,30], E. Licata[20], T. Lister[31], W. Löffler[5], J. Marchant[32], J.-M. Martin-Fleitas[11], R. Messineo[33], F. Mignard[34], R. Morbidelli[13], E. Poggio[35,13], A. Riva[13], N. Rowell[9], E. Salguero[36], M. Sarasso[13], E. Sciacca[18], H. Siddiqui[7], R.L. Smart[13], A. Spagna[13], I. Steele[32], F. Taris[15], J. Torra[8], A. van Elteren[37], W. van Reeven[11], and A. Vecchiato[13]

(Affiliations can be found after the references)

ABSTRACT

Context. *Gaia* Data Release 2 (*Gaia* DR2) contains results for 1693 million sources in the magnitude range 3 to 21 based on observations collected by the European Space Agency *Gaia* satellite during the first 22 months of its operational phase.

Aims. We describe the input data, models, and processing used for the astrometric content of *Gaia* DR2, and the validation of these results performed within the astrometry task.

Methods. Some 320 billion centroid positions from the pre-processed astrometric CCD observations were used to estimate the five astrometric parameters (positions, parallaxes, and proper motions) for 1332 million sources, and approximate positions at the reference epoch J2015.5 for an additional 361 million mostly faint sources. These data were calculated in two steps. First, the satellite attitude and the astrometric calibration parameters of the CCDs were obtained in an astrometric global iterative solution for 16 million selected sources, using about 1% of the input data. This primary solution was tied to the extragalactic International Celestial Reference System (ICRS) by means of quasars. The resulting attitude and calibration were then used to calculate the astrometric parameters of all the sources. Special validation solutions were used to characterise the random and systematic errors in parallax and proper motion.

Results. For the sources with five-parameter astrometric solutions, the median uncertainty in parallax and position at the reference epoch J2015.5 is about 0.04 mas for bright ($G < 14$ mag) sources, 0.1 mas at $G = 17$ mag, and 0.7 mas at $G = 20$ mag. In the proper motion components the corresponding uncertainties are 0.05, 0.2, and 1.2 mas yr^{-1}, respectively. The optical reference frame defined by *Gaia* DR2 is aligned with ICRS and is non-rotating with respect to the quasars to within 0.15 mas yr^{-1}. From the quasars and validation solutions we estimate that systematics in the parallaxes depending on position, magnitude, and colour are generally below 0.1 mas, but the parallaxes are on the whole too small by about 0.03 mas. Significant spatial correlations of up to 0.04 mas in parallax and 0.07 mas yr^{-1} in proper motion are seen on small (< 1 deg) and intermediate (20 deg) angular scales. Important statistics and information for the users of the *Gaia* DR2 astrometry are given in the appendices.

Key words. astrometry – parallaxes – proper motions – methods: data analysis – space vehicles: instruments

1. Introduction

Gaia DR2 (Gaia Collaboration et al. 2018a), the second release of data from the European Space Agency mission *Gaia* (Gaia Collaboration et al. 2016b), contains provisional results based on observations collected during the first 22 months since the start of nominal operations in July 2014. The astrometric data in *Gaia* DR2 include the five astrometric parameters (position, parallax, and proper motion) for 1332 million sources, and the approximate positions at epoch J2015.5 for an additional 361 million mostly faint sources with too few observations for a reliable five-parameter solution. The limiting magnitude is $G \simeq 21.0$. The bright limit is $G \simeq 3$, although stars with $G \lesssim 6$ generally have inferior astrometry due to calibration issues. The data are publicly available in the online *Gaia* Archive at https://archives.esac.esa.int/gaia.

This paper gives an overview of the astrometric processing for *Gaia* DR2 and describes the main characteristics of the results. Further details are provided in the online documentation of the *Gaia* Archive and in specialised papers. In contrast to the Tycho-*Gaia* astrometric solution (TGAS; Lindegren et al. 2016) in *Gaia* DR1 (Gaia Collaboration et al. 2016a), the present solution does not incorporate any astrometric information from Hipparcos and Tycho-2, and the results are therefore independent of these catalogues. Similarly to *Gaia* DR1, all sources are treated as single stars and thus representable by the five astrometric parameters. For unresolved binaries (separation $\lesssim 100$ mas), the results thus refer to the photocentre, while for resolved binaries the results may refer to either component and are sometimes spurious due to confusion of the components. For a very small number

⋆ Corresponding author: L. Lindegren
e-mail: lennart@astro.lu.se

of nearby sources, perspective effects due to their radial motions were taken into account.

The input data for the astrometric solutions are summarised in Sect. 2. A central part of the processing carried out by the *Gaia* Data Processing and Analysis Consortium (DPAC; Gaia Collaboration et al. 2016b) is the astrometric global iterative solution (AGIS) described in Lindegren et al. (2012, hereafter the AGIS paper), and the present results were largely computed using the models and algorithms described in that paper. However, a few major additions have been made since 2012, and they are outlined in Sect. 3. Section 4 describes the main steps of the solutions. The validation of the results carried out by the astrometry team of DPAC primarily aimed at estimating the level of systematic errors; this is described in Sect. 5, with the main conclusions in Sect. 6. Three appendices give statistics and other information of potential interest to users of the *Gaia* DR2 astrometry.

2. Data used

The main input to the astrometric solutions are one- or two-dimensional measurements of the locations of point-source images on *Gaia*'s CCD detectors, derived by the image parameter determination (Sect. 2.2) in the pre-processing of the raw *Gaia* data (Fabricius et al. 2016). The CCD measurements must be assigned to specific sources, so that all the measurements of a given source can be considered together in the astrometric solution. This is achieved by a dedicated cross-matching procedure following the same overall three-step scheme as for *Gaia* DR1. First all sources close to a detection – the candidate matches – are found. This is done for the full set of observations, using updated calibrations and an extended attitude covering also time intervals that may later be excluded. Next, the detections are divided into isolated groups consisting of the smallest possible sets of detections with candidate matches to the same sources, such that a given candidate source only appears in one group. Finally, each group is resolved into clusters of detections and each cluster assigned to one source. What is done differently from *Gaia* DR1 is the way the clusters are formed. For *Gaia* DR1, this involved a simple nearest-neighbour algorithm, applied to one detection at a time, without a global view of the group. For *Gaia* DR2, a more elaborate clustering algorithm was used, giving better results in dense areas and performing much better for sources with high proper motions as it includes the detection of linear motion. The overall cross-match scheme is described in Castañeda et al. (in prep.). For *Gaia* DR2, about 52 billion detections were processed, but 11 billion were considered spurious and therefore did not take part in the cross matching. The remaining 41 billion transits were matched to 2583 million sources, of which a significant number could still be spurious. Even among the clearly non-spurious sources, many had too few or too poor observations to make it to the release, which therefore has a total of 1693 million sources.

A second important input to the astrometric solution for *Gaia* DR2 is the colour information, available for most of the sources thanks to the early photometric processing of data from the blue and red photometers (BP and RP; van Leeuwen et al. 2017; Riello et al. 2018; Evans et al. 2018). This processing used astrometric data (source and attitude parameters) taken from a provisional astrometric solution (Sect. 4.1).

Additional input data are obtained from the basic angle monitor (BAM; Sect. 2.4) and the orbit reconstruction and time synchronisation data provided by the Mission Operations Centre (Sect. 5.3 in Gaia Collaboration et al. 2016b).

2.1. Time coverage

Gaia DR2 is based on data collected from the start of the nominal observations on 2014 July 25 (10:30 UTC) until 2016 May 23 (11:35 UTC), or 668 days. However, the astrometric solution for this release did not use the observations during the first month after commissioning, when a special scanning mode (the ecliptic pole scanning law, EPSL) was employed. The data for the astrometry therefore start on 2014 Aug 22 (21:00 UTC) and cover 640 days or 1.75 yr, with some interruptions mentioned below.

Hereafter we use the onboard mission timeline (OBMT) to label onboard events; it is expressed as the number of nominal revolutions of exactly 21 600 s (6 h) onboard time from an arbitrary origin. The approximate relation between OBMT (in revolutions) and barycentric coordinate time (TCB, in Julian years) at *Gaia* is

$$\mathrm{TCB} \simeq \mathrm{J2015.0} + (\mathrm{OBMT} - 1717.6256\ \mathrm{rev})/(1461\ \mathrm{rev\ yr^{-1}})\,. \quad (1)$$

The nominal observations start at OBMT 1078.38 rev. The astrometric solution used data in the interval OBMT 1192.13–3750.56 rev, with major gaps at OBMT 1316.49–1389.11 rev and 2324.90–2401.56 rev due to mirror decontamination events and the subsequent recovery of thermal equilibrium. Planned maintenance operations (station-keeping manoeuvres, telescope refocusing, etc.), micrometeoroid hits, and other events caused additional gaps that rarely exceeded a few hours.

The reference epoch used for the astrometry in *Gaia* DR2 is J2015.5 (see Sect. 3.1), approximately half-way through the observation period used in the solution. This reference epoch, chosen to minimise correlations between the positions and proper motions, is 0.5 Julian year later than the reference epoch for *Gaia* DR1; this difference must be taken into account when comparing positional data from the two releases.

2.2. Image parameters

Image parameters are obtained by fitting a model profile to the photon counts in the observation window centred on the source in the CCD pixel stream. The model profile is a point spread function (PSF) for a two-dimensional window and a line spread function (LSF) in the more common case of a one-dimensional window (for details on the CCD operations, see Sect. 3.3.2 in Gaia Collaboration et al. 2016b). The main image parameters are the estimated one- or two-dimensional location of the image centroid (defined by the origin of the fitted PSF or LSF) and the integrated flux of the image. The image parameter determination for *Gaia* DR2 is essentially the same as for *Gaia* DR1 (see Sect. 5 in Fabricius et al. 2016). In particular, the fitted PSF and LSF were assumed to be independent of time and of the colour and magnitude of the source, which means that centroid shifts depending on time, colour, and magnitude need to be modelled in the astrometric solution (Sect. 3.3). For *Gaia* DR2, all image parameters have been re-determined in a uniform way and recovering observations that for various reasons did not enter *Gaia* DR1. The sky background has been recalibrated, and we now have a far more detailed calibration of the electronic bias of the CCDs (Hambly et al. 2018). Important for sources brighter than $G \simeq 12$ is a more reliable identification of saturated samples, which are not used in the PSF fitting.

All observations provide an along-scan (AL) measurement, consisting of the precise time at which the image centroid passes a fiducial line on the CCD. The two-dimensional windows, mainly used for bright sources ($G \lesssim 13$), provide in addition

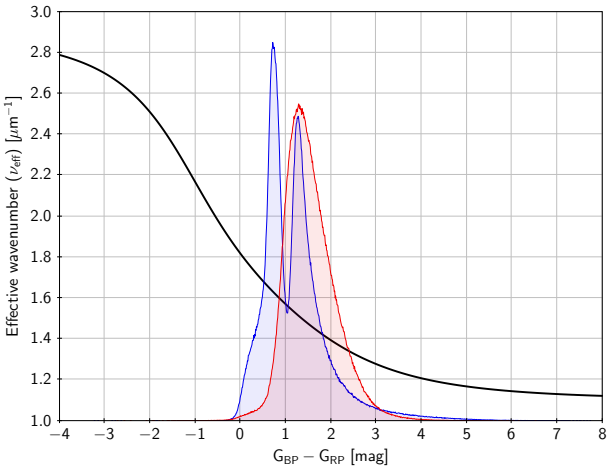

Fig. 1. Effective wavenumber as a function of colour index. The curve is the analytical relation in Eq. (2). We also show the distribution of $G_{\rm BP} - G_{\rm RP}$ for a random selection of bright ($G < 12$ mag, bluish histogram with two peaks) and faint ($G > 18$ mag, reddish histogram) sources.

a less precise across-scan (AC) measurement from the pixel column of the image centroid. A singe transit over the field of view thus generates ten AL measurements and one or ten AC measurements, although some of them may be discarded in the subsequent processing. The first observation in a transit is always made with the sky mapper (SM); it is two-dimensional, but less precise in both AL and AC than the subsequent observations in the astrometric field (AF) because of the special readout mode of the SM detectors. Only AF observations are used in the astrometric solutions. All measurements come with a formal uncertainty estimated by the image parameter determination. Based on the photon-noise statistics, the median formal AL uncertainty is about 0.06 mas per CCD observation in the AF for $G < 12$ mag, 0.20 mas at $G = 15$ mag, and 3.8 mas at $G = 20$ mag (cf. Fig. 9).

2.3. Colour information

The chromaticity calibration (Sect. 3.3) requires that the effective wavenumber $\nu_{\rm eff} = \langle \lambda^{-1} \rangle$ is known for all primary sources. For *Gaia* DR2, this quantity was computed from the mean integrated $G_{\rm BP}$ and $G_{\rm RP}$ magnitudes provided by the photometry pipeline (Riello et al. 2018), using the formula

$$\nu_{\rm eff}\,[\mu{\rm m}^{-1}] = 2.0 - \frac{1.8}{\pi}\arctan(0.331 + 0.572C - 0.014C^2 + 0.045C^3)\,, \quad (2)$$

where $C = G_{\rm BP} - G_{\rm RP}$ (Fig. 1). The arctan transformation constrains $\nu_{\rm eff}$ to the interval $[1.1,\ 2.9]\,\mu{\rm m}^{-1}$ (roughly corresponding to the passband of G, or $\simeq$ 340–910 nm) as a safeguard against spurious extreme values of C. The polynomial coefficients are based on pre-launch calibrations of the photometric bands and standard stellar flux libraries. In future releases, more accurate values of $\nu_{\rm eff}$ may be computed directly from the calibrated BP and RP spectra.

2.4. BAM data

The basic angle monitor (BAM) is an interferometric device measuring short-term ($\lesssim$ 1 day) variations of the basic angle at μas precision (Mora et al. 2016). Similarly to what was done for *Gaia* DR1 (Appendix A.2 in Lindegren et al. 2016), the BAM data are here used to correct the astrometric measurements for the rapid variations (in particular the ~1 mas amplitude 6 h oscillations) not covered by the astrometric calibration model. However, the corrections are considerably more detailed for *Gaia* DR2, taking advantage of several improvements in the processing and analysis of the BAM data: cosmic-ray filtering at pixel level of the raw BAM data; use of cross-correlation to determine very precise relative fringe phases; improved modelling of discontinuities and other variations that cannot be represented by the simple harmonic model used for *Gaia* DR1 (cf. Figs. A.2 and A.3 in Lindegren et al. 2016). Some 370 basic-angle jumps with a median amplitude of 45 μas are corrected in this way. The jumps appear seemingly at random times, but at a much increased rate in the weeks following a decontamination event. The jumps, plus the smoothed BAM data between jumps, provided the basic-angle corrector for *Gaia* DR2 in the form of a spline function of time.

The spin-related distortion model (Sect. 3.4) provides certain global corrections to the BAM data, derived from the astrometric observations, but cannot replace the BAM data, which contain a host of more detailed information such as the jumps.

3. Models

3.1. Source model

The *Gaia* data processing is based on a consistent theory of relativistic astronomical reference systems (Soffel et al. 2003). Relevant components of the model are gathered in the *Gaia* relativity model (GREM; Klioner 2003, 2004). The primary coordinate system is the Barycentric Celestial Reference System (BCRF) with origin at the solar system barycentre and axes aligned with the International Celestial Reference System (ICRS). The time-like coordinate of the BCRS is the barycentric coordinate time (TCB).

The astrometric solutions described in this paper always assume that the observed centre of the source moves with uniform space motion relative to the solar system barycentre. (Non-linear motions caused by binarity and other perturbations require special solutions that will be included in future *Gaia* releases.) The relevant source model is described in Sect. 3.2 of the AGIS paper and is not repeated here. It depends on six kinematic parameters per source, that is, the standard five astrometric parameters (α, δ, ϖ, $\mu_{\alpha*}$, and μ_δ), and the radial velocity v_r. The astrometric parameters in *Gaia* DR2 refer to the reference epoch J2015.5 = JD 2457 206.375 (TCB) = 2015 July 2, 21:00:00 (TCB). The positions and proper motions refer to the ICRS thanks to the special frame alignment procedure (Sect. 5.1).

The source model allows taking into account perspective acceleration through terms depending on the radial velocity v_r. The accumulated effect over a time interval T is $\Delta = |v_r|\mu\varpi T^2/A_{\rm u}$, where $\mu = (\mu_{\alpha*}^2 + \mu_\delta^2)^{1/2}$ is the total proper motion and $A_{\rm u}$ is the astronomical unit. This is negligible except for some very nearby high-velocity stars, and for nearly all sources we ignore the effect by setting $v_r = 0$ in the astrometric processing. Only for 53 nearby Hipparcos sources was it taken into account by assuming non-zero values of v_r taken from the literature (SIMBAD; Wenger et al. 2000). These sources were selected as having a predicted $\Delta > 0.023$ mas for $T = 1.75$ yr, calculated from Hipparcos astrometry (van Leeuwen 2007). (The somewhat arbitrary limit 0.023 mas corresponds to an RMS modelling error below 0.002 mas, which is truly insignificant for this release.) The top ten cases are listed in Table 1. In future releases, perspective acceleration will be taken into account whenever possi-

Table 1. Ten Hipparcos sources in Gaia DR2 with the largest predicted perspective acceleration.

Designation	HIP	v_r [km s^{-1}]	Δ [mas]	Name
Gaia DR2 4472832130942575872	87937	−110.51	1.975	Barnard's star
Gaia DR2 4810594479417465600	24186	245.19	1.694	Kapteyn's star
Gaia DR2 2552928187080872832	3829	263.00	0.573	Van Maanen 2
Gaia DR2 1872046574983507456	104214	−65.74	0.313	61 Cyg A
Gaia DR2 1872046574983497216	104217	−64.07	0.297	61 Cyg B
Gaia DR2 4034171629042489088	57939	−98.35	0.239	Groombridge 1830
Gaia DR2 5853498713160606720	70890	−22.40	0.208	α Cen C (Proxima)
Gaia DR2 6412595290592307840	108870	−40.00	0.163	ϵ Ind
Gaia DR2 3340477717172813568	26857	105.83	0.144	Ross 47
Gaia DR2 4847957293277762560	15510	87.40	0.141	e Eri

Notes. The table gives the assumed radial velocity, v_r (taken from SIMBAD, Wenger et al. 2000), for 10 of the 53 Hipparcos sources where the perspective acceleration was taken into account in the astrometric solutions. Δ is the predicted size of the effect calculated as described in the text. The complete table of the 53 sources is given in the *Gaia* DR2 online documentation.

ble, using radial-velocity data from *Gaia*'s onboard spectrometer (RVS; Sartoretti et al. 2018). We note that 34 of the 53 sources have radial velocities from the RVS in this release, with a median absolute deviation of 0.6 km s^{-1} from the values used here. The absolute difference exceeds 5 km s^{-1} in only four cases, the most extreme being HIP 47425 = *Gaia* DR2 5425628298649940608 with $v_r = +142 \pm 21$ km s^{-1} from SIMBAD, based on Rodgers & Eggen (1974), and $v_r = +17.8 \pm 0.2$ km s^{-1} in *Gaia* DR2. In none of the cases will the error in v_r cause an astrometric effect exceeding 0.02 mas in the present reduction.

The final secondary solution (Sect. 4.2) requires knowledge of $\nu_{\rm eff}$ for all sources in order to take the chromaticity into account. For most but not all sources, this is known from the photometric processing as described in Sect. 2.3. Given the calibrated chromaticity, it is also possible, however, to obtain an astrometric estimate of $\nu_{\rm eff}$ for every source by formally introducing it as an additional (sixth) astrometric source parameter. The resulting estimate of $\nu_{\rm eff}$, called pseudo-colour, is much less precise than the $\nu_{\rm eff}$ calculated from $G_{\rm BP} - G_{\rm RP}$ using Eq. (2), but has the advantage that it can be obtained for every source allowing a five-parameter solution. Moreover, it is not affected by the BP/RP flux excess issue (Evans et al. 2018), which tends to make faint sources in crowded areas too blue as measured by the $G_{\rm BP} - G_{\rm RP}$.

To ensure the most uniform astrometric treatment of sources, the pseudo-colour was consistently used as a proxy for $\nu_{\rm eff}$ in all cases where *Gaia* DR2 provides a five-parameter solution, that is, even when photometric colours are available. Because it is so important for the astrometry, the pseudo-colour is given in the *Gaia* Archive as `astrometric_pseudo_colour`. Normally, it does not provide an astrophysically useful estimate of the colour because its precision is much lower than the photometric data.

Our treatment of the pseudo-colour as a sixth source parameter should not be confused with the use of the radial proper motion $\mu_r = v_r\varpi/A_{\rm u}$ in the kinematic source model (e.g. Eq. 2 of Lindegren et al. 2016). This quantity, sometimes referred to as the "sixth astrometric parameter", is used internally in AGIS to take into account the perspective acceleration, but is never explicitly estimated as an astrometric parameter.

3.2. Attitude model

The attitude specifies the orientation of the optical instrument in ICRS as a function of time. Mathematically, it is given by the unit quaternion $\mathbf{q}(t)$. The attitude model described in Sect. 3.3 of the AGIS paper represents the time-dependent components of $\mathbf{q}(t)$ as cubic splines. For *Gaia* DR1, a knot interval of about 30 s was used in the splines, but it was noted that a much shorter knot interval (i.e. more flexible splines) would actually be needed to cope with the considerable attitude irregularities on shorter timescales, including a large number of "micro-events" such as the very frequent micro-clanks (see Appendices C.4 and E.4 in Lindegren et al. 2016) and less frequent micrometeoroid hits. Decreasing the knot interval of the splines is not a good way forward, however, as it would weaken the solution by the increased number of attitude parameters. Moreover, this cannot adequately represent the CCD-integrated effects of the micro-events, which depend also on the gate (g) used for an observation. For *Gaia* DR2 the attitude model includes a new layer, known as the corrective attitude $\mathbf{q}_{\rm c}(t, g)$, such that the (gate-dependent) effective attitude becomes

$$\mathbf{q}_{\rm e}(t,\, g) = \mathbf{q}_{\rm p}(t)\, \mathbf{q}_{\rm c}(t,\, g)\,. \tag{3}$$

Here $\mathbf{q}_{\rm p}(t)$ is the primary attitude: this uses the same spline representation as the old attitude model, and its parameters are estimated in the primary solution in a similar way as before, the main difference being that the field angle residuals (Eqs. 25–26 in the AGIS paper) are now computed using the effective attitude $\mathbf{q}_{\rm e}(t,\, g)$ for the relevant gate. The effective attitude represents the mean pointing of the instrument during the CCD integration interval, which is different depending on g.

In Eq. (3) the corrective attitude $\mathbf{q}_{\rm c}$ represents a small time- and gate-dependent rotation that takes care of attitude irregularities that are too fast for the spline model. It is calculated in the AGIS pre-processor and remains fixed during subsequent astrometric solutions. For details about its calculation, we refer to the *Gaia* DR2 online documentation. Briefly, the procedure includes the following steps:

1. Given two successive CCD observations in the astrometric field (AF) of the same source, with observation times t_k and t_{k+1}, an estimate of the inertial angular rate along the nominal spin axis z (in the scanning reference system, SRS) is obtained as

$$\overline{\omega}_z = -\frac{\eta_{k+1} - \eta_k}{t_{k+1} - t_k} + \left(\omega_x \cos\varphi + \omega_y \sin\varphi\right) \tan\zeta\,, \tag{4}$$

where η_k and η_{k+1} are the AL field angles calculated from a preliminary geometrical model of the instrument. The minus

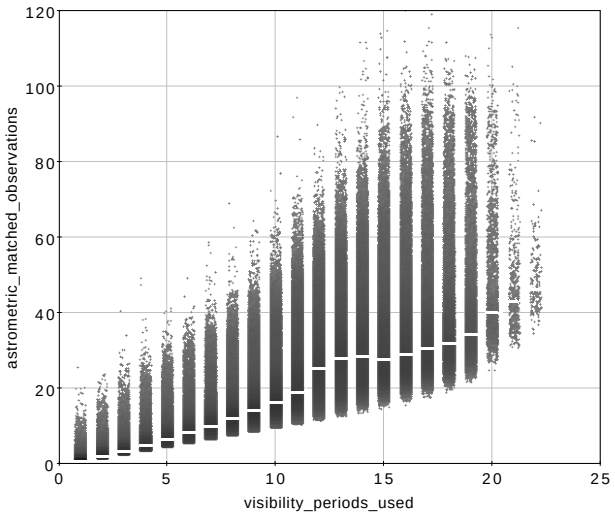

Fig. 2. Relation between the number of visibility periods and field-of-view transits (matched observations) per source used in the secondary astrometric solutions. A small random number was added to the integer number of visibility periods to widen the vertical bars. The white horizontal line through each bar shows the location of the median. The diagram was constructed for a random subset of about 2.5 million sources.

ondary solution for a particular source. A visibility period is a group of observations separated from other groups by a gap of at least four days. This statistic is a better indicator of an astrometrically well-observed source than for example `astrometric_matched_observations` (the number of field-of-view transits used in the solution): while a five-parameter solution is in principle possible with fewer than ten field-of-view transits, such a solution will be very unreliable unless the transits are well spread out in time. As illustrated in Fig. 2, there are many sources with >10 transits concentrated in just a few visibility periods.

– `astrometric_sigma5d_max` is a five-dimensional equivalent to the semi-major axis of the position error ellipse and is useful for filtering out cases where one of the five parameters, or some linear combination of several parameters, is particularly bad. It is measured in mas and computed as the square root of the largest singular value of the scaled 5×5 covariance matrix of the astrometric parameters. The matrix is scaled so as to put the five parameters on a comparable scale, taking into account the maximum along-scan parallax factor for the parallax and the time coverage of the observations for the proper motion components. If $\boldsymbol{C}$ is the unscaled covariance matrix, the scaled matrix is $\boldsymbol{SCS}$, where $\boldsymbol{S} = \mathrm{diag}(1, 1, \sin\xi, T/2, T/2)$, $\xi = 45°$ is the solar aspect angle in the nominal scanning law, and $T = 1.75115$ yr the time coverage of the data used in the solution. `astrometric_sigma5d_max` was not corrected for the DOF bug, as that would obscure the source selection made at an earlier stage based on the uncorrected quantity.

The five-parameter solution was accepted if the following conditions were all met for the source:

$$\left.\begin{array}{ll} \text{(i)} & \text{mean magnitude } G \le 21.0 \\ \text{(ii)} & \texttt{visibility\_periods\_used} \ge 6 \\ \text{(iii)} & \texttt{astrometric\_sigma5d\_max} \le (1.2\ \text{mas}) \times \gamma(G) \end{array}\right\}, \tag{11}$$

where $\gamma(G) = \max[1,\ 10^{0.2(G-18)}]$. The upper limit in (iii) gradually increases from 1.2 mas for $G \le 18$ to 4.78 mas at $G = 21$. This test was applied using preliminary G magnitudes, with the result that some sources in *Gaia* DR2 have five-parameter solutions even though they do not satisfy (iii).

If the five-parameter solution was rejected by Eq. (11), a fall-back solution was attempted as previously described. The resulting position, referring to the epoch J2015.5, was accepted provided that the following conditions are all met:

$$\left.\begin{array}{ll} \text{(i)} & \texttt{astrometric\_matched\_observations} \ge 5 \\ \text{(ii)} & \texttt{astrometric\_excess\_noise} < 20\ \text{mas} \\ \text{(iii)} & \sigma_{\mathrm{pos,\,max}} < 100\ \text{mas} \end{array}\right\}. \tag{12}$$

`astrometric_excess_noise` is the excess source noise ϵ_i introduced in Sect. 3.6 of the AGIS paper, and $\sigma_{\mathrm{pos,\,max}}$ is the semi-major axis of the error ellipse in position given by Eq. (B.1). Sources rejected also by Eq. (12) are mostly spurious and no results are published for them.

These criteria resulted in 1335 million sources with a five-parameter solution and 400 million with a fall-back solution, that is, without parallax and proper motion. About 18 million sources were subsequently removed as duplicates, that is, where the observations of the same physical source had been split between two or more different source identifiers. Duplicates were identified by positional coincidence, using a maximum separation of 0.4 arcsec. To decide which source to keep, the following order of preference was used: unconditionally keep any source (quasar) used for the reference frame alignment; otherwise prefer a five-parameter solution before a fall-back solution, and keep the source with the smallest `astrometric_sigma5d_max` to break a tie.

Gaia DR2 finally gives five-parameter solutions for 1332 million sources, with formal uncertainties ranging from about 0.02 mas to 2 mas in parallax and twice that in annual proper motion. For the 361 million sources with fall-back solutions, the positional uncertainty at J2015.5 is about 1 to 4 mas. Further statistics are given in Appendix B.

5. Internal validation

This section summarises the results of a number of investigations carried out by the DPAC astrometry team in order to validate the astrometric solutions. This aimed in particular at characterising the systematic errors in parallax and proper motion, and the realism of the formal uncertainties. Some additional quality indicators are discussed in Appendix C.

5.1. Reference frame

The celestial reference frame of *Gaia* DR2, known as *Gaia*-CRF2 (Gaia Collaboration et al. 2018b), is nominally aligned with ICRS and non-rotating with respect to the distant universe. This was achieved by means of a subset of 492 006 primary sources assumed to be quasars. These included 2843 sources provisionally identified as the optical counterparts of VLBI sources in a prototype version of ICRF3, and 489 163 sources found by cross-matching AGIS02.1 with the AllWISE AGN catalogue (Secrest et al. 2015, 2016). The unpublished prototype ICRF3 catalogue (30/06/2017, solution from GSFC) contains accurate VLBI positions for 4262 radio sources and was kindly made available to us by the IAU Working Group *Third Realisation of International Celestial Reference Frame*.

The radius for the positional matching was 0.1 arcsec for the VLBI sources and 1 arcsec for the AllWISE sample. Apart from

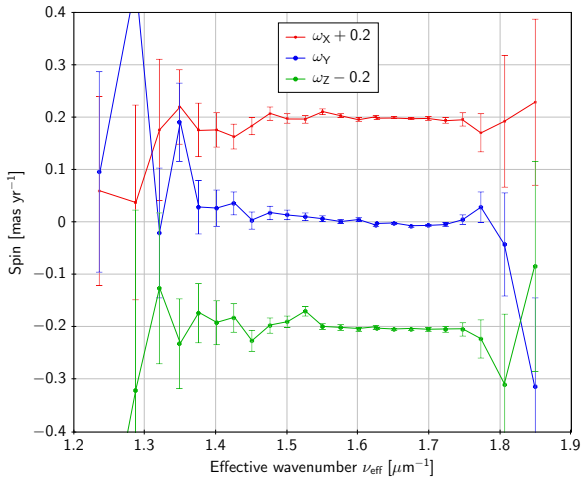

Fig. 3. Dependence of the faint reference frame on colour. The diagram shows the components of spin ω_X, ω_Y, and ω_Z around the ICRS axes, as estimated for faint ($G \simeq 15$–21) quasars subdivided by effective wavenumber. The components in X and Z were shifted by ±0.2 mas yr^{-1} for better visibility. Error bars are at 68% confidence intervals for the estimated spin.

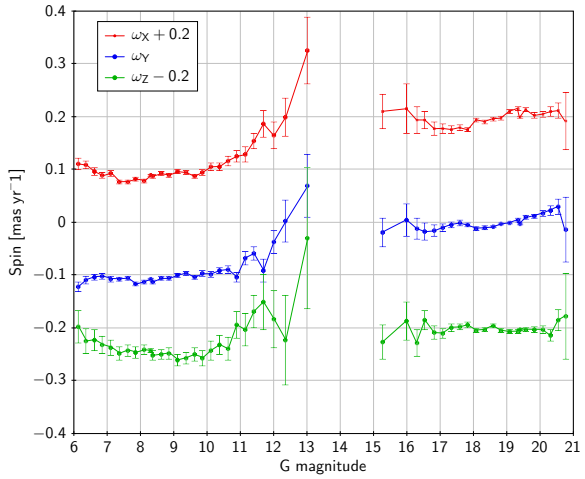

Fig. 4. Dependence of the reference frame on magnitude. The diagram shows the spin components as in Fig. 3, but subdivided by magnitude. The points at the faint end ($G \gtrsim 15$) are estimated from the proper motions of quasars. At the bright end ($G \lesssim 13$), the spin is estimated from the differences in stellar proper motions between *Gaia* DR2 and the Hipparcos subset of TGAS in *Gaia* DR1.

the positional coincidence, the joint application of the following conditions reduced the risk of contamination by Galactic stars:

$$\left.\begin{array}{ll} \text{(i)} & \texttt{astrometric\_matched\_observations} \geq 8 \\ \text{(ii)} & \varsigma_\varpi < 1 \text{ mas} \\ \text{(iii)} & |\varpi/\varsigma_\varpi| < 5 \\ \text{(iv)} & (\mu_{\alpha*}/\varsigma_{\mu\alpha*})^2 + (\mu_\delta/\varsigma_{\mu\delta})^2 < 25 \\ \text{(v)} & |\sin b| > 0.1 \end{array}\right\}, \qquad (13)$$

where b is Galactic latitude. We used the formula $\sin b = (-0.867666\cos\alpha - 0.198076\sin\alpha)\cos\delta + 0.455984\sin\delta$, which is accurate to about 0.1 arcsec. These conditions were applied to both samples, except that (v) was not used for the VLBI sample where the risk of contamination is much lower thanks to the smaller positional match radius.

