*Draft of 2018.08.12–17:46 -0400.* **CONFIDENTIAL**

MNRAS , 1–9 (2018) Preprint 12 August 2018 Compiled using MNRAS LATEX style file v3.0

# A quantitative analysis of systematic differences in positions and proper motions of Gaia DR2 with respect to VLBI

L. Petrov[1,2⋆], Y. Y. Kovalev[2,3,4] and A. V. Plavin[2,3]

[1] *Astrogeo Center, 7312 Sportsman Dr., Falls Church, VA 22043, USA*

[2] *Astro Space Center of Lebedev Physical Institute, Profsoyuznaya 84/32, 117997 Moscow, Russia*

[3] *Moscow Institute of Physics and Technology, Dolgoprudny, Institutsky per., 9, Moscow, Russia*

[4] *Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, 53121 Bonn, Germany*

Accepted xxx. Received yyy; in original form zzz

**ABSTRACT**

We have analyzed the differences in positions of 9081 matching sources between the Gaia DR2 and VLBI catalogues. The median position uncertainty of matching sources in the VLBI catalogue is a factor of two larger than the median position uncertainty in the Gaia DR2. There are 9% matching sources with statistically significant offsets between two catalogues. We found that after removal of the statistically significant outliers, the reported positional errors should be re-scaled by a factor of 1.3 for VLBI and 1.06 for Gaia, and in addition, Gaia errors should be multiplied by the square root of chi square per degree of freedom in order to best match the normalized position differences to the Rayleigh distribution. We have established that the major contributor to statistically significant position offsets is the presence of optical jets. Among the sources for which the jet direction was determined, the position offsets are parallel to the jet directions for a 62% of the outliers. Among the matching sources with significant proper motion, the fraction of objects with proper motion directions parallel to jets is a factor of 3 greater. Such sources have systematically higher chi square per degree of freedom. We explain these proper motions as a manifestation of the source position jitter that we have predicted earlier. Therefore, the assumption that quasars are fixed points and therefore, differential proper motions determined with respect to quasar photocenters can be regarded as absolute proper motions, should be treated with a great caution.

**Key words:** galaxies: active – galaxies: jets – quasars: general – radio continuum: galaxies – astrometry: reference systems

## 1 INTRODUCTION

Since 1980s very long baseline interferometry (VLBI) has been the most accurate absolute astrometry technique. The accuracy of VLBI absolute positions can reach the 0.1 mas level. With few exceptions, VLBI is able to provide absolute positions only for active galactic nuclei (AGNs). In 2016, the Gaia Data Release 1 (DR1) (Lindegren et al. 2016) ushered an emergence of the technique that rivals VLBI in accuracy. A quick analysis by Mignard et al. (2016) found that in general, the differences between common AGNs in VLBI and Gaia DR1 catalogues are close to their uncertainties, except for a 6% of common objects. Mignard et al. (2016) claims that "close examination a number of these cases shows that a likely explanation for the offset can often be found, for example in the form of a bright host galaxy or nearby star". They conclude (page 13) that "the overall agreement between the optical and radio positions is excellent". We see it differently. If two independent observing campaigns produced small (negligible) differences, that also implies that the contribution of a new campaign is also small (negligible) with respect to what has been known before. Science does not emerge from agreements. It emerges from disagreements. Therefore, we focused our analysis on differences between VLBI and Gaia AGN positions.

Our analysis of Gaia DR1 confirmed the existence of a population of sources with statistically significant VLBI/Gaia offsets (Petrov & Kovalev 2017a). We found that such factors as the failures in quality control in both VLBI and Gaia, blended nearby stars, or bright host galaxies can account at maximum 1/3 of that population. This analysis, as well as recent works of others (Mignard et al. 2016; Makarov et al. 2017; Frouard et al. 2018; Liu et al. 2018a,b,c), used arc lengths of VLBI/Gaia differences. Including the second dimension, the position angle of VLBI/Gaia offsets, resulted in a breakthrough. Though the distribution of the position angle counted from the declination axis turned out to be close to uniform, the distribution of the position angles of respect to the jet direction determined from analysis of VLBI images of matching sources re-

⋆ E-mail: Leonid.Petrov@lpetrov.net

vealed a strong anisotropy (Kovalev et al. 2017): they have have a preferable direction along the jet, and at a smaller extent in the direction opposite to the jet. We interpret it as a manifestation of a presence of optical jets at scales finer than the Gaia point spread function (PSF), i.e. 100–300 mas. Known optical jets in AGNs resolved with Hubble Space Telescope are cospatial (Gabuzda et al. 2006; Perlman et al. 2010; Meyer et al. 2018). Even in that case still there will be position differences. Since the response to an extended structure of a power detector used by Gaia and an interferometer that records voltage is fundamentally different, as it was shown in (Petrov & Kovalev 2017b), Gaia positions correspond to the location of the optical centroid, while the VLBI positions are associated to the most compact feature and bright feature of the jet base. Therefore, the physical meaning of the VLBI/Gaia offset is a displacement of the optical centroid with respect to the jet base.

In April 2018, the Gaia DR2 was published (Lindegren et al. 2018). It has 48% more sources than Gaia DR1 and a significantly higher accuracy. Mignard et al. (2018) reported that in general, the differences VLBI/Gaia DR2 are small with some exceptions. They set out five reasons for discrepancies (page 10): 1) real offsets between the centres of emission at optical and radio wavelengths; 2) error in matching of VLBI and Gaia objects; 3) an extended galaxy around the quasar; 4) double or lensed quasars; or 5) simply statistical outliers. The presence of optical jets was not put in the list as a likely explanation.

In Petrov & Kovalev (2017b) we examined the consequences of our interpretation of the VLBI/Gaia offsets due to the presence of optical jets. Among others, we made two predictions: 1) "further improvement in the position accuracy of VLBI and Gaia will not result in a reconciliation of radio and optical positions, but will result in improvement of the accuracy of determination of these position differences", 2) "we predict a jitter in the Gaia centroid position estimates for radio- loud AGNs" (pages 3785–3786). Since the Gaia DR2 accuracy is noticeably better than the Gaia DR1 accuracy, this motivated us to extend our previous analysis to the Gaia DR2 and check whether these predictions came true. To answer the question what is the most significant contributor to systematic position differences is the goal of this article.

