A&A, 711, A299 (2026)
https://doi.org/10.1051/0004-6361/202659515

Astronomy & Astrophysics

Follow the wobble: Statistical methods to detect astrometric binary asteroids in *Gaia* FPR

L. Liberato[1,⋆], P. Tanga[1], D. Mary[1], R. Lallemand[2], Z. Liu[2], B. Carry[1], J. Desmars[3,2], D. Hestroffer[2], K. Minker[4], and A. Siakas[5]

[1] Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Bd de l'Observatoire, CS 34229, 06304 Nice Cedex 4, France
[2] Laboratoire Temps Espace (LTE), Observatoire de Paris, Université PSL, Sorbonne Université, 77 avenue Denfert Rochereau 75014 Paris, France
[3] Polytechnic Institute of Advanced Sciences-IPSA, 63 Boulevard de Brandebourg, 94200 Ivry-sur-Seine, France
[4] Lowell Observatory, 1400 Mars Hill Rd. Flagstaff, Arizona 86001, USA
[5] Department of Physics, Aristotle University of Thessaloniki, University Campus, Thessaloniki, 54124, Greece

Received 19 February 2026 / Accepted 21 May 2026

ABSTRACT

Context. In a previous study, we leveraged the astrometric accuracy of *Gaia* DR3 to obtain the first list of astrometric binary asteroid candidates. Some of these candidates have now been confirmed. However, that work did not provide the details of the statistical methods.

Aims. Our first aim is to provide methodological details and a performance evaluation of the approach used for detecting binaries. Our second aim is to establish an updated list of binary asteroid candidates from *Gaia* FPR astrometric residual exploration, accounting for the statistical properties of the *Gaia* FPR data.

Methods. We accounted for the astrometric uncertainties from *Gaia* FPR and refined the statistical model of the data. We used this model in Monte Carlo simulations to evaluate the strength of the individual detections. We implemented a trend-detection method in the residuals and applied a dedicated period-search algorithm. In addition, we updated the statistical selection process to build the list of candidates. We introduced a method for detecting objects in multiple windows of consecutive observation and refined the estimation of confidence intervals for these parameters, thereby better constraining the physical parameter selection.

Results. We detect 343 binary asteroid candidates, corresponding to 410 windows of consecutive observations, in the astrometric data. We show that in noise-only control simulations, the typical number of detections is 88% lower than in the *Gaia* FPR data. We also detect nine known binaries, 25 candidates that overlap with the Pan-STARRS survey, and 99 candidates that overlap with our previous binary search in *Gaia* DR3. Finally, we report 45 objects exhibiting residual trends suggestive of wide binary systems.

Conclusions. Our results and analyses demonstrate that although detecting binary asteroids is a difficult problem due to their low signal levels, the proposed method provides a reliable list of detections, including systems that are poorly accessible to conventional techniques. This set of targets is valuable for future confirmation with stellar occultations, light curves, and forthcoming LSST data.

Key words. methods: numerical – methods: statistical – catalogs – astrometry – minor planets, asteroids: general

1. Introduction

Asteroids constitute a rich source of information. They provide insights into the material composition and distribution in the protoplanetary disc (Carry 2012; DeMeo et al. 2015), the processes of planetary formation (Kleine et al. 2002; Izidoro et al. 2015), and the evolution of the Solar System to its current state (DeMeo & Carry 2014; Morbidelli et al. 2015). Binary asteroids provide the easiest way to gather this information because they act as small-scale laboratories of planetary formation and may carry footprints from the primordial Solar System. However, detecting binary asteroids is challenging, as reflected by the relatively small number of known binary systems, whereas they are expected to represent about 20% of the asteroid population (Pravec & Harris 2007; Margot et al. 2015).

In a previous article (Liberato et al. (2024), hereafter L24), we presented the results of a new method to discover binary asteroids in the Solar System by detecting their astrometric wobble, that is, the periodic variations in the position of a minor body induced by the gravitational perturbation from a companion. We analysed astrometric data from *Gaia* Data Release 3 (DR3) for more than 150 000 asteroids over 34 months of operation. This analysis yielded a list of more than 350 binary asteroid candidates. Since the publication of our initial astrometric binary exploration, several objects in our list had their binary nature confirmed, such as (3220) Murayama (Benishek et al. 2025; Sato 2013), (720) Bohlinia (Gorshanov et al. 2025), (1879) Broederstroom (Benishek et al. 2024), (1967) Menzel (Monteiro et al. 2024).These confirmations demonstrate that our method is effective in detecting astrometric signals from binary asteroids. Additionally, stellar occultations results for other objects show ambiguous evidence of their binarity, suggesting the possibility of contact binary systems (Lallemand et al. 2026).

In the present paper, we provide methodological details for binary asteroid detection and apply it to the *Gaia* Focused Product Release (FPR) (David et al. 2023) astrometric data, which contains the same ∼152 000 objects as in *Gaia* DR3 but with observations spanning 66 months.

⋆ Corresponding author: luana.liberato@oca.eu

The approach used in L24 can be summarised as follows. Our data sample consisted of transit–averaged residuals from the orbital fit of the astrometric data from *Gaia* DR3, projected in the along-scan (AL) direction of *Gaia*. The uncertainties affecting these residuals were taken as the standard deviation of the residuals per transit, also projected in the AL direction. To avoid changes in the geometry of observations that could drastically affect the phase and amplitude of the astrometric wobble detection, we performed the period search in windows of observations (WO) that contained at least ten transits over a maximum span of ten days[1].

We used the generalised Lomb-Scargle periodogram (GLSP, VanderPlas 2018; Lomb 1976; Scargle 1982) to perform a period search in each WO, considering the largest peak of the periodogram as the potential signal. We estimated a significance for the peak value by computing its p-value, which indicates how likely it is to obtain a similar peak value in a noise-only situation. We also estimated confