

Bla-bla-bla

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

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

1 INTRODUCTION

Since 80s very long baseline interferometry (VLBI) was the most accurate absolute astrometry technique. Accuracy of VLBI absolute positions can reach 0.1 mas level. With few exceptions, the objects VLBI is able to provide absolute positions are active galactic nuclei (AGNs). In 2016 the Gaia Data Release 1 (DR1) (Lindgren et al. 2016) ushered an appearance of the technique that rivals VLBI in accuracy. 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 6% 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 do not think that if two independent observing campaigns produced small (negligible) differences, such an outcome should be called excellent, because it 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 existence of a population of sources with a statistically significant VLBI/Gaia offsets (Petrov & Kovalev 2017a). We found that such factors as failures in quality control in both VLBI, Gaia blended nearby stars or bright host galaxies can account at maximum 1/3 of the population. This analysis, as well as work of others ref, used arc lengths of VLBI/Gaia differences. Including the second dimension, 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 with respect to the jet direction determined from analysis of VLBI images of matching sources revealed strong anisotropy

(Kovalev et al. 2017): Gaia offsets associated with the position of image centroid with respect to the VLBI position associated with the most compact feature of the jet base have a preferable direction along the jet and at a smaller extent along the opposite direction. 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. It should be noted that even if radio and optical jets are perfectly co-spatial, as ground observations of some AGNs with very large jets resolved by the HST suggest ref, there will be position differences since a response of a power detector used by Gaia and an interferometer that records voltage to an extended structure is fundamentally different (Petrov & Kovalev 2017b).

In April 2018, the Gaia DR2 was published (Lindgren et al. 2018). It has 48% more sources than Gaia DR1 and significantly higher accuracy than Gaia DR1. Mignard & Klioner (2018) found 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. Though the authors were aware of results in Kovalev et al. (2017), they did not mention the presence of optic jet as a likely explanation, tacitly assuming this factor so insignificant that it is not worth mentioning.

In Petrov & Kovalev (2017b) we examined consequences of our interpretation of VLBI/Gaia offsets due to 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 Gaia centroid position estimates for radio-loud AGNs” (pages 3785–3786). Since accuracy of Gaia DR2 is noticeably better than accuracy Gaia DR1, this motivated us to extend our previous analysis to Gaia DR2 and check whether these predictions came true. We predicted the impact of optical structure in VLBI/Gaia DR2 differences will

* E-mail: Leonid.Petrov@lpetrov.net

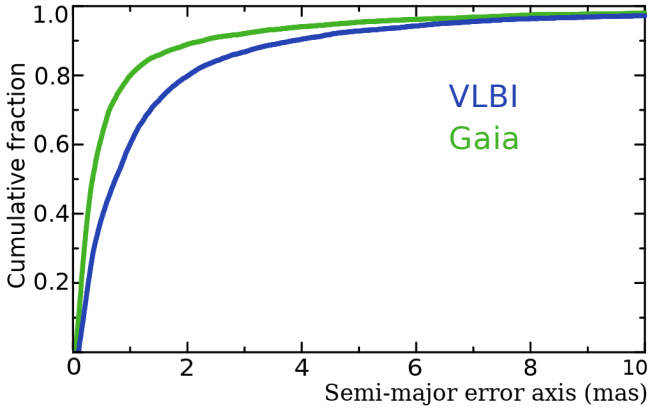


Figure 1. Cumulative distribution function of semi-major error axes $P(\sigma_{\text{maj}} < a)$: green (upper) curve for Gaia and blue (low) curve for VLBI.

be more significant than in VLBI/Gaia DR1, while [Mignard & Klioner \(2018\)](#) tacitly assume this factor is insignificant. Answering the question which interpretation is correct is the goal of this letter.

2 COMPARISON OF VLBI/GAIA POSITIONS

We matched Gaia DR2 catalogue of 1,692,919,135 objects against the Radio Fundamental Catalogue rfc_2018b (Petrov and Kovalev in preparation, 2018)¹ (RFC) of 15,077 sources. The RFC catalogue is derived using all VLBI observations under astrometric programs publicly available by July 15 2018. We used the same procedure of matching describe in more details in [Petrov & Kovalev \(2017a\)](#) and got 9030 matches with the probability of false association below $2 \cdot 10^{-4}$ level. Immediate comparison of formal uncertainties *among matches* showed that Gaia uncertainties 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 absolute position accuracy of strong sources 0.1 mas can be reached, the majority of the sources were observed only once for 60 seconds, which is insufficient to derive their position with that level of accuracy. It is fair to say Gaia uncertainties of matches are roughly 2 smaller than errors of VLBI matches, though there is no grounds for generalization of this statement to the entire Gaia or VLBI catalogues.