## 2 COMPARISON OF VLBI/GAIA POSITIONS

We matched the Gaia DR2 catalogue of 1,692,919,135 objects against the Radio Fundamental Catalogue rfc_2018b (L. Petrov and Y.Y. Kovalev in preparation, 2018)[1] (RFC) of 15,155 sources. The RFC catalogue is derived using all VLBI observations under astrometric programs publicly available by August 01 2018. We used the same procedure of matching of Gaia objects against the VLBI catalogue, described in detail in Petrov & Kovalev (2017a) and got 9081 matches with the probability of false association below the $2 \cdot 10^{-4}$ level. The immediate comparison of formal uncertainties *among matches* showed that the Gaia uncertainties

[1] Available online at http://astrogeo.org/rfc

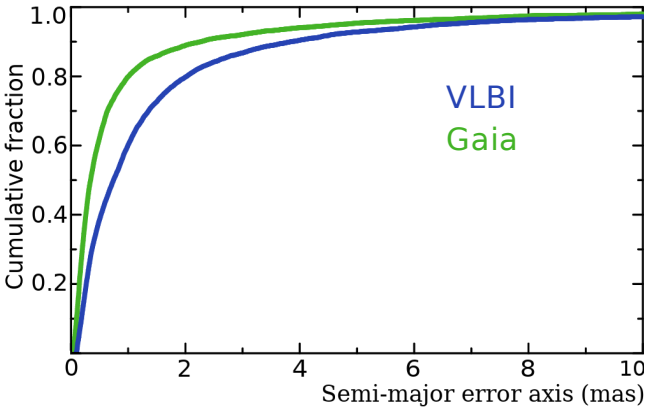


**Figure 1.** Cumulative distribution function of semi-major error axes $P(\sigma_{\rm maj} < a)$: green (upper) curve for Gaia and blue (low) curve for VLBI.

are smaller (see Figure 1). The median semi-major error ellipse of the VLBI sample is 0.74 mas against the 0.34 mas of the Gaia sample. Although VLBI can reach accuracies of 0.1 mas in absolute positions of strong sources, the majority of the sources were observed only once for 60 seconds, which is insufficient to derive their position with that level of accuracy. The Gaia uncertainties of matches are roughly twice smaller than the VLBI uncertainties, though there is no grounds for generalization of this statement to the entire Gaia or VLBI catalogues.

Among 9081 matches, 8112 have radio images at milliarcsecond resolution. Using these images, we have evaluated the jet directions for 4030 sources, i.e. for one half of the sample. We removed 48 sources that include 13 radio stars, 1 supernova remnant in the nearby star-forming galaxy, 10 gravitational lenses, and 24 double objects.

### 2.1 Analysis of VLBI/Gaia position angles with respect to the jet direction

We examined the arc lengths $a$ between the VLBI and Gaia source position estimates and the position angle of VLBI offset with respect to Gaia offset $\phi$ counted counter-clockwise with respect to the declination axis. Using reported position uncertainties and correlations between right ascensions and declinations, we computed the semi-major and semi-minor axes error ellipse, as well as position angles $\theta$ for both VLBI and Gaia position estimates. Then, assuming VLBI and Gaia errors are independent, we computed the uncertainties of the arc lengths $\sigma_a$ and position offsets $\sigma_\phi$ in the linear approximation:

$$\begin{aligned}
\sigma_a^2 &= \frac{1+\tan^2(\theta_v-\phi)}{1+\dfrac{\sigma^2_{v,\rm maj}}{\sigma^2_{v,\rm min}}}\,\sigma^2_{v,\rm maj} + \frac{1+\tan^2(\theta_g-\phi)}{1+\dfrac{\sigma^2_{g,\rm maj}}{\sigma^2_{g,\rm min}}}\,\sigma^2_{g,\rm maj} \\
\sigma_\phi^2 &= \Delta(\alpha_g-\alpha_v)^2(\sigma^2_{v,\delta}+\sigma^2_{g,\delta})\cos^2\delta_v/a^4 + \\
&\quad \Delta(\delta_g-\delta_v)^2(\sigma^2_{v,\alpha}+\sigma^2_{g,\alpha})\cos^2\delta_v/a^4 - \\
&\quad 2\Delta(\alpha_g-\alpha_v)\Delta(\delta_g-\delta_v)\cdot \\
&\quad (\mathrm{Corr}_v\sigma_{v,\alpha}\sigma_{v,\delta}+\mathrm{Corr}_g\sigma_{g,\alpha}\sigma_{g,\delta})\cos^2\delta_v/a^4,
\end{aligned} \tag{1}$$

where Corr is the correlation between right ascension and declination and the uncertainties in right ascensions are as-

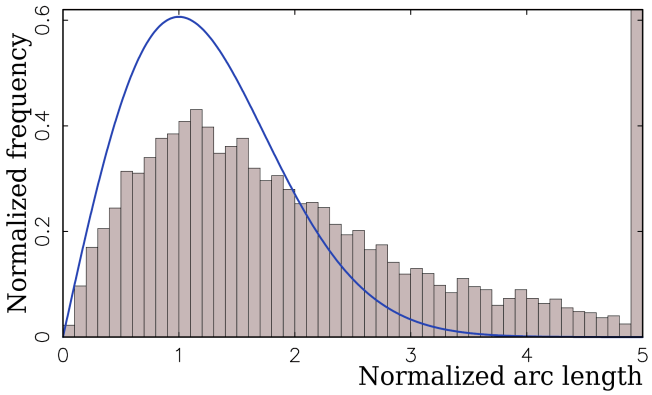


**Figure 2.** Distribution of the normalized VLBI/Gaia arc-lengths over 9033 matching sources. The last bin that holds normalized arc lengths $> 5$ exceeds the plot bounding box. The blue smooth curve shows Rayleigh distribution with $\sigma = 1$.

sumed reported without the $\cos\delta$ factor. Labels $v$ and $g$ stand for VLBI and Gaia respectively.

Figure 2 shows the distribution of the normalized arc-lengths $a/\sigma_a$ among all the matches. The last bin contains 1067 sources with normalized arcs greater than 5, or 11.4%. The number of sources with normalized arcs greater than 4, what for this study we consider statistically significant, is 16.3%, or 1/6. The goal of our study is to explain these outliers.

We computed the histograms of the distribution of the position angle offsets with respect to the jet directions determined from analysis of VLBI images of matches at milliarcsecond scales. We denote this quantity as $\psi$. Figure 3a shows such a histogram. Comparing this Figure with the upper left Figure 3 in Kovalev et al. (2017) shows that the anisotropy is revealed even more clearly: the peaks became sharper and narrower. The height of the peak with respect to the background is 2.8 versus 1.7 and the full width at half maximum (FWHM) is 0.42 rad versus 0.62 rad. We confirmed that anisotropy of $\psi$ angle is not an artifact of Gaia DR1, and follow the prediction made in Petrov & Kovalev (2017b) has come true.

We should note that the histogram of $\psi$ is affected by its measurement errors that depend on $a/\sigma_\phi$. We assume $\sigma_\psi = \sigma_\phi$ ignoring errors in the determination of jet direction angle that at the moment we cannot precisely characterize. At large $a/\sigma_\phi$ (say, more than 4) the distribution of the $\psi$ errors for a given measurement converges to the normal distribution. At low $a/\sigma_\phi$ (say less than 0.25), the distribution is converging to the uniform distribution. The analytic expression for the $\psi$ measurement errors can be found in page 233 of Thompson et al. (2017). Including measurements of $\psi$ with large errors smears the histogram. In order to mitigate smearing, we filtered out matches with $\sigma_\psi > 0.3$ rad. We found empirically, that reducing the threshold further results in the histograms to degrade as a consequence of the scarcity of remaining points, though not to change their shape noticeably.

Figure 3b shows the histogram of $\psi$ angles for all the matches with $\sigma_\psi < 0.3$ rad. The peaks at $0^\circ$ and $180^\circ$ became much stronger. A further analysis revealed that the histograms are different for short and long arc distances between VLBI and Gaia positions as shown in Figure 3c and 3d.

**Table 1.** Results of fitting the model in eq. 2 to the histograms in Figures 3a–d.

| Case | $\alpha$ | FWHM$_1$ rad | $\beta$ | FWHM$_2$ rad | $1-\alpha-2\beta$ | # src |
|---|---|---|---|---|---|---|
| a | 0.08 | 0.42 | 0.17 | 2.03 | 0.58 | 4017 |
| b | 0.23 | 0.40 | 0.22 | 1.48 | 0.33 | 985 |
| c | 0.07 | 0.35 | 0.17 | 1.01 | 0.47 | 423 |
| d | 0.24 | 0.40 | 0.17 | 1.84 | 0.28 | 565 |

To characterize histograms, we fitted a mathematical model to the histograms as follows:

$$f(\psi) = \alpha N(0,\sigma_1) \; + \; \beta N(0,\sigma_2) \; + \; \beta N(\pi,\sigma_2) + \frac{1-\alpha-2\beta}{2\pi}, \quad (2)$$

where $N(a,\sigma)$ is the normalized Gaussian function with first two moments $a$ and $\sigma$. We have selected a model that is as simple as possible. In the context of this study a choice of functions to represent the empirical distribution is irrelevant as far as the mathematical mode fits the distribution. Parameter $\alpha$ describes the contribution of the main narrow peak, parameter $\beta$ describes the contribution of the secondary wide peaks that has the maximum at both 0 and $\pi$, and the last term describes the contribution of the uniform component of the distribution. We noticed that the broad peaks at $\psi = 0$ and $\pi$ has a similar shape and fitting them separately with two additional parameters does not improve the fit. The results of fitting this 4-parametric model to the histograms in Figures 3a–d are shown in Table 1.

We see that the main peak at $\psi = 0$ with a FWHM around 0.4 rad value is rather insensitive to the way how a subsample is drawn. We tentatively conclude that the fitted FWHM is the intrinsic width of the peak. The peak at $\psi = 0$ is contributed predominately by the matches with large position offsets. This peak is related to the presence of optical jets. Several factors contributed to the peak broadening: a) the intrinsic width of the jet; b) errors in determination of the jet direction; c) curvature of the jet, which makes jet direction determination problematic. Perturbations in jet shape are magnified because of Doppler boosting. Typically, only a beginning of a jet is discernible at VLBI images due to limited dynamic range, while Gaia is sensitive to the jet up to scales comparable with the PSF.

Two secondary peaks are broad, with maxima at $\psi = 0$ and $\pi$. They are formed by matches almost exclusively with offsets shorter than 2–2.5 mas. The fraction of these secondary peaks in the distribution is relatively insensitive to the way how the subsample is drawn, 0.17–0.22, but its FWHM varies. We interpret it as an indication that a simplistic 4-parameter model is too coarse to fully describe the empirical distribution which shape depends on the VLBI/Gaia offset length.

The fifth column in Table 1 shows the fraction of the sources which offsets position angles have the uniform distribution, i.e. is not related to the core-jet morphology. This fraction is 0.58 for the histogram build using all the observations. The fraction is reduced to 0.33 for the subsample of observations with $\sigma_\psi < 0.3$ rad and to 0.25 for the subsample of observations with $\sigma_\psi < 0.2$ rad. This reduction occurs

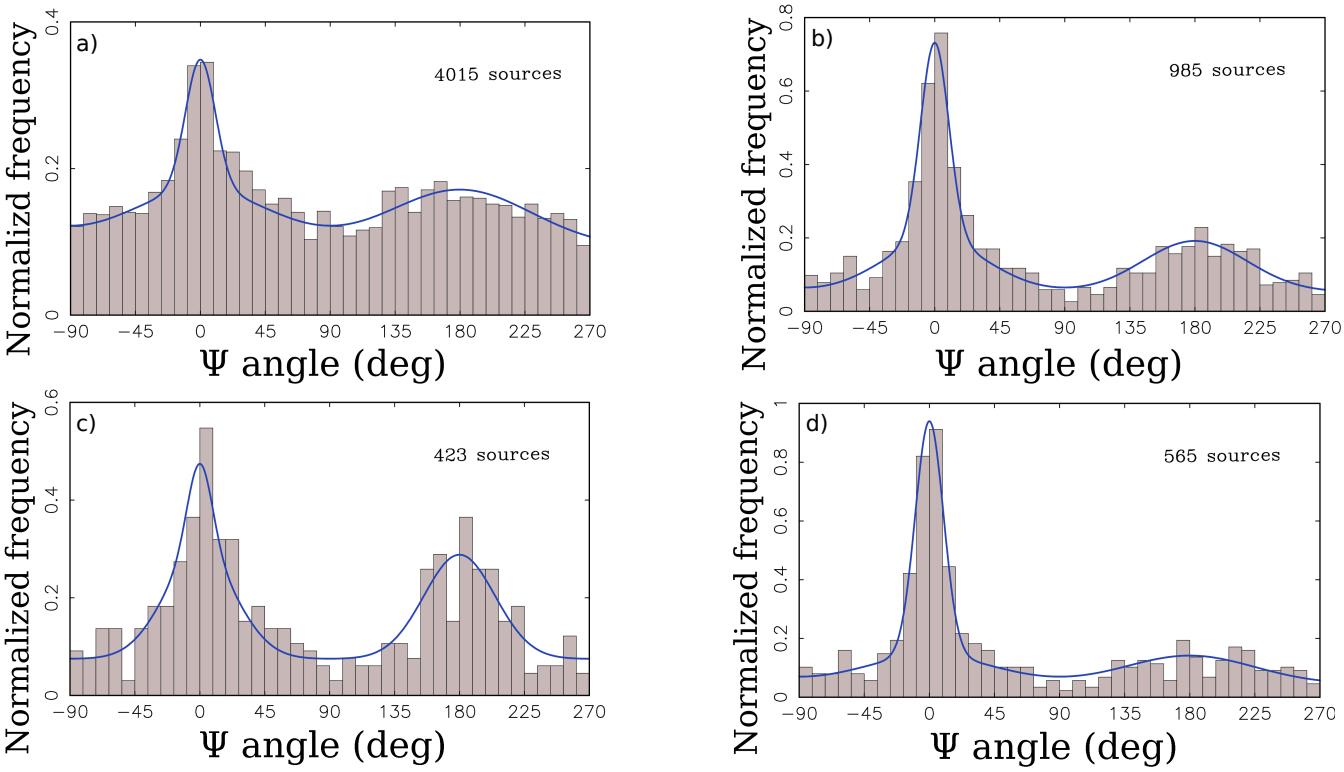


**Figure 3.** The histograms of the distribution of the position angle of Gaia offset with respect to VLBI position counted with respect to jet direction counter-clockwise. *Top left (a):* all the matches with known jet directions. *Top right (b):* the matches with $\sigma_\psi < 0.3$ rad. *Bottom left (c):* the matches with $\sigma_\psi < 0.3$ rad and arc-lengths $< 2.5$ mas. *Bottom right (d):* the matches with $\sigma_\psi < 0.3$ rad and arc-lengths $> 2.5$ mas. Blue curves are the best approximation of a three-component model.