The selection of sources for the frame rotator described above was made before the final solution had been computed and therefore used preliminary values for the various quantities in Eq. (13), including standard uncertainties (ς) not yet corrected for the DOF bug. The resulting subsets of sources are indicated in the *Gaia* Archive by the field `frame_rotator_object_type`, which is 2 for the 2843 sources matched to the ICRF3 prototype, 3 for the 489 163 sources matched to the AllWISE AGN catalogue, and 0 for sources not used by the frame rotator. The magnitude distributions of these subsets are shown in Fig. B.1. It can be noted that the AllWISE sample (labelled "QSO" in the diagram) contains three bright sources ($G < 12$) that are probably distant Galactic stars of unusual colours (the brightest being the Herbig AeBe star HD 37357). These objects are not included in the larger but cleaner quasar sample analysed in Sect. 5.2, obtained by applying the stricter Eq. (14) to the final data.

The adjustment of the reference frame was done in the primary solution (step 4 of Sect. 4.2) using the frame rotator described in Sect. 6.1 of the AGIS paper. At the end of an iteration, the frame rotator estimated the frame orientation parameters $[\epsilon_X, \epsilon_Y, \epsilon_Z]$ at J2015.5, using the VLBI sources, and the spin parameters $[\omega_X, \omega_Y, \omega_Z]$ using the AllWISE and VLBI sources. The attitude and the positions and proper motions of the primary sources were then corrected accordingly. The acceleration parameters $[a_X, a_Y, a_Z]$ were not estimated as part of this process, as they are expected to be insignificant compared with the current level of systematics (see below).

At the end of the primary solution, the attitude was thus aligned with the VLBI frame, and the subsequent secondary solutions (step 5 of Sect. 4.2) should then result in source parameters in the desired reference system. This was checked by a separate off-line analysis, using independent software and more sophisticated algorithms. This confirmed the global alignment of the positions with the VLBI to within ±0.02 mas per axis. This applies to the faint reference frame represented by the VLBI sample with a median magnitude of $G \simeq 18.8$. The bright reference frame was checked by means of some 20 bright radio stars with accurate VLBI positions and proper motions collected from the literature. Unfortunately, their small number and the sometimes large epoch difference between the VLBI observations and *Gaia*, combined with the manifestly non-linear motions of many of the radio stars, did not allow a good determination of the orientation error of the bright reference frame of *Gaia* DR2 at epoch J2015.5. No significant offset was found at an upper (2σ) limit of about ±0.3 mas per axis.

Concerning the spin of the reference frame relative to the quasars, estimates of $[\omega_X, \omega_Y, \omega_Z]$ using various weighting schemes and including also the acceleration parameters confirmed that the faint reference frame of *Gaia* DR2 is globally non-rotating to within ±0.02 mas yr^{-1} in all three axes. Particular attention was given to a possible dependence of the spin parameters on colour (using the effective wavenumber $\nu_{\rm eff}$) and magnitude (G). Figure 3 suggests a small systematic dependence on colour, for example, by ±0.02 mas yr^{-1} over the range $1.4 \lesssim \nu_{\rm eff} \lesssim 1.8\ \mu$m^{-1} corresponding to roughly $G_{\rm BP} - G_{\rm RP} = 0$ to 2 mag. As this result was derived for quasars that are typically fainter than 15th magnitude, it does not necessarily represent the quality of the *Gaia* DR2 reference frame for much brighter objects.

Figure 4 indeed suggests that the bright ($G \lesssim 12$) reference frame of *Gaia* DR2 has a significant (~0.15 mas yr^{-1}) spin relative to the fainter quasars. The points in the left part of the diagram were calculated from stellar proper motion differences between the current solution and *Gaia* DR1 (TGAS). Only 88 091 sources in the Hipparcos subset of TGAS were used for this comparison owing to their superior precision in TGAS.

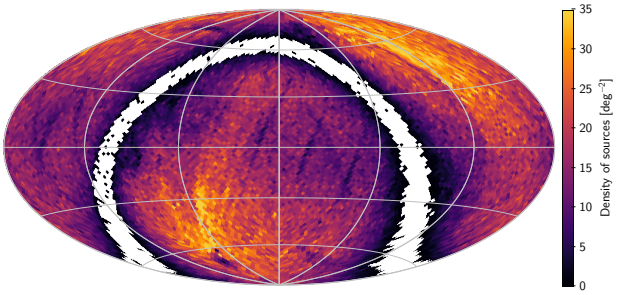

Fig. 5. Density map of the full quasar sample (union of AllWISE AGNs and VLBI sources) at a resolution of 1.8×1.8 deg^2. The scatter of points in the Galactic band are VLBI sources. This and following full-sky maps use a Hammer–Aitoff projection in equatorial (ICRS) coordinates with $\alpha = \delta = 0$ at the centre, north up, and α increasing from right to left.

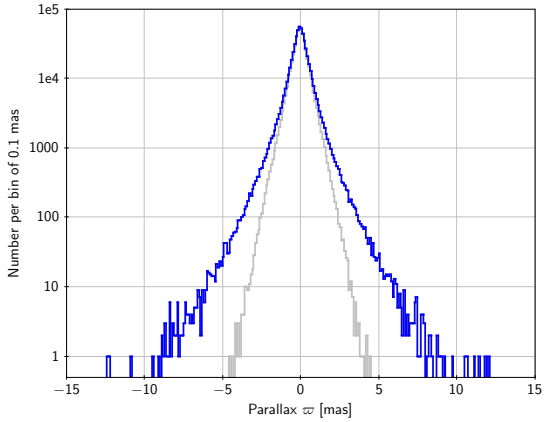

Fig. 6. Parallax distribution for 556 869 sources identified as quasars. Outer (blue) curve: the whole sample; inner (grey) curve: the subsample of 492 928 sources with $\sigma_\varpi < 1$ mas.

Although based on a much shorter stretch of *Gaia* observations than the present solution, TGAS provides a valuable comparison for the proper motions thanks to its ∼24 yr time difference from the Hipparcos epoch. If the spin difference of 0.15 mas yr^{-1} between the two catalogues were to be explained as systematics in TGAS, it would require an alignment error of ∼3.6 mas in the positions either in TGAS at epoch J2015.0 or in Hipparcos at epoch J1991.25. Given the way these catalogues were constructed, both hypotheses are very unlikely. The most reasonable explanation for the offsets in Fig. 4 is therefore systematics in the *Gaia* DR2 proper motions of the bright sources. The gradual change between magnitudes 12 and 10 suggests an origin in the gated observations, which dominate for $G \lesssim 12$, or possibly in observations of window class 0, which dominate for $G \lesssim 13$.

Formally, *Gaia*-CRF2 is materialised by the positions in *Gaia* DR2 of the 556 869 sources identified as quasars in Sect. 5.2. A separate list of these sources is provided in the *Gaia* Archive. A more comprehensive analysis of *Gaia*-CRF2 is given by Gaia Collaboration et al. (2018b).

5.2. Parallax zero point

Global astrometric satellites like Hipparcos and *Gaia* are able to measure absolute parallaxes, that is, without zero-point error, but this capability is susceptible to various instrumental effects, in particular, to a certain kind of basic-angle variations. As discussed by Butkevich et al. (2017), periodic variations of the basic angle (Γ) of the form $\delta\Gamma(t) = A_1 d(t)\cos\Omega(t)$, where $d(t)$ is the distance of *Gaia* from the solar system barycentre in au and $\Omega(t)$ is the spin phase relative to the barycentre, are observationally almost indistinguishable from a global parallax shift of $\delta\varpi = A_1/[2\sin\xi\sin(\Gamma/2)] \simeq 0.883A_1$. This is clearly reminiscent of the first term in Eq. (10). Although d, ξ, and Ω in that equation are heliotropic quantities, while the present formula uses barytropic quantities, and d appears with different powers in the two expressions, the differences are small enough to cause a near-degeneracy between A_1 and $\delta C_{1,0}$. This is the reason why the latter parameter was not estimated in the spin-related distortion model.

It is believed that the basic-angle corrector derived from BAM data (Sect. 2.4) eliminates basic-angle variations very efficiently, but a remaining small variation corresponding to the undetermined $\delta C_{1,0}$ cannot be excluded. This would then show up as a small offset in the parallaxes. For this reason, it is extremely important to investigate the parallax zero point by external means, that is, using astrophysical sources with known parallaxes. It is also important to check possible dependences of the zero point on other factors such as position, magnitude, and colour, which could be created by errors in the calibration model.

The quasars are almost ideal for checking the parallax zero point thanks to their extremely small parallaxes (< 0.0025 μas for redshift $z > 0.1$), large number, availability over most of the celestial sphere, and, in most cases, nearly point-like appearance. The main drawbacks are their faintness and peculiar colours.

In order to create the largest possible quasar sample for validation purposes, a new cross-match of the final *Gaia* DR2 data with the AllWISE AGN catalogue (Secrest et al. 2015) was made, choosing in each case the nearest positional match. The further selection used the criteria

$$\left.\begin{array}{ll}
\text{(i)} & \texttt{astrometric_matched_observations} \geq 8 \\
\text{(ii)} & \texttt{astrometric_params_solved} = 31 \\
\text{(iii)} & |(\varpi + 0.029\ \text{mas})/\sigma_\varpi| < 5 \\
\text{(iv)} & (\mu_{\alpha*}/\sigma_{\mu\alpha*})^2 + (\mu_\delta/\sigma_{\mu\delta})^2 < 25 \\
\text{(v)} & |\sin b| > 0.1 \\
\text{(vi)} & \rho < (2\ \text{arcsec}) \times |\sin b|
\end{array}\right\}, \qquad (14)$$

which is somewhat similar to Eq. (13), but stricter and applied to the final data. Step (ii) selects five-parameter solutions ($31 = 11111_2$), and step (iii) takes into account the median offset of the parallaxes (see below). The combination of steps (v) and (vi) makes the probability of a chance match with a Galactic star generally lower than $\sim10^{-4}$ at all Galactic latitudes. A reality check of the resulting selection against SIMBAD revealed that the two brightest sources (at $G = 8.85$ and 11.72 mag) are stars; removing them leaves 555 934 sources in the sample. The fraction of stars among the AllWISE AGN sources is estimated at $\leq 0.041\%$ (Secrest et al. 2015), or $\lesssim 230$ in this sample, but only a fraction of them may pass the criteria in Eq. (14).