Among 9030 matches, 8080 have images. Using these images we have evaluated jet directions for 4030 sources, i.e. for one half. We removed 40 radio stars, supernova remnants in nearby star-forming galaxies, and double objects.

We examine arc length a between VLBI and Gaia source positions estimates and the position angle of VLBI offset with respect to Gaia offset ϕ counted counter-clockwise with respect to the declination axis. Using reported position uncertainties and correlations between right ascensions and declinations, we computed semi-major, and semi-minor axes error ellipse and position angle θ for both VLBI and Gaia position estimates. Then, assuming VLBI and Gaia errors

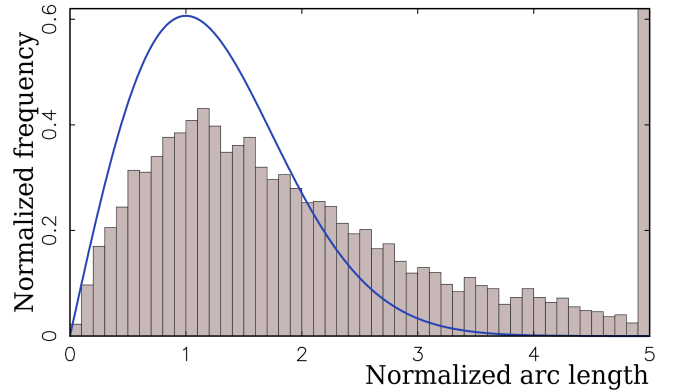


Figure 2. Distribution of normalized VLBI/Gaia arc-lengths over 8990 matching sources. The last bin that holds normalized arc lengths > 5 exceeds the plot bounding box. The blue smooth curve shows Rayleigh distribution with $\sigma = 1$.

are independent we computed uncertainties of arc lengths σ_a and position offsets σ_ϕ :

$$\sigma_a^2 = \frac{1 + \tan^2(\theta_v - \phi)}{1 + \frac{\sigma_{v,\text{maj}}^2}{\sigma_{v,\text{min}}^2}} \sigma_{v,\text{maj}}^2 + \frac{1 + \tan^2(\theta_g - \phi)}{1 + \frac{\sigma_{g,\text{maj}}^2}{\sigma_{g,\text{min}}^2}} \sigma_{g,\text{maj}}^2$$

$$\sigma_\phi^2 = \frac{\Delta(\alpha_g - \alpha_v)^2 (\sigma_{v,\delta}^2 + \sigma_{g,\delta}^2) \cos^2 \delta_v / a^4 + \Delta(\delta_g - \delta_v)^2 (\sigma_{v,\alpha}^2 + \sigma_{g,\alpha}^2) \cos^2 \delta_v / a^4 - 2\Delta(\alpha_g - \alpha_v)\Delta(\delta_g - \delta_v) \cdot (\text{Corr}_v \sigma_{v,\alpha} \sigma_{v,\delta} + \text{Corr}_g \sigma_{g,\alpha} \sigma_{g,\delta}) \cos^2 \delta_v / a^4}{\Delta(\alpha_g - \alpha_v)^2 (\sigma_{v,\delta}^2 + \sigma_{g,\delta}^2) \cos^2 \delta_v / a^4 + \Delta(\delta_g - \delta_v)^2 (\sigma_{v,\alpha}^2 + \sigma_{g,\alpha}^2) \cos^2 \delta_v / a^4 - 2\Delta(\alpha_g - \alpha_v)\Delta(\delta_g - \delta_v) \cdot (\text{Corr}_v \sigma_{v,\alpha} \sigma_{v,\delta} + \text{Corr}_g \sigma_{g,\alpha} \sigma_{g,\delta}) \cos^2 \delta_v / a^4}, \quad (1)$$

where Corr is correlation between right ascension and declination and uncertainties in right ascensions are assumed reported without $\cos \delta$ factor.

Figure 2 shows the distribution of the normalized arc-lengths a/σ_a among all the matches. The last bin contains 1029 sources with normalized arcs greater 5, or 11.4%. There number of sources with normalized arcs greater 4 that for this study we consider statistically significant, is 16.3%, or 1/6. The goal of our study is to explain the outliers.

We computed the histograms of the distribution of position angle offsets with respect to the jet direction determined from analysis of VLBI images of matches at milliarcsecond scales. We denote this quantity as ψ . Figure 3a shows such a histogram. Comparing with the upper left Figure 3 in [Kovalev et al. \(2017\)](#) shows 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. The anisotropy is not an artifact of Gaia DR1, but the real phenomena, and more precise observations measured it more precisely. The prediction made in [Petrov & Kovalev \(2017b\)](#) has come true.

Uncertainty of position angle depends on a/σ_ϕ . At large a/σ_ϕ (say, more than 4) the distribution of ψ measurement errors converges to the normal distribution. Ignoring errors in determination of jet direction angle that at the moment we cannot precisely characterize, we assume $\sigma_\psi = \sigma_p \text{hi}$. At low a/σ_ϕ (say less than 0.25) the distribution is converging to the uniform. The analytic expression for ψ measurement errors can be found in page 233 of [Thompson et al. \(2017\)](#). Measurements of ψ with large errors smear the distribution.

¹ Available online at <http://astrogeo.org/rfc>

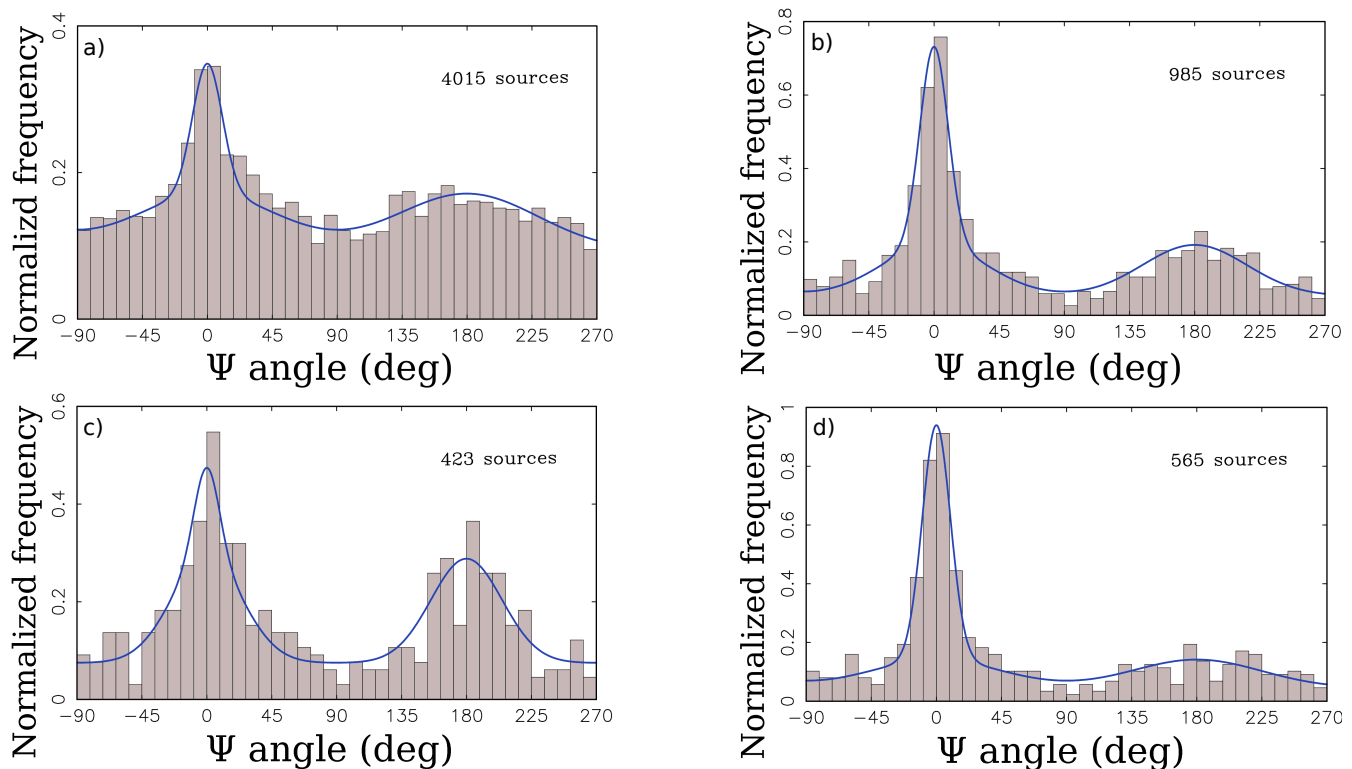


Figure 3. The histograms of the distribution of the position angle of Gaia offset with respect to VLBI position counted with respect to jet direction counter-clockwise. *Top left (a)*: all the matches with known jet directions. *Top right (b)*: the matches with $\sigma_\psi < 0.3$ rad. *Bottom left (c)*: 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.

In order to mitigate the effect of 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 due to scarcity of remaining points, though not to change their shape noticeably.

Figure 3b shows the histogram of all the matches with $\sigma_\psi < 0.3$ rad. The peaks at 0° and 180° became much stronger. Further analysis revealed that the histograms are different for short and long arc distance between VLBI and Gaia positions. Figure 3c and 3d show the histogram for $\sigma_\psi < 0.3$ rad and short and long arc lengths respectively.

To characterize histograms, we fitted it to a model. We have selected a model that is as simple as possible:

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

where $N(a, \sigma)$ is the normal distribution. Parameter α describes the contribution of the main narrow peak, parameter β describes the contribution of the secondary wide peaks that has the maximum at both 0 and π , and the last term describes the contribution of the uniform component of the distribution. Results of fitting this 4-parametric model to the histograms in Figures 3a–d is shown in Table 1

We see that the main peak at $\psi = 0$ and FWHM around 0.4 rad that is rather insensitive to the way how the subsample is drawn. We tentatively conclude that this is the intrinsic width of the peak. The peak at $\psi = 0$ is formed by predominately matches with large position offsets. We sur-

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

Case	α	FWHM ₁ rad	β	FWHM ₂ rad	$1 - \alpha - 2\beta$	# src
a	0.08	0.42	0.17	2.03	0.58	4015
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

mise that this peak is related to optic jet and its width is determined by jet collimation.

Two secondary peaks is broad, has maximum at $\psi = 0$ and π , and is formed by matches exclusively with offsets shorter than 2–2.5 mas. This can be explained by more turbulent environment or with clouds of obscuring matter at distances within 2.5 mas of the jet base. The share 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 not adequate to describe the empirical distribution.

The fifth column in Table 1 shows the share of the the component with the uniform distribution, i.e. the component that is not related to the core-jet morphology. This share is 0.58 for the histogram build using all the observations. The share is reduced to 0.33 for the subsample of observations with $\sigma_\psi < 0.3$ rad and to 0.25 for the subsam-

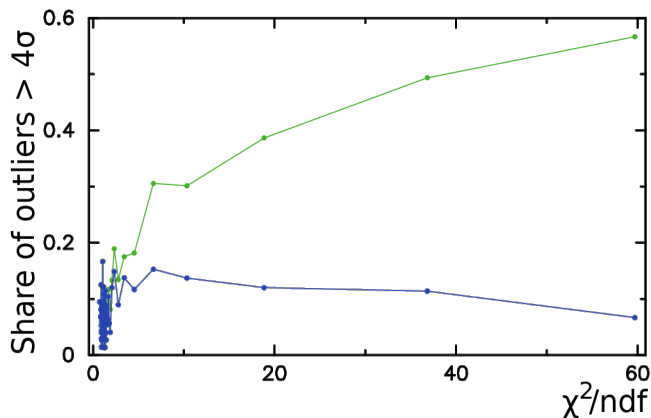


Figure 4. The share of outliers with normalized arc length of VLBI and Gaia matches > 4 for 1% percentiles of χ^2/ndf . The horizontal axis along the median value of χ^2/ndf within each 1% percentile. The upper green curve was computed using original Gaia position uncertainties. The low blue curve was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

ple of observations with $\sigma_\psi < 0.2$ rad. This reduction occurs partly due to mitigation of the histogram smearing, partly due to the selection bias. Since σ_ψ depends both on uncertainties of position estimates and the arc-length, setting the upper limit for σ_ψ disproportionately favours the matches with long VLBI/Gaia offsets.