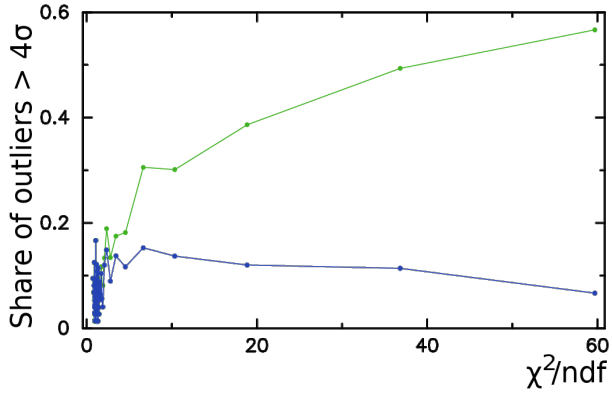


**Figure 4.** The fraction of outliers with normalized arc length of VLBI and Gaia matches $> 4$ for 1% percentiles of $\chi^2$/ndf. The horizontal axis is along the median value of $\chi^2$/ndf within each 1% percentile. The upper green curve was computed using original Gaia position uncertainties. The low blue curve was computed using Gaia uncertainties multiplied by $\sqrt{\chi^2/\mathrm{ndf}}$ factor.

partly due to the mitigation of the histogram smearing, and partly due to the selection bias. Since $\sigma_\psi$ depends on both uncertainties of position estimates and the arc-length, setting the upper limit for $\sigma_\psi$ disproportionately favours the matches with long VLBI/Gaia offsets that for a given position uncertainties have high chances to have low $\sigma_\psi$.

The distribution of the VLBI/Gaia position offset angles was studied by Plavin et al. (2018) for different purposes applying a different fraction analysis approach. The outcome qualitatively agrees with results presented here.

### 2.2 Re-scaling VLBI and Gaia reported position uncertainties

The presence of strong peaks at histograms in Figures 3 means these matches are affected by systematic errors. These systematic errors also affect the distribution of normalized arc lengths shown in Figure 2. In order to mitigate their impact, we re-drew the histogram and excluded the sources with offset position angles with respect to jet directions within 0.5 rad of peaks at 0 and $\pi$. As a result, we got a clean sample that is not affected by the systematic errors due to the presence of optical jet. We used this clean sample for characterizing Gaia and VLBI reported position uncertainties. We wanted to answer the question how realistic the uncertainties are.

We noticed that the number of outliers, i.e. the matches with the normalized arc $> 4$, grows with an increase of $\chi^2$/ndf, where ndf is the number of degrees of freedom. $\chi^2$ is provided in variable astrometric_chi2_al of the Gaia DR2 archive. The number of degrees of freedom was computed as the difference of the variables astrometric_n_good_obs_al and astrometric_params_solved. We split the dataset into 1% percentiles of $\chi^2$/ndf and computed the fraction of outliers for each percentile. The dependence of the fraction of outliers as a function of the mean $\chi^2$/ndf within a percentile is shown with a green curve in Figure 4. It grows approxi-

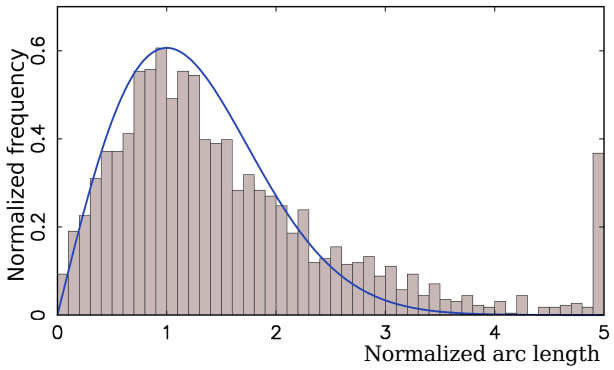


**Figure 5.** Distribution of normalized VLBI/Gaia arc-lengths over 2313 matching sources. The sample includes all the sources with known jet directions and excludes the sources with $\psi \in [-0.5, -0.5]$ and $\psi \in [\pi - 0.5, \pi + 0.5]$ rad. Scaling factors 1.05 and 1.30 were applied to Gaia and VLBI position uncertainties respectively. Gaia uncertainties were also multiplied by $\sqrt{\chi^2/\mathrm{ndf}}$. The blue smooth curve shows Rayleigh distribution with $\sigma = 1$.

mately as $\sqrt{\chi^2/\mathrm{ndf}}$ when $\chi^2/\mathrm{ndf}$ >1.5–2. Since the number of degrees of freedom is the mathematical expectation of $\chi^2$, in a case if all uncertainties of Gaia observables of a given source are underestimated by a common factor, multiplying the uncertainties in parameter estimates by $\sqrt{\chi^2/\mathrm{ndf}}$ corrects the impact of the underestimation of measurements errors. The blue curve in Figure 4 demonstrates that after re-scaling Gaia position uncertainties, the dependence of the number of outliers as a function of $\chi^2/\mathrm{ndf}$ has disappeared. Scaling position errors by $\chi^2/\mathrm{ndf}$ makes them larger, which makes the normalized arc-length smaller. We argue that re-scaling Gaia position errors makes them more realistic by accounting for the additional noise that increases $\chi^2/\mathrm{ndf}$.

In addition to source-dependent re-scaling that is based on $\chi^2/\mathrm{ndf}$ statistics of a given source, we evaluated global scaling factors for both VLBI and Gaia that affect every source. This is the simplest way to mitigate the impact of systematic errors on uncertainties and make them more realistic without re-running a solution. Since the normalized arc lengths are affected by both uncertainties of VLBI and Gaia positions, we estimated the scaling factors of VLBI uncertainties by processing the subset of observations with the Gaia position uncertainties a factor of 5 greater than the VLBI uncertainties and vice versus: we estimated scaling factors for the Gaia uncertainties (after scaling them by $\sqrt{\chi^2/\mathrm{ndf}}$) by processing the subset of observations with Gaia position uncertainties by a factor of 5 smaller than VLBI uncertainties. We adjusted the scaling factors in such a way that the distribution of normalized arc-lengths of the subsample be approximated with Rayleigh distribution $\sigma = 1$. The scaling factors are 1.06 for Gaia and 1.30 for VLBI. Applying scaling parameters to adjust uncertainties for accounting for the influence of systematic errors is a common technique. For instance, a scaling factor 1.5 was used to inflate source position uncertainties in the ICRF1 catalogue (Ma et al. 1998).