Applying conditions (i)–(iv) to the sources matched to the ICRF3 prototype (Sect. 5.1) gave 2820 sources, 1885 of which were already in the AllWISE sample. The union set thus contains a total of 556 869 sources, which also define the celestial reference frame of *Gaia* DR2 (Gaia Collaboration et al. 2018b). A density map of this quasar sample (Fig. 5) shows imprints of the *Gaia* and AllWISE scanning laws as well as the effects of Galactic extinction and confusion. In the following, the high-precision subset of 492 928 sources with $\sigma_\varpi < 1$ mas is sometimes used instead of the full quasar sample.

Figure 6 shows the distribution of parallaxes for the full quasar sample and the high-precision subset. For the full sample,

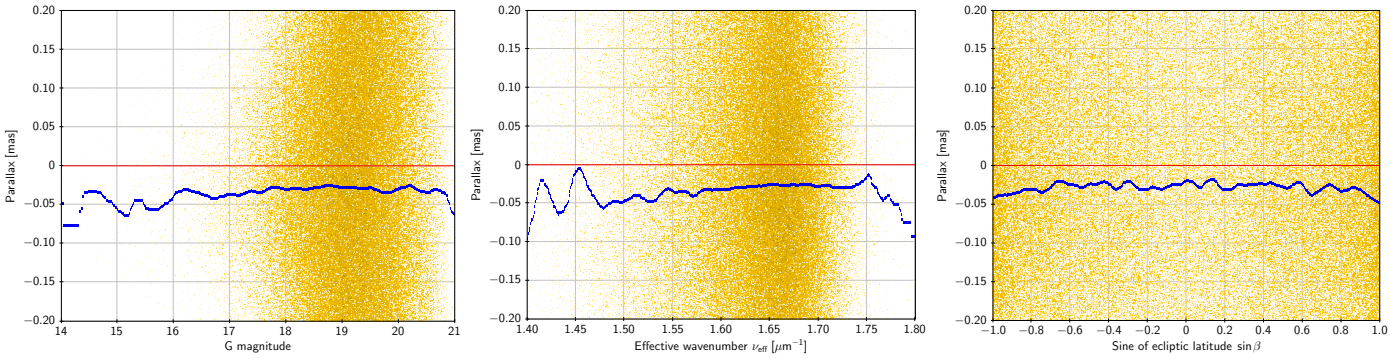

Fig. 7. Parallaxes for the full quasar sample plotted against magnitude (left), colour (middle), and ecliptic latitude (right). Because of the chosen scale, only about one-third of the data points are shown as yellow dots; the blue curves are the running medians.

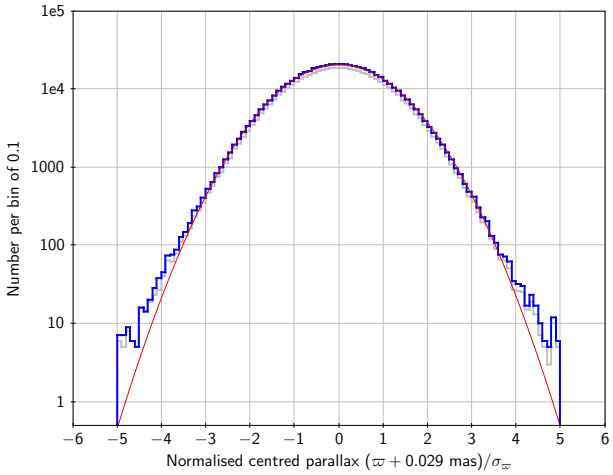

Fig. 8. Distributions of the normalised centred parallaxes for the same samples as in Fig. 6. The red curve is a Gaussian distribution with the same standard deviation (1.081) as the normalised centred parallaxes for the full sample.

the mean and median parallax is −0.0308 mas and −0.0287 mas, respectively; for the high-precision subset, the corresponding values are −0.0288 mas and −0.0283 mas. For the subsequent analysis we adopt −0.029 mas as the global zero point of the parallaxes. Scatter plots of the parallaxes versus magnitude and colour (left and middle panels of Fig. 7) show systematic trends with a change of ∼0.02 mas over the ranges covered by the data. A plot against ecliptic latitude (right panel) shows a roughly quadratic variation with ∼0.010 mas smaller parallaxes towards the ecliptic poles. Thus, while the global mean offset of −0.029 mas is statistically well-determined, the actual offset applicable for a given combination of magnitude, colour, and position may be different by several tens of μas. Spatial variations of the parallax zero point are further analysed in Sect. 5.4.

Figure 8 shows the distribution of $(\varpi + 0.029\ \mathrm{mas})/\sigma_\varpi$, that is, the parallaxes corrected for the global offset and normalised by the formal uncertainties. Ideally, this should follow a normal distribution with zero mean and unit variance. The actual sample standard deviation of this quantity is 1.081. Similarly, the sample standard deviations of the normalised proper motions, $\mu_{\alpha*}/\sigma_{\mu\alpha*}$ and $\mu_\delta/\sigma_{\mu\delta}$, are 1.093 and 1.115, respectively. The distributions are very close to normal, as suggested by the red curve in Fig. 8, although it should be noted that the selection in Eq. (14) removed any point beyond ±5 units in the normalised quantities. The conclusion is that the accidental errors are close to normal, but with a standard deviation some 8–12% larger than the formal uncertainties. This applies to the faint sources ($G \gtrsim 15$) beyond the Galactic plane ($|\sin b| > 0.1$) represented by the quasar subset.

The observations contributing to the parallax determinations are distributed roughly uniformly over the 62 CCDs in the central 0.7° × 0.7° astrometric field of the *Gaia* instrument. The basic-angle variation relevant for the parallax zero point is therefore effectively given by the average variation in this field. On the other hand, the CCD generating the BAM data is situated about 0.7° from the centre of the astrometric field, that is, well outside the field near one of its corners. The corrections given in Table 3 show that the variations measured by the BAM are not fully representative of the variations present in the astrometric field. It is noted that a parallax zero point of −29 μas corresponds to a value $\simeq -33$ μas for the undetermined correction $\delta C_{1,0}$ in Table 3.

Differential variations within the astrometric field depending on Ω are described by the global parameters c_{fklm}, s_{fklm} in Eq. (9), which are estimated in the primary solution. In principle, this allows the differential variations to be extrapolated to the location of the BAM. Although such a procedure is clearly problematic, it could provide an independent estimate of the crucial parameter $\delta C_{1,0}$ and important consistency checks for other parameters. A detailed investigation along these lines will only be meaningful at a later time when other calibration errors have been substantially eliminated. With the current solution, we note that the largest amplitudes $|c_{fklm}|$, $|s_{fklm}|$ are associated with the lowest temporal (k) and spatial ($l + m$) orders, as would be expected for a physical instrument. Moreover, their sizes (0.01 to 0.05 mas) are in the approximate range needed to account for the corrections to the BAM data reported in Table 3 as well as the global parallax offset of −0.029 mas. However, there could be many other explanations for this offset; in particular, it appears that unmodelled AL centroid shifts related to the transverse smearing of the images during a CCD integration (depending on the AC rate $\mathrm{d}\zeta/\mathrm{d}t$) could be an important contributor (Sect. 5.3).

5.3. Residual analysis

Analysis of the astrometric residuals can reveal inadequacies in the calibration model, for example where a new effect needs to be added or where the time granularity of some effect already included in the model has insufficient resolution. It is particularly interesting to look for model deficiencies that might explain the systematics seen in the astrometric results, for instance, the parallax zero point error. In this section we first estimate the total size of the unmodelled errors, and then give two examples of effects that contribute to the errors in the present solutions, but could be eliminated in future releases.

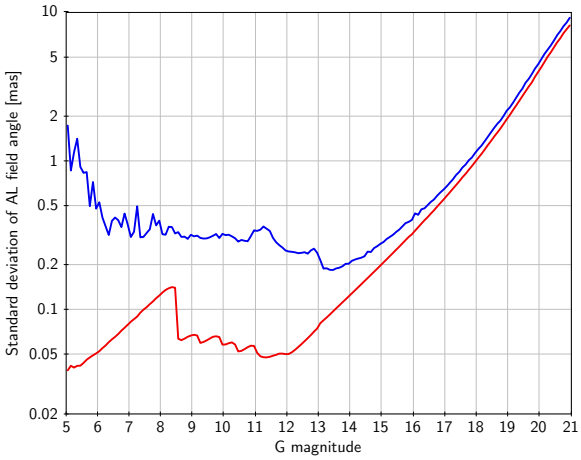

Fig. 9. Precision of along-scan astrometric measurements as a function of magnitude. The red (lower) curve is a running median of the formal precision from the image parameter determination; the blue (upper) curve is a robust estimate of the actual standard deviation of the post-fit residuals. The difference between the two curves represents the combination of all unmodelled errors.

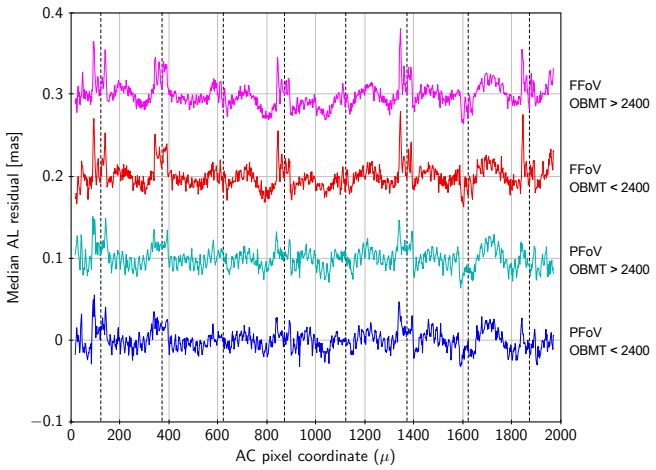

Fig. 10. Small-scale distortion for ungated observations on one of the astrometric CCDs (strip 7, row 4). The curves show the median AL residual for sources in the magnitude range G = 13–16 plotted against the AC pixel coordinate μ, and subdivided according to field of view (preceding PFoV, or following FFoV) and time (before or after the decontamination at OBMT ≃ 2400). For better visibility, the successive curves were vertically displaced by 0.1 mas. The vertical dashed lines show the stitch block boundaries, which divide the 1966 pixels in blocks of 250 pixels, except for the two outermost blocks that are 108 pixels.

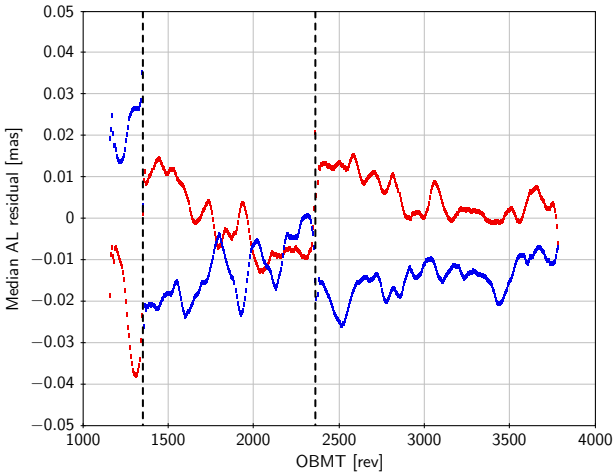

Fig. 11. Residual systematics depending on the AC scan rate. The curves show the median residual as a function of OBMT for observations of window class 1 (G ≃ 13–16) in the preceding field of view. The red curve is for observations with positive AC rate, and the blue curve for negative AC rate. The vertical dashed lines show the approximate times of the two decontamination events.