The presence of strong peaks at histograms in Figures 3 means these matches are affected by systematic errors. This is one of the factors that affects the distribution of normalized arc lengths shown in Figure 2. In order to mitigate this factor, we re-drew the histogram and eliminated the sources with offset position angles with respect to jet direction within 0.5 rad of peaks at 0 and π . The next step is to adjust scaling parameters of position uncertainties. We noticed the number of outliers, i.e. the matches with the normalized arc > 4 , grows with the growth of χ^2/ndf , where ndf is the number of degrees of freedom. χ^2 is provided in variable `astrometric_chi2_al` of the Gaia DR2 archive and the number of degrees of freedom was computed as the difference of the variables `astrometric_n_good_obs_al` - `astrometric_params_solved`. We split the dataset into 1% percentiles of χ^2/ndf and computed the share of outliers for each percentile. The dependence of the share of outliers as a function of the mean χ^2/ndf within a percentile is shown with a green curve in Figure 4. It grows approximately as $\sqrt{\chi^2/\text{ndf}}$ when $\chi^2/\text{ndf} > 1.5-2$. Since the number of degrees of freedom is the mathematical expectation of χ^2 , in a case if all uncertainties of Gaia observables of a given source are underestimated by a common factor and the absence of other systematic errors, multiplying the uncertainties in parameters estimation by $\sqrt{\chi^2/\text{ndf}}$ corrects the impact of underestimation of measurements errors. The blue curve in Figure 4 demonstrates that after re-scaling Gaia position uncertainties, dependence of the number of outliers as a function of χ^2/ndf has disappeared.

In addition to source-dependent re-scaling that is based on χ^2/ndf statistics of a given source, we evaluated global scaling factors for both VLBI and Gaia that effect every source for accounting both systematic errors and biases in reported uncertainties. This is the simplest way to mitigate

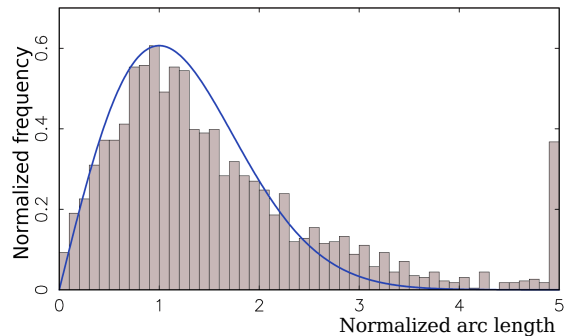


Figure 5. Distribution of normalized VLBI/Gaia arc-lengths over 2313 matching sources. The sample includes all the sources with known jet directions and excludes the sources with $\psi \in [-0.5, -0.5]$ and $\psi \in [\pi - 0.5, \pi + 0.5]$ rad. Scaling factors 1.05 and 1.30 were applied to Gaia and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

the make uncertainties more realistic that does not require 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 that have Gaia position uncertainties by a factor of 5 greater than VLBI uncertainties and vice versa estimated scaling factors for Gaia uncertainties (after scaling them by $\sqrt{\chi^2/\text{ndf}}$) by processing the subset of observations with Gaia position uncertainties by a factor of 5 smaller than VLBI uncertainties. We adjusted scaling factors in such a way the distribution of the 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 the influence of systematic errors is a common technique. For instance, f.e. 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 optic structure have a strong concentration towards $\psi = 0$ and $\psi = \pi$, we expect that removal 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 share of matches with normalized residuals > 4 for several sub-samples. The first row of Table 2 shows that excluding the sources within the peaks of the distribution of ψ 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 structure. Counting 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 the statistics for the the low 50% percentile of VLBI errors within the sub-samples after errors re-scaling. The reduction of the number of outliers is 1.77 for the 1/2 of

Table 2. Table with the share 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 row 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 ψ in these ranges and exclude everything else.

	all	off-peak	on-peak
	r # src	r # src	r # src
all	9.0 8990	6.6 7288	19.4 1702
$\sigma_v \leq 0.963$ mas	10.0 4496	5.9 3169	19.7 1323
all with known ψ	11.2 4015	5.4 2313	22.1 1702
$\sigma_v \leq 0.455$ mas	11.4 1997	4.3 1109	20.3 888

the overall sample of matching sources and 2.65 for the sub-sample of the sources with known jet directions. We used a subsample for the the low 50% percentile of VLBI errors to filter out the sources with large errors that prevented us to determine reliably ψ angle and therefore discriminate the “on-peak” and “off-peak” sources.