Since as we have established, the Gaia systematic errors caused by optical structure have a strong concentration towards $\psi = 0$ and $\psi = \pi$, we expect that removal of the matches with $\psi \in [-0.5, -0.5]$ and $\psi \in [\pi - 0.5, \pi + 0.5]$ rad and keeping only "off-peak" matches should affect the statistics of the number of outliers. We computed the fraction of matches with normalized residuals > 4 for for several sub-samples. Since we applied error re-scaling, the number of outliers has reduced with respect to our initial estimate mentioned above. The first row of Table 2 shows that excluding the sources within the peaks of the distribution of $\psi$ angle reduces the number of outliers by a factor of 1.36, but considering only the sources within 0.5 rad of the peaks doubles the number of outliers. Since the jet directions were determined only for 45% of the matches, these statistics underestimate the impact of the systematic errors caused by the presence of optical jets. If to count only the sources with known jet directions, excluding the sources within the peaks reduces the number of outliers by a factor of 2.07. Rows 2 and 4 of Table 2 shows also the statistics for the subsamples of low 50% percentile of VLBI re-scaled errors. The reduction of the number of outliers is 1.77 for the 50% percentile of the overall sample of matching sources and 2.65 for the sub-sample of the sources with known jet directions. The reduction of the number of outliers is greater for the lower 50% percentile because the sources with smaller position uncertainties have smaller errors in determining the $\psi$ angle, what makes discrimination of the "on-peak" and "off-peak" sources more reliable.

**Table 2.** Table with the fraction of matches with normalized residuals > 4 for a number sub-samples in pro cents (column r). The last two rows show the sub-samples of matches with known jet directions. The second and fourth raw use a sub-sample of matches with VLBI semi-major error ellipse less than median among all matches and the matches with known jet directions respectively. Column "off-peak" excludes the sources with $\psi \in [-0.5, -0.5]$ and $\psi \in [\pi - 0.5, \pi + 0.5]$ rad. Column "on-peak" include the sources with $\psi$ in these ranges and exclude everything else.

| | all | | off-peak | | on-peak | |
|---|---|---|---|---|---|---|
| | r | # src | r | # src | r | # src |
| all | 9.0 | 9033 | 6.6 | 7288 | 19.4 | 1702 |
| $\sigma_v \leq 0.963$ mas | 10.0 | 4496 | 5.9 | 3169 | 19.7 | 1323 |
| all with known $\psi$ | 11.2 | 4017 | 5.4 | 2313 | 22.1 | 1702 |
| $\sigma_v \leq 0.455$ mas | 11.4 | 1997 | 4.3 | 1109 | 20.3 | 888 |

Results in Table 2 show that the presence of optical structure parallel to the jet explains 62% VLBI/Gaia position offsets significant at the $4\sigma$ level for a sub-sample of 23% VLBI/Gaia matches that have known jet directions and VLBI position errors lower than the median. In order to generalize this result to the entire population of radio-loud AGN, we need assume that the significance of VLBI/Gaia does not depend on VLBI position error and does not depend on the measurability of the radio jet directions. The VLBI position errors above 0.2–0.3 mas level are limited by the thermal noise, and thus, the first assumption is valid. The validity of the second assumption is questionable. The detectability of parsec-scale radio jet depends on the jet brightness and the dynamic range of observations that in turn depends on the source flux density. Since the correlation between radio and optical fluxes is low, missing a jet just because the source was weak does not create a selection bias. However, if a jet direction for a given source was not detected because its radio jet is intrinsically weaker, missing

**Table 3.** Estimates of rotation angles around axes 1,2,3 of the Gaia positions of matches with respect to VLBI positions of four sub-samples. Units are milliarcseconds.

| | | | | |
|---|---|---|---|---|
| all | 9033 | $-0.030 \pm 0.004$ | $0.090 \pm 0.004$ | $-0.030 \pm 0.005$ |
| with jets | 4016 | $-0.010 \pm 0.005$ | $0.092 \pm 0.005$ | $-0.010 \pm 0.006$ |
| off-peak | 2647 | $-0.013 \pm 0.006$ | $0.095 \pm 0.006$ | $0.008 \pm 0.007$ |
| on-peak | 1369 | $-0.005 \pm 0.008$ | $0.091 \pm 0.007$ | $-0.037 \pm 0.009$ |

such a source may create a selection bias, because a weak radio jet may imply a weak optical jet. A sub-sample of sources with determined jet direction may have a selection bias towards jets brighter in radio and optic with respect to the overall population.

### 2.3 Impact of systematic errors on determination of the orientation of the Gaia catalogue with respect to the VLBI catalogue

Any source catalogue can be rotated at an arbitrary angle and the observables, f.e. group delays, remain the same. The orientation of a catalogue can be described by three angles. These angles for a given catalogue cannot be determined from observation and are *set* by imposing certain conditions. The orientation of the RFC catalogue is set to require the net rotation with respect to 212 so-called "defining" sources in the ICRF1 catalogue (Ma et al. 1998) be zero. Gaia DR2 catalogue was aligned with to respect to 2843 matching source the ICRF3-prototype catalogue using the frame rotator technique described in detail in Lindegren et al. (2012). The systematic differences caused by the optical structure affect the procedure for establishing the catalogue orientation. To provide a quantitative measure of sensitivity of the orientation angles to systematic errors, we computed the three angles of Gaia DR2 orientation with respect to the RFC VLBI catalogue (See Table 3). We see that selecting different samples, including those the most affected by systematic errors (on-peak) and least affected (off-peak), resulted in differences in orientation angles around 0.02 mas. A large value of the orientation angle around axis 2 is somewhat unexpected, but since the ICRF3-prototype catalogue used for alignment of the Gaia DR2 is not publicly available, the origin of this somewhat large value cannot be established.

## 3 ANALYSIS OF GAIA AND VLBI PROPER MOTIONS

The Gaia DR2 provides proper motions and parallaxes for 78% sources. Among 9081 matches, proper motion estimates are available for 7774 sources. Since the AGNs are located at cosmological distances, their proper motions considered as a bulk tangential motion are supposed to be well below the Gaia detection limit. A flare at the accretion disk or jet will change position of the centroid. A flare will cause a shift in position of the centroid, and therefore, results in a non-zero estimate of proper motion. Such a proper motion may be statistically significant even at Gaia level of accuracy. To check it, we made histograms of proper motions as a function of the position angles of the proper motion with respect to the jet directions denoted as $\bar{\psi}$. We analyzed the

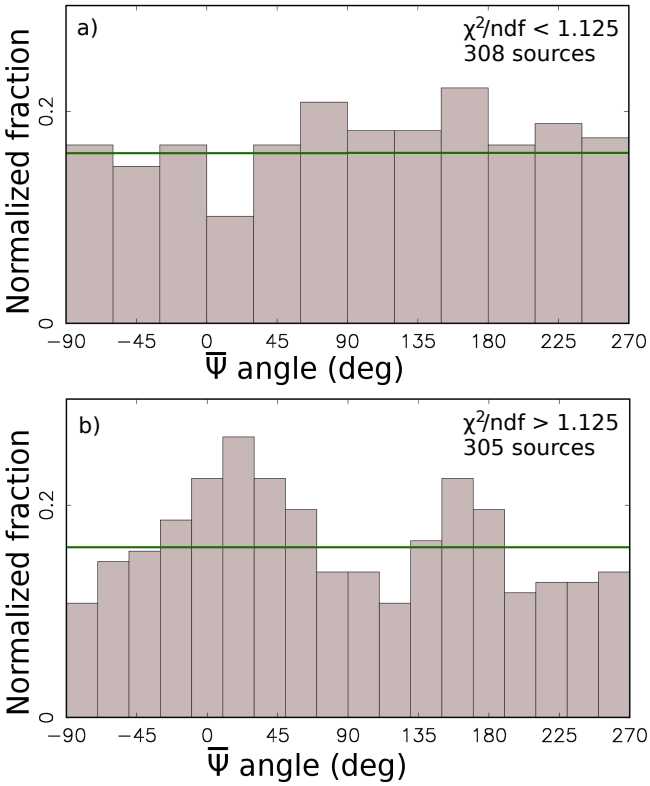


**Figure 6.** The histograms of the distribution of Gaia proper motion position angle with errors $< 0.4$ rad among the matches with different $\chi^2$/ndf. *Left figure* uses the matches with $\chi^2$/ndf less than the median in this subsample 1.125. *Right figure* uses the matches with $\chi^2$/ndf greater than the median in this subsample 1.125. For comparison, green line shows the uniform distribution.