Figure 9 compares the photon-statistical uncertainties of the AL angular measurements with the scatter of post-fit residuals in the astrometric solution. The red curve is the formal precision from the image parameter determination, derived from the assumed Poissonian character of the individual CCD sample values. This curve has three domains, depending on the number of photons (N) in the stellar image: for moderately bright sources ($G \simeq$ 12–17), the centroiding precision is limited by the photon noise in the stellar image, or $\sigma \propto N^{-1/2}$, leading to a slope of about 0.2 dex mag^{-1}; for fainter sources ($G \gtrsim 17$), the background gradually becomes more important, leading to a higher slope in the red curve; finally, for the bright stars ($G \lesssim 12$), the use of the gates limits N and hence the centroiding precision to a value roughly independent of G.

The blue curve in Fig. 9 is the robust scatter estimate (RSE)[3] of the post-fit residuals, computed in bins of 0.1 mag. For faint sources, it agrees reasonably well with the formal uncertainties (for $G > 17$ the RSE is on average 15% higher than the formal uncertainties), but for brighter sources, there is a strong discrepancy. The difference between the blue and red curves represents the combination of all unmodelled source, attitude, and calibration errors. The quadratic difference amounts to about 0.3 mas for $G \simeq$ 6–12, 0.25 mas for $G \simeq$ 12–13, and 0.15 mas for $G \gtrsim 12$. Part of this may be attributable to the sources (e.g. binarity), part to residual attitude irregularities, but a major part is clearly due to inadequacies of the calibration models, including the LSF and PSF models used for the image parameter determination. A main task in preparation for future *Gaia* data releases will be to improve these models and hence reduce the gap between the two curves.

The astrometric calibration model (Sect. 3.3) currently does not include small-scale irregularities of the CCDs. To assess the importance of such errors, we plot in Fig. 10 the median AL residual, subdivided by field of view and time, as a function of the AC pixel coordinate μ. Comparing the four curves, it is seen that the pattern is extremely stable in time, but slightly different in the two fields of view. The rms amplitude is only 0.013 mas in the preceding and 0.015 mas in the following field of view, far too small to explain the discrepancy seen in Fig. 9. While the small-scale irregularities are therefore unimportant in the current solution, they will be included in future calibration models.

One of the most interesting trends revealed by the residual analysis concerns a hitherto unmodelled dependence on the across-scan rate $\mathrm{d}\zeta/\mathrm{d}t$, where ζ is the AC field angle. In the nominal scanning law, the AC rate varies sinusoidally over the 6 hr spin period with an amplitude of about ±0.18 arcsec s^{-1}, or ±0.3% of the constant AL rate (60 arcsec s^{-1}). It is in general different in the two fields of view. The AC motion of stellar images by up to 0.8 arcsec during its motion across a CCD smears the PSF in the AC direction. While this obviously has a strong

[3] The RSE is a robust measure of the dispersion of a distribution, defined as $\left(2\sqrt{2}\,\mathrm{erf}^{-1}(4/5)\right)^{-1} \simeq 0.390152$ times the difference between the 90th and 10th percentiles. For a normal distribution, the RSE equals the standard deviation.

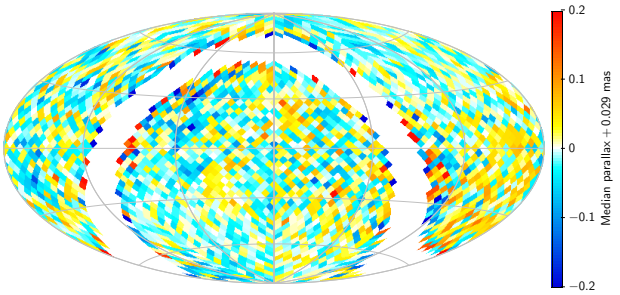

Fig. 12. Map of the median parallaxes for the full quasar sample, showing large-scale variations of the parallax zero point. See Fig. 5 for the coordinate system and density of sources. Median values are calculated in cells of about 3.7×3.7 deg^2. Only cells with $|\sin b| > 0.2$ are plotted.

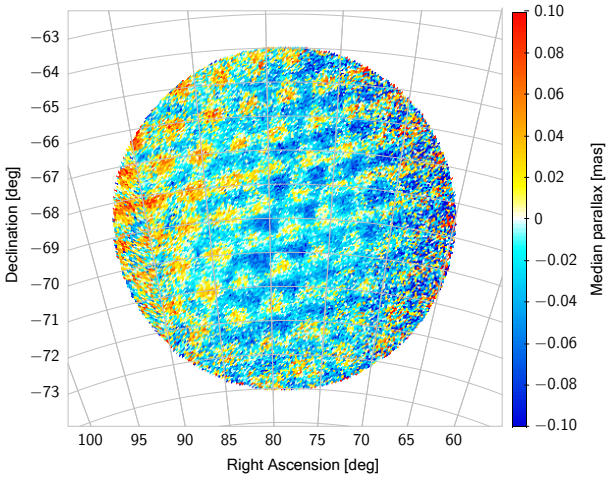

Fig. 13. Map of the median parallaxes for a sample of sources in the LMC area, showing small-scale variations of the parallax zero point. Median values are calculated in cells of about 0.057×0.057 deg^2.

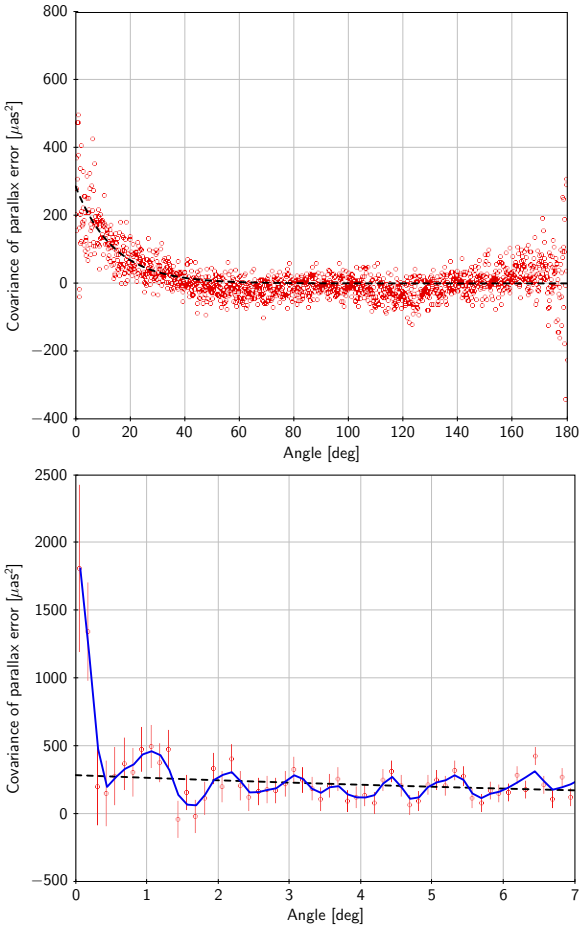

Fig. 14. *Top:* Spatial covariance $V_\varpi(\theta)$ of parallax errors in the high-precision quasar sample. Red circles are the individual estimates, and the dashed black curve shows a fitted exponential. *Bottom:* The same data for separations $< 7°$ with errors bars (68% confidence intervals) and a running triangular mean (blue curve). The two highest points, for separations $< 0.25°$, are outside the plot in the top panel.

effect on the AC location of the image, it should, to a first approximation, not affect the AL location of the centroid. However, secondary effects involving a non-symmetric PSF or non-linear response to the photon flux could easily generate a small dependence of the precise AL location on the AC rate. Figure 11 shows that this is indeed the case. Test solutions including astrometric calibration terms depending on the AC rate show reduced levels of systematics, for example in terms of the ∼ 1 deg scale correlations discussed in Sect. 5.4. AL centroiding errors depending on the AC rate are particularly insidious, as the AC rate exhibits a strong correlation with the AL parallax factor in the current nominal scanning law.

5.4. Spatial correlations

Figure 12 is a map of the median quasar parallax, adjusted for the median offset −0.029 mas, at a resolution of a few degrees. Away from the Galactic plane, where there is a sufficient density of quasars (cf. Fig. 5) for estimating a local zero point, there are several areas of a few tens of degrees where the parallaxes are systematically offset by about ±0.05 mas from the global mean. This demonstrates the presence of correlated errors on spatial scales of 10–20 deg and RMS values of a few tens of μas. Irregularities on smaller scales cannot be probed in this way using quasars, owing to their low average density.

However, distant stars in dense regions reveal significant variations on much smaller scales. As an example, Fig. 13 shows the median parallaxes for about 2.5 million sources in the area of the LMC. To remove most foreground stars, we selected sources with magnitudes between $G = 17$ and 19, within 5 deg of the LMC centre $(\alpha, \delta) = (78.77°, -69.01°)$, and with proper motions $(\mu_{\alpha*} - 1.850)^2 + (\mu_\delta - 0.233)^2 < 1$ mas^2 yr^{-2} (cf. Gaia Collaboration et al. 2018c). The mean and median values of their parallaxes are −0.014 mas, roughly consistent with the parallax zero point from quasars at the LMC location near the South Ecliptic Pole (Fig. 7, right), assuming a true parallax of 0.020 mas for the LMC (Freedman et al. 2001). The quasi-regular triangular pattern in Fig. 13 has a period of about 1 deg and a typical amplitude of about ±0.03 mas. The left part of the circular area seems to be offset by 0.02 mas from the rest with a straight and rather sharp boundary. These patterns are clearly related to *Gaia*'s scanning law with its precessional motion of about 1 deg per revolution. Similar (unphysical) patterns are seen in parallax maps of high-density areas around the Galactic centre, and also in the proper motions. Thus strong correlated errors (or systematics) also exist on spatial scales much below 1 deg.

A global, quantitative characterisation of these correlations can be obtained by calculating the covariance of the quasar parallax errors as a function of angular separation,

$$V_\varpi(\theta) = \langle (\varpi_i - \overline{\varpi})(\varpi_j - \overline{\varpi}) \rangle \,. \quad (15)$$

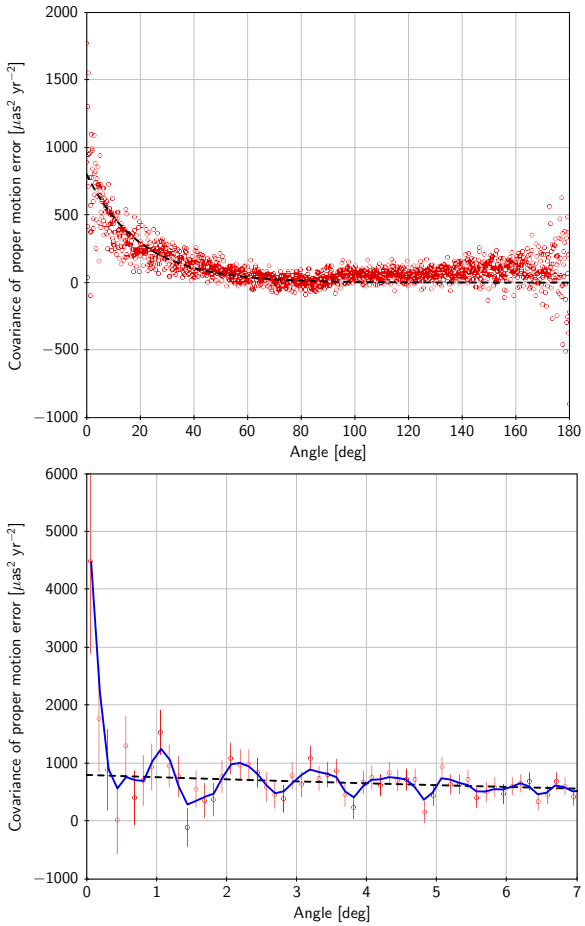

Fig. 15. Same as Fig. 14, but for the proper motions of the high-precision quasar sample ($V_\mu(\theta)$). The highest point, for the smallest separation, is outside the plot in the top panel.