Results in Table 2 show that the presence of optic structure aligned along the jet explains 62% VLBI/Gaia position offsets significant at 4σ level for a sub-sample of 23% VLBI/Gaia matches with known jet direction and VLBI position error 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 measurability of the radio jet direction. 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. 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 optic fluxes is low, missing jet just because the source was weak does not create a selection bias. However, if a source direction for a given source was not detected because its radio jet is intrinsically weaker, missing such a source may create a selection bias, because weak radio jet may imply a weak optic 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.

3 ANALYSIS OF GAIA PROPER MOTIONS

4 SUMMARY AND CONCLUSIONS

Other topics

We have established the main reason of statistically significant VLBI/Gaia position offset is the presence of optical structure. Among the sources with measured jet direction, 52–62% matching sources with the normalized arc length exceeding 4, i.e. *the majority*, is due to the presence of optical structure. Although this share for the entire population of matching AGNs may be somewhat lower, we got the firm lower limit: 27%. Other reasons mentioned by Mignard

& Klioner (2018) can explain only a tiny fraction of outliers. Error in matching of VLBI and Gaia objects are easily controlled by computing the probability of false association based on source density in the vicinity of 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 make increase the number of false associations, but very unlikely can increase their count by an order of magnitude. Double and/or lensed AGNs are easily identified in radio images and can be thoroughly accounted. The presence of extended galaxies around the AGN shifts the optic centroid, but this shift is independent on ψ angle. Table 1 provides the upper limit of the share of matches which centroid might be shifted by the hosting galaxy: 33%.

We found that scaling Gaia position uncertainties by $\sqrt{\chi^2}/\text{ndf}$ eliminated the dependence of the share of the number of outliers on χ^2/ndf . Examining the subset of matches with dominating VLBI or Gaia errors allowed us to evaluate scaling factors for VLBI uncertainties, 1.30, and Gaia position uncertainties: $1.06 \sqrt{\chi^2}/\text{ndf}$. 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 close to the Rayleigh distribution: compare Figures 2 and 5.

We do not claim that we have solved the problem of establishing the nature of outliers, we distribution in Figure 5 still deviates from Rayleigh and we still did not uncover the nature of 1/3 outliers, but we made a quite substantial progress.

Other topics

REFERENCES

Kovalev Y. Y., Petrov L., Plavin A. V., 2017, *A&A*, **598**, L1
 Lindegren L., et al., 2016, *A&A*, **595**, A4
 Lindegren L., et al., 2018, preprint, ([arXiv:1804.09366](https://arxiv.org/abs/1804.09366))
 Ma C., et al., 1998, *AJ*, **116**, 516
 Mignard F., Klioner S., 2018, preprint, ([arXiv:1804.09377](https://arxiv.org/abs/1804.09377))
 Mignard F., et al., 2016, *A&A*, **595**, A5
 Petrov L., Kovalev Y. Y., 2017a, *MNRAS*, **467**, L71
 Petrov L., Kovalev Y. Y., 2017b, *MNRAS*, **471**, 3775
 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](https://doi.org/10.1007/978-3-319-44431-4)

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.

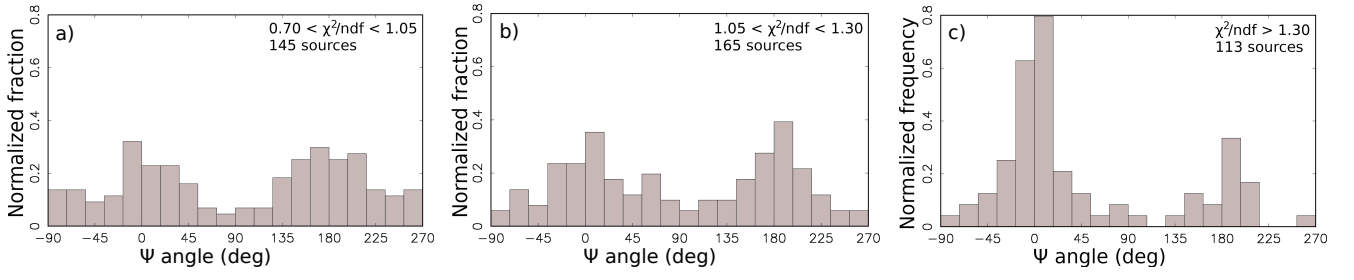


Figure 6. The histograms of the distribution of the position angle of Gaia offset with respect to VLBI position for matches with $\sigma_\psi < 0.3$ rad and arc-lengths < 2.5 mas and different ranges of χ^2/ndf .