sample of 613 matching sources with $\sigma(\bar{\psi}) < 0.4$ rad. The histograms showed weak peaks. The peaks become much sharper when we split the sample into two subsets: the subset with $\chi^2$/ndf less then the median 1.125 and the subset with $\chi^2$/ndf greater than the median (See Figure 6).

We see that the subsample of matches with large $\chi^2$/ndf shows two peaks at $\bar{\psi} = 0$ and $\bar{\psi} = \pi$ that are significant, while the subsample with $\chi^2$/ndf below the median does not. A non-linear motion is one of the reasons why $\chi^2$/ndf deviates from 1. The histogram at Figure 6-b tells us that among the sources with non-linear motion, the fraction of objects with proper motions along the jet or in the direction opposite the jet is disproportionately high. This dependence on angle $\bar{\psi}$ implies that at least for a fraction of the sources the proper motion is caused by the photocenter changes parallel to the jet direction.

We expect that most of optical flares happen close to the center of the AGN, either in the accretion disk or in the jet base. We can not directly see where the optical jet flares occur. However, the following arguments apply. The radio variability is associated with the apparent jet base — the core (e.g., Kovalev et al. 2005; Lister et al. 2016). Optical synchrotron emission is more transparent with even brighter core and steeper jet spectrum (e.g., Mimica et al. 2009). As a result, the jet base is expected to be the prime source of optical flares. The correlation between direction of linear

polarization between optical flares and radio core reported by Jorstad et al. (2007) confirms this.

Brightening a jet component shifts the centroid temporarily and irregularly. We call this behavior jitter and we predicted it in Petrov & Kovalev (2017b). Unlike to proper motions of stars, extending the interval of observations does not result in a convergence of a proper motion estimate to some value with small uncertainty. Instead, it slowly converges to zero. Peaks at 0 and $\pi$ in the histogram of the number of sources with $\sigma(\bar{\psi})$ as a function of $\bar{\psi}$ at a sub-samples with high $\chi^2$/ndf provides us the first evidence that predicted jitter indeed takes place. We used here estimates of AGN proper motions and $\chi^2$/ndf as a proxy for jitter detection.

We explored further the impact of a selection based on $\chi^2$/ndf on the distribution of position offset angles with respect to jet direction. We did not find a noticeable impact of $\chi^2$/ndf for VLBI/Gaia offsets longer 2.5 mas, but we found this selection affects the matches with VLBI/Gaia offsets shorter 2.5 mas. Figure 7 shows the distributions of $\psi$ angles of matches with $\sigma(\psi) < 0.3$ rad divided into three sub-samples approximately equally distributed over $\chi^2$/ndf. The peaks at $\psi = 0$ and $\psi = \pi$ are broad for the sub-sample of low $\chi^2$/ndf. They are getting sharper for the sub-sample of intermediate $\chi^2$/ndf. The sub-sample with large $\chi^2$/ndf is strikingly different than the sub-sample with low $\chi^2$/ndf: the histogram has a very strong peak at $\psi = 0$, i.e. along the jet direction and a smaller fraction of matches outside the main peaks.

Analysis of the connection of the Gaia DR2 proper motions with $\chi^2$/ndf suggests that the matches with large $\chi^2$/ndf are more prone to exhibit the jitter. This allows us to conclude tentatively that among the sub-sample of sources with VLBI/Gaia offsets shorter than 2.5 mas, flares and jitter occur predominantly in the objects that have Gaia offsets along the jet direction. This indicates that the mechanism that causes an increase of $\chi^2$/ndf may not work or at least is not dominating for sources with $\psi = \pi$. At the same time, Figures 6 and 8 suggest there is no strong preferable sign of the motion direction, either along or opposite to the jet. Such a pattern is consistent with a jitter caused by flares: depending when a flare has happened, at the beginning or the end of observing interval, the direction of the proper motion may be opposite.

It is instructive to examine whether proper motions in AGN positions derived from VLBI data analysis show the same pattern. We have run a special VLBI solution and estimated proper motions of 3039 sources using ionosphere-free linear combinations of group delays at 8.4 and 2.3 GHz. Source structure was considered as a $\delta$-function in processing VLBI observations at both fringe fitting and computation of theoretical group delay. We selected the sources that were observed in at least 2 sessions over an interval of at least 3 years and each observing session had at last 20 usable combinations of group delays. We applied the data reduction for the acceleration of the barycenter of the Solar system towards the Galactic center with right ascension 17h45m36.6s, declination $-28^\circ 56' 00''.0$, and magnitude $1.845 \cdot 10^{-10}$ m/s$^2$. We applied net-rotation constraints among 628 sources with strong history of observations, namely, observed in at least 8 sessions over 4 years or longer and have at least 128 usable linear combinations of group delays.

Figure 8 shows the histograms of the proper motion position angles $\bar{\psi}_g$ with respect to jet directions among those matching sources from Gaia and VLBI that have magnitude of the proper motions and position offsets significant at $3\sigma$ level for Gaia and $4\sigma$ for VLBI. There are 75 such sources in Gaia dataset and 284 in the VLBI dataset. The fraction of Gaia sources in bins at $\bar{\psi}_g = 0$ and $\bar{\psi}_g = \pi$ is a factor of 3 greater than on average. The median proper motions in these samples is 1.15 mas in the Gaia subset and 0.022 mas in the VLBI subset, i.e. a factor of 52 less. The Gaia proper motions were evaluated of 1.15 year time interval. The VLBI proper motions were evaluated over a time span in the range of 7.9 to 38.2 years with the median 26.5 years, a factor of 22.8 longer. The median magnitude of proper motions parallel to jet directions does not differ from the median magnitudes over the entire populations for both VLBI and Gaia.

While the histogram of the Gaia proper motion position angles shows peaks at both $\bar{\psi}_g = 0$ and $\bar{\psi}_g = \pi$, a similar histogram of the VLBI proper motion position angles shows only a peak at $\bar{\psi}_v = 0$. Explanation of this pattern in VLBI proper motions requires further investigation. As we showed in Petrov & Kovalev (2017b), unlike to a power detector, e.g. a CCD, an interferometer is not sensitive to the centroid change. Unaccounted contribution of an extended jet affects source position estimates at scales of tens microarcseconds. The unaccounted contribution of source structure to VLBI positions may reach a level of 0.1–1 mas if the image has more than one compact component, especially if the compact component is located at a distance comparable with a resolution of the interferometer. Change of relative brightness of distance between component due to flares causes in changes of position estimates at given epochs, and as result, causes proper motion. The peak at the low plot in Figure 8 confirms that at least for some of the sources this mechanisms works.