Here $\overline{\varpi}$ is the mean parallax of all the quasars in the sample, and the average is taken over all non-redundant pairs of quasars ($i > j$) with angular separation $\theta \pm \Delta\theta/2$. Figure 14 shows the result of this calculation for the high-precision quasar sample, using a bin width of $\Delta\theta = 0.125$ deg. The positive covariance for angles $\lesssim 40$ deg is a signature of large-scale systematics and is reasonably well approximated by the fitted exponential

$$V_\varpi(\theta) \simeq (285\ \mu\mathrm{as}^2) \times \exp(-\theta/14^\circ)\,, \tag{16}$$

shown by the dashed curve. This function corresponds to errors with an RMS amplitude of $285^{1/2} \simeq 17\ \mu$as and a characteristic spatial scale of 14 deg, both of which are consistent with the large-scale patterns seen in Fig. 12. The dip in $V_\varpi(\theta)$ around $\theta = 120$ deg may be related to the basic angle, although it is centred on a slightly higher value than $\Gamma = 106.5$ deg.

The lower panel of Fig. 14 shows $V_\varpi(\theta)$ for $\theta < 7$ deg. The blue curve connects the slightly smoothed values. Although Eq. (16), shown by the dashed curve, well describes the mean covariance averaged over a few degrees, the detailed curve shows multiple oscillations around the exponential with a period of about 1 deg, and for the smallest angles (< 0.125 deg), the covariance becomes much larger, about 1850 μas^2 (with a large statistical uncertainty), corresponding to an RMS amplitude of 43 μas. These features are clearly produced by small-scale patterns similar to what is seen in the LMC area (Fig. 13).

Qualitatively similar correlations on both large and small angular scales are found by analysing the proper motions of the

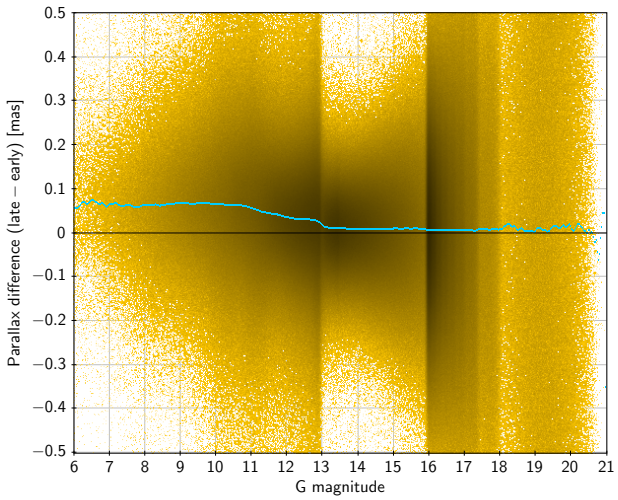

Fig. 16. Difference in parallax between the "late" and "early" solutions as a function of magnitude. The cyan curve is the median. Only results for primary sources are plotted; discontinuities in the density of points at $G = 13$, 16, etc. are caused by the way the primary sources are selected.

quasars. We define

$$V_\mu(\theta) = \frac{1}{2}\langle \boldsymbol{\mu}_i' \boldsymbol{\mu}_j \rangle\,, \tag{17}$$

where $\boldsymbol{\mu}_i = \boldsymbol{p}_i \mu_{\alpha*i} + \boldsymbol{q}_i \mu_{\delta i}$ is the proper motion vector of source i, with unit vectors $\boldsymbol{p}_i$ and $\boldsymbol{q}_i$ towards increasing α and δ, respectively (e.g. Eq. 3 in Lindegren et al. 2016). The prime denotes the scalar product. The vector formulation was chosen in order to combine the two components of proper motion in a frame-independent way. For small separations $\boldsymbol{p}_i \simeq \boldsymbol{p}_j$ and $\boldsymbol{q}_i \simeq \boldsymbol{q}_j$, which gives $V_\mu(\theta) \simeq (\mu_{\alpha*i}\mu_{\alpha*j} + \mu_{\delta i}\mu_{\delta j})/2$; thus V_μ is the covariance averaged between the two components of the proper motion. Figure 15 shows $V_\mu(\theta)$ for the high-precision quasar sample. The dashed curve is the fitted exponential

$$V_\mu(\theta) \simeq (800\ \mu\mathrm{as}^2\mathrm{yr}^{-2}) \times \exp(-\theta/20^\circ)\,. \tag{18}$$

The value at $\theta = 0$ corresponds to an RMS amplitude of about 28 μas yr^{-1} for the large-scale systematics. At small separations similar features are seen as for $V_\varpi(\theta)$, including the 1 deg oscillations; for $\theta < 0.125$ deg the covariance is 4400 μas^{2}yr^{-2}, corresponding to an RMS value of 66 μas yr^{-1} per component of the proper motions. Again, this is consistent with small-scale proper motion patterns seen, for example, in the LMC (Gaia Collaboration et al. 2018c).

The RMS values derived above and summarised in Table 4 for the different angular scales can be interpreted as the noise floor when averaging the parallaxes or proper motions for a large number of sources in areas of the corresponding sizes. The numbers should be seen as indicative and not necessarily as representative for sources that are much brighter than the quasars.

5.5. Split-field solutions

The internal consistency of the astrometric solution can be examined by comparing solutions based on complementary subsets of the observations. The observations can for example be divided depending on the CCD strip in the astrometric field (AF). Normally a source is observed in nine consecutive CCD strips, denoted AF1–AF9, as its image moves over the focal plane (see e.g. Fig. 3 in the AGIS paper), thus generating up to nine AL observations per field-of-view transit. The photon noise component

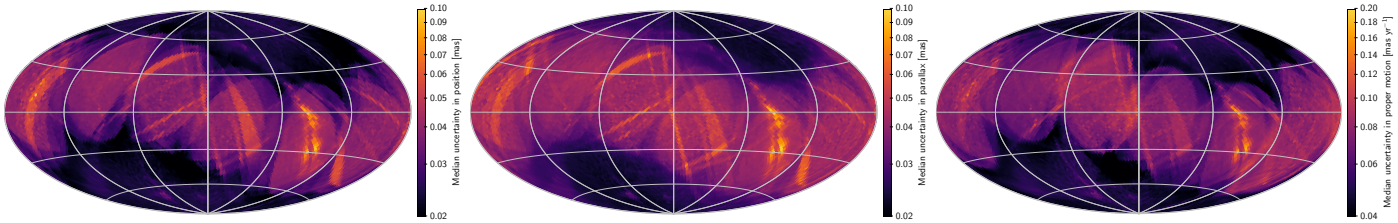

Fig. B.3. Formal uncertainties at $G \simeq 15$ for sources with a five-parameter astrometric solution. *Left:* Semi-major axis of the error ellipse in position at epoch J2015.5. *Middle:* Standard deviation in parallax. *Right:* Semi-major axis of the error ellipse in proper motion. This and all other full-sky maps in this paper use a Hammer–Aitoff projection in equatorial (ICRS) coordinates with $\alpha = \delta = 0$ at the centre, north up, and α increasing from right to left.

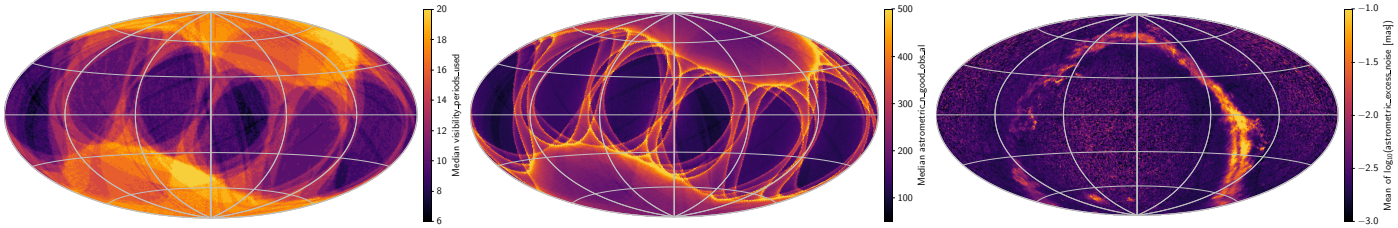

Fig. B.4. Observation statistics at $G \simeq 15$ for sources with a five-parameter astrometric solution. These statistics are main factors governing the formal uncertainties of the astrometric data. *Left:* Number of visibility periods used. *Middle:* Number of good CCD observations AL. A map of the number of used field-of-view transits is very similar, with a factor nine smaller numbers. *Right:* Mean excess source noise.

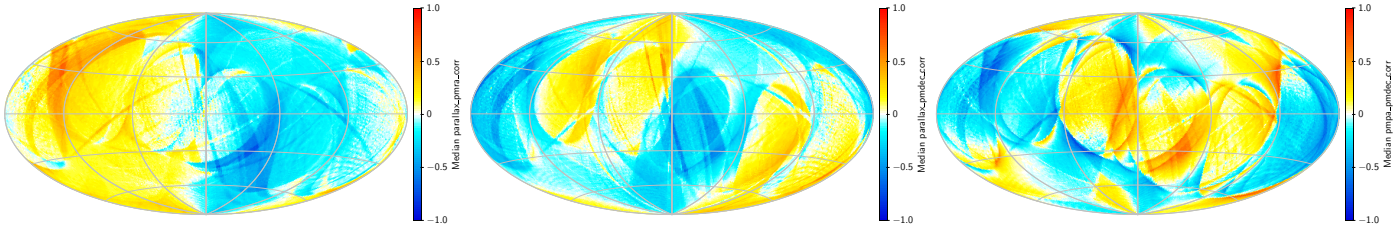

Fig. B.5. Correlation coefficients at $G \simeq 15$ for sources with a five-parameter astrometric solution. Maps of the correlations at other magnitudes are very similar to these. *Left:* Correlation between ϖ and $\mu_{\alpha*}$. *Middle:* Correlation between ϖ and μ_δ. *Right:* Correlation between $\mu_{\alpha*}$ and μ_δ.

brighter sources; and a general scatter of large u at all magnitudes, which could be caused by partially resolved or astrometric binaries. If we want to keep the sources with $G < 6$ (which include most of the giants) but remove the blob at $G > 18$, a possible cut is given by the black lines, i.e. the function

$$u < 1.2 \times \max(1, \exp(-0.2(G - 19.5))) \,. \quad \text{(C.1)}$$

Adding this criterion to Selection A gives Selection B with 249 793 sources and the much cleaner HR diagram in the middle panel of Fig. C.1. (A similar filtering could be obtained by using the excess source noise instead of u, for example by selecting `astrometric_excess_noise` < 1 mas, but the behaviour of the excess noise for $G \lesssim 15$ is less discriminating due to the DOF bug.) Selection B still contains two sources with $\varpi > 800$ mas.