## 4 OTHER KNOWN CAUSES OF VLBI/GAIA POSITIONS OFFSETS

A number of authors (Mignard et al. 2016; Makarov et al. 2017; Frouard et al. 2018) suggested alternative explanations of statistically significant offsets:

- Error in matching VLBI and Gaia objects. They are easily controlled by computing the probability of false association based on source density in the vicinity of the candidates to association. The cutoff of the probability of false association $2 \cdot 10^{-4}$ results in the mathematical expectation of the total number of false association to 2. The coarseness of the model of source density may increase the number of false associations, but very unlikely can increase their count by an order of magnitude.
- An extended galaxy around the quasar. Position estimates of extended objects may suffer from deficiencies of the Gaia PSF model. To examine to which extent this affected VLBI/Gaia offsets, we examined a subsample of brightest galaxies from the NGC catalogue (Sinnnott & de Jager 1990). We used positions of these sources from Simbad database (Wenger et al. 2000), cross-matched them with against the RFC catalogue, and found 167 associations. Of them, 49, or 29%, had a counterpart in Gaia DR2. It is worth noting the fraction of VLBI/Gaia matches is twice less than

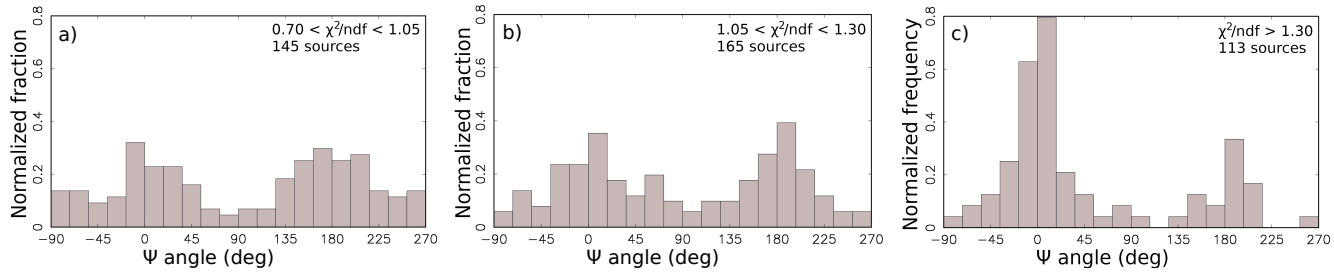


**Figure 7.** The histograms of the distribution of the position angle of Gaia offset with respect to VLBI position for matches with $\sigma_\psi < 0.3$ rad and arc-lengths $< 2.5$ mas and different ranges of $\chi^2$/ndf.

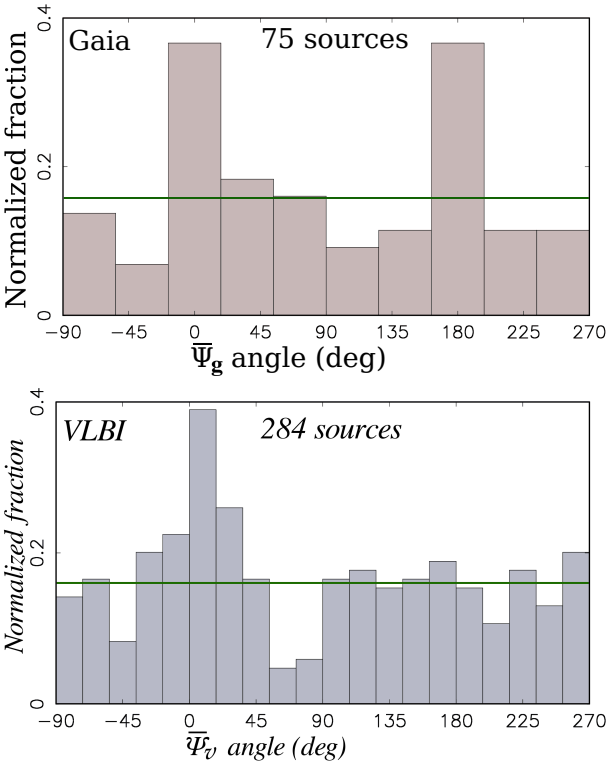


**Figure 8.** The histograms of significant AGNs proper motions among matching sources. *Up:* the proper motions from Gaia DR2 with magnitudes $> 3\sigma$ in both proper motions and position offsets. *Down:* the proper motions from the VLBI global solution. The horizontal green line shows the uniform distribution.

in the full sample. Without re-scaling the Gaia position uncertainties by the $\sqrt{\chi^2/\mathrm{ndf}}$ factor, approximately one half of these counterparts, 26 objects, have normalized arc-length exceeding 4. However, these objects had large $\chi^2$/ndf. After re-scaling the Gaia position uncertainties by $\chi^2$/ndf only one object (J1210+3924) has normalized arc-length above 4.

We conclude that extended galaxies may have large VLBI/Gaia offsets, but they also have large $\chi^2$/ndf. Scaling the uncertainties by the $\sqrt{\chi^2/\mathrm{ndf}}$ factor makes the normalized arc-lengths of galaxies indistinguishable from the rest of the sample.

• Lensed quasars. There are 10 known gravitational lenses in the sample of VLBI/Gaia matches. Since gravitational lenses were extensively hunted using radio surveys (e.g., Browne et al. 2003), it is unlikely that the RFC has even one missed gravitational lens.

• Double quasars. Makarov et al. (2017) presented a list of 28 sources with VLBI/Gaia DR1 significant offsets that have a close component on PanSTARRS images. Of them, 24 were found in Gaia DR2 and passed our test of the probability of false association $2 \cdot 10^{-4}$. Of them, 11 have significant VLBI/Gaia DR2 offsets. The second component may be either a star or a merging galaxy. During galaxy mergers, the nuclei may be dislodged with respect to the galaxy center of mass. A study of such systems may help to constraint theories of galaxy mergers. However, the number of such systems is small (11 out of 2293 identified in Makarov et al. (2017), i.e. 0.5%).

## 5 SUMMARY AND CONCLUSIONS

Here we summarize the main results of our comparison of AGN positions and proper motions the Gaia DR2 against the most complete catalogue of VLBI positions to date, the RFC.

(i) The Gaia DR2 AGN position uncertainties of VLBI matching sources are a factor of two smaller than the VLBI position uncertainties. The VLBI position catalogues and not the most precise any more.

(ii) We predicted in Petrov & Kovalev (2017b) that the improvement in accuracy of VLBI and/or Gaia will not reconcile the VLBI and Gaia positions, but will make these differences more significant. This prediction has come true. The fraction of outliers grew from 6 to 9%, the distribution of the position offset directions as a function of $\psi$ angle became sharper.

(iii) We demonstrated that the main reason for the statistically significant VLBI/Gaia position offset is the presence of optical structure. Among the matching sources with the normalized arc lengths exceeding 4 that have measured jet directions, 52–62%, i.e. *the majority*, have the position offsets are parallel to the jet direction. Therefore, we conclude that the optical jet is the cause. Although this fraction may be somewhat lower for the entire population of matching AGNs, we got its firm lower limit: 27%. Other reasons mentioned by Mignard et al. (2018) can explain only a small fraction of outliers.

The presence of emission from the hosting galaxy within the Gaia point spread function of an AGN may shift the centroid with respect to the nucleus if the galaxy central region structure is asymmetric or the AGN is dislodged with

respect to the galaxy center of mass, but such a shift is independent on $\psi$ angle. Table 1 provides the upper limit of the fraction of outliers which position offset do not depend on $\psi$, 33%. It does not seem likely that all of these offsets are caused by the contribution of hosting galaxies, because the fraction of AGNs with discernible hosting galaxies is much less.

We found that scaling the Gaia position uncertainties by $\sqrt{\chi^2/\mathrm{ndf}}$ eliminated the dependence of the fraction of the number of outliers on $\chi^2/\mathrm{ndf}$. Examining the subset of matches with dominating VLBI or Gaia errors allowed us to evaluate the scaling factors for the VLBI uncertainties, 1.30, and the Gaia position uncertainties: $1.06\sqrt{\chi^2/\mathrm{ndf}}$. Eliminating the observations within 0.5 rad of $\psi = 0$ and $\psi = \pi$ and using re-scaled uncertainties, made the distribution of normalized VLBI/Gaia arc-lengths much closer to the Rayleigh distribution: compare Figures 2 and 5.

(iv) The contribution of VLBI and/or Gaia systematic errors on estimates of the orientation angles of the Gaia DR2 catalogue with respect the VLBI catalogue does not exceed 0.02 mas.

(v) We predicted in Petrov & Kovalev (2017b) that flares in AGNs would cause a jitter in AGN positions. The analysis of Gaia proper motions provided us an indirect confirmation of this prediction: the sources with excessive Gaia residuals, i.e. large $\chi^2/\mathrm{ndf}$, have proper motion directions predominately parallel to the jet directions. The median magnitude of statistically significant proper motions is larger than 1 mas/yr over a 1.16 year interval, which is significantly higher than $< 0.05$ mas/yr over 5 years anticipated before the Gaia launch (Perryman et al. 2014). Although AGNs proper motions should not be interpreted as a bulk tangential motion, at the same time, these proper motions are not always an artifacts of Gaia data analysis, but some of them are real. The photo-centers of at least some quasars are not fixed points and the possibility of quasar proper motion should be taken into account in interpreting results of differential astrometry.

(vi) We found that VLBI proper motions have a preferable direction along with the jet. Median VLBI proper motions of AGNs are more than a factor of 50 smaller than Gaia proper motions.

We do not claim that we have solved the problem of establishing the nature of *all* outliers. The distribution in Figure 5 still deviates from Rayleigh and we still did not uncover the nature of the 1/3 outliers, but we made a quite substantial progress. We anticipate that a study of VLBI/Gaia position offsets will become a power tool for probing properties of the accretion disk and the relativistic jet in the AGNs, in line with the pioneering work of Plavin et al. (2018).

## ACKNOWLEDGMENTS

We used in our work the Astrogeo VLBI FITS image database[2] that contains radio images contributed by A. Bertarini, L. Garcia, N. Corey, Y. Cui, L. Gurvits, X. He, S. Lee, R. Lico, E. Liuzzo, A. Marscher, S. Jorstad, C. Marvin, D. Homan, Y. Kovalev, M. Lister, A. B. Pushkarev, E. Ros, T. Savolainen, A. Pushkarev, K. Sokolovski, G. Taylor, A. de Witt, M. Xu, B. Zhang, and the authors.

It is out pleasure to thank Eduardo Ros for suggestions that led to improvements of the manuscript.

This project is supported by the Russian Science Foundation grant 16-12-10481. This work has made use of data from the European Space Agency (ESA) mission Gaia[3], processed by the Gaia Data Processing and Analysis Consortium (DPAC[4]). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement. We used in our work VLBA data provided by the Long Baseline Observatory that is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. This research has made use of the SIMBAD database, operated at CDS, Strasbourg, France

## REFERENCES

Browne I. W. A., et al., 2003, MNRAS, 341, 13
Frouard J., Johnson M. C., Fey A., Makarov V. V., Dorland B. N., 2018, AJ, 155, 229
Gabuzda D. C., Rastorgueva E. A., Smith P. S., O'Sullivan S. P., 2006, MNRAS, 369, 1596
Jorstad S. G., et al., 2007, AJ, 134, 799
Kovalev Y. Y., et al., 2005, AJ, 130, 2473
Kovalev Y. Y., Petrov L., Plavin A. V., 2017, A&A, 598, L1
Lindegren L., Lammers U., Hobbs D., O'Mullane W., Bastian U., Hernández J., 2012, A&A, 538, A78
Lindegren L., et al., 2016, A&A, 595, A4
Lindegren L., et al., 2018, A&A, p. A14
Lister M. L., et al., 2016, AJ, 152, 12
Liu J.-C., Zhu Z., Liu N., 2018a, AJ, 156, 13
Liu J.-C., Malkin Z., Zhu Z., 2018b, MNRAS, 474, 4477
Liu N., Zhu Z., Liu J.-C., 2018c, A&A, 609, A19
Ma C., et al., 1998, AJ, 116, 516
Makarov V. V., Frouard J., Berghea C. T., Rest A., Chambers K. C., Kaiser N., Kudritzki R.-P., Magnier E. A., 2017, ApJ, 835, L30
Meyer E. T., Petropoulou M., Georganopoulos M., Chiaberge M., Breiding P., Sparks W. B., 2018, ApJ, 860, 9
Mignard F., et al., 2016, A&A, 595, A5
Mignard F., Klioner S., Lindegren L., Hernandez J., Bastian U., Bombrun A., 2018, A&A, 616, A14
Mimica P., Aloy M.-A., Agudo I., Martí J. M., Gómez J. L., Miralles J. A., 2009, ApJ, 696, 1142
Perlman E. S., et al., 2010, ApJ, 708, 171
Perryman M., Spergel D. N., Lindegren L., 2014, ApJ, 789, 166
Petrov L., Kovalev Y. Y., 2017a, MNRAS, 467, L71
Petrov L., Kovalev Y. Y., 2017b, MNRAS, 471, 3775
Plavin A., Kovalev Y. Y., Petrov L., 2018, preprint (arXiv:1899.99999)
Sinnnott R. W., de Jager C., 1990, Space Sci. Rev., 54, 190
Thompson A. R., Moran J. M., Swenson Jr. G. W., 2017, Interferometry and Synthesis in Radio Astronomy, 3rd Edition. Springer, doi:10.1007/978-3-319-44431-4
Wenger M., et al., 2000, A&AS, 143, 9

This paper has been typeset from a TEX/LATEX file prepared by the author.

[2] Available at http://astrogeo.org/vlbi_images

[3] https://www.cosmos.esa.int/gaia
[4] https://www.cosmos.esa.int/web/gaia/dpac/consortium