Additional scatter in the HR diagram is produced by photometric errors mainly in the BP and RP bands, affecting in particular faint sources in crowded areas. An indicator of possible issues with the BP and RP photometry is the flux excess factor $E = (I_{\rm BP} + I_{\rm RP})/I_{\rm G}$ (`phot_bp_rp_excess_factor`), where I_X is the photometric flux in band X (Evans et al. 2018). Adding the criterion (Gaia Collaboration et al. 2018d)

$$1.0 + 0.015(G_{\rm BP} - G_{\rm RP})^2 < E < 1.3 + 0.06(G_{\rm BP} - G_{\rm RP})^2 \quad \text{(C.2)}$$

to Selection B gives Selection C with 242 582 sources and the HR diagram in the right panel of Fig. C.1. The remaining scatter of points between the main and white-dwarf sequences may be partly real, consisting of binaries with white-dwarf and main-sequence companions of roughly equal magnitude. In Selection C the source with the largest parallax is Proxima Centauri.

The chance matching mechanism discussed above, where different observations of the same *Gaia* source are matched to two (or more) physically distinct objects, should produce a roughly equal number of positive and negative spurious parallaxes. Further insight into the mechanism can therefore be gained by inspecting a sample of sources with significantly negative parallaxes. The selection

$$\left.\begin{array}{ll} \text{(i)} & \varpi < -10 \text{ mas} \\ \text{(ii)} & \varpi/\sigma_\varpi < -10 \end{array}\right\} \quad \text{(Selection N)}$$

gives 113 393 sources with manifestly unphysical parallaxes. A plot of u versus G for this sample is shown in the right panel of Fig. C.2. The similarity to the "blob" in the left plot is striking, and supports the idea that most of the spurious large (positive or negative) parallaxes can be removed by a judicious cut in the (G, u) plane. In fact 90% of the the sources in Selection N are removed by the cut in Eq. (C.1).

Selection N includes 61 sources with $\varpi < -800$ mas, the smallest being −1857 mas. For comparison, if the photometric criteria (iii) and (iv) are removed from Selection A, the number of sources with $\varpi > 800$ mas is 46. Conversely, if (iii) and (iv) are imposed on Selection N, the number of sources with $\varpi < -800$ mas is reduced to 6. The similar number of very large negative and positive parallaxes, when similar criteria are

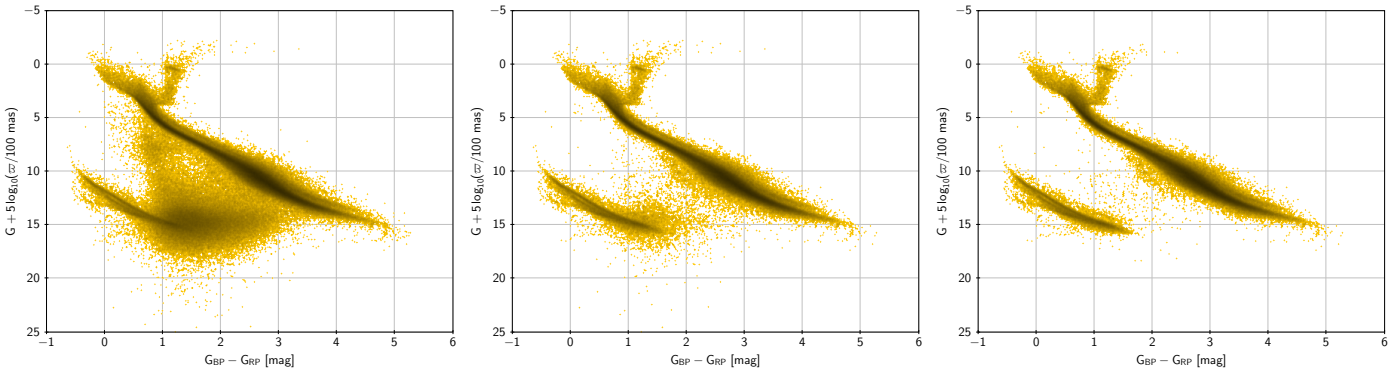

Fig. C.1. HR diagram of sources nominally within 100 pc and with relative distance error less than 10%. *Left:* Raw diagram (Selection A). *Middle:* Sources filtered by unit weight error (Selection B). *Right:* Sources filtered by unit weight error and flux excess ratio (Selection C).

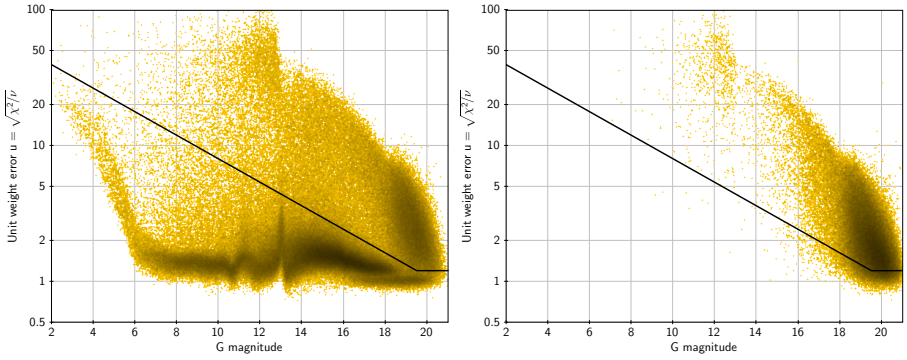

Fig. C.2. Unit weight error $u = (\chi^2/\nu)^{1/2}$ versus magnitude for two samples. *Left:* Selection A (positive parallaxes). *Right:* Selection N (negative parallaxes). The black line is the threshold defined in Eq. (C.1).

applied, broadly supports the hypothesis that most of the spurious large parallaxes result from the previously described chance matching of the observations to distinct objects. (The same thing can of course happen with the resolved components of a physical double star, if the separation is $\lesssim 1$ arcsec.) The probability that it happens should decrease steeply with an increased number of available observations, or rather with the number of visibility periods (Sect. 4.3). That this is indeed the case is illustrated in Fig. C.3, where the tail of normalised negative parallaxes is plotted for Selection N and for some subsets of it. Nominally, if the parallax errors were truly unbiased and Gaussian, we would expect to have no source at all with $-\varpi/\sigma_\varpi > 6$. The blue curve shows the distribution for the sources in Selection N, which by Eq. (11) all have at least six visibility periods. Requiring at least 7 or 10 visibility periods (green line/rings, and grey line/squares, respectively) drastically reduces the negative tail while retaining 85% and 41% of the sources. Requiring even more visibility periods only shrinks the sample without changing the shape of the tail. If these criteria are applied to Selection A, the HR diagram gets cleaner at the faint end, but most of the points between the main sequence and white dwarfs around colour index 1 are still present. Increasing the minimum number of visibility periods is therefore efficient for eliminating the most extreme spurious parallaxes, but not for cleaning the middle and upper part of the HR diagram. The red curve in Fig. C.3 shows the distribution of negative parallaxes after the cut in Eq. (C.1), which is clearly more effective in removing the many parallaxes that are only moderately wrong.

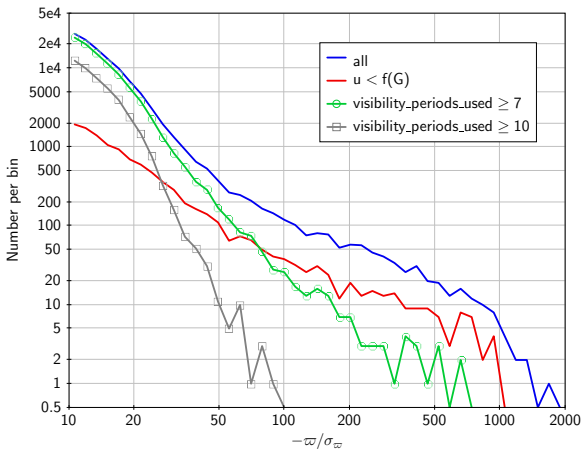

Fig. C.3. Distribution of the negative tail of normalised parallaxes.

The effectiveness of the filters described above is also illustrated in Fig. C.4. The left map shows the celestial distribution of the 73 246 sources in *Gaia* DR2 that are nominally within 50 pc from the Sun, i.e. with $\varpi > 20$ mas. Stars in this volume should have a rather uniform distribution on the sky; yet the map shows strong features correlated with the density of faint stars (e.g. along the Galactic equator) or related to the scanning law (e.g. the triangular patch in the left part of the map). Much of these features disappear after applying the cut in Eq. (C.1), as shown in the middle map. Applying in addition the cut in Eq. (C.2) leaves 34 001 sources with a nearly uniform distribution (right map). The remaining concentration of points at $(\alpha, \delta) \simeq (67°, +16°)$ is the Hyades cluster. It can be inferred that most of the remaining sources are real. Inevitably, however, the filtering eliminates also some real sources with valid solutions. In this example the 39 245 sources removed by Eqs. (C.1)–(C.2) include at least some 700 actual nearby stars, among them Sirius B, Kruger 60, Ross 614, η Cas, π^3 Ori, and δ Eri.

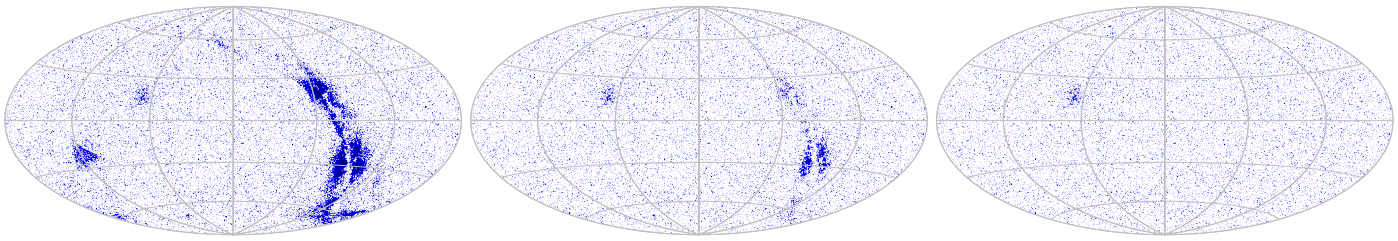

Fig. C.4. Distribution in equatorial (ICRS) coordinates of sources formally within 50 pc. *Left:* All 73 246 sources with $\varpi > 20$ mas. *Middle:* The subset of 39 478 sources satisfying Eq. (C.1). *Right:* The subset of 34 001 sources satisfying both Eqs. (C.1) and (C.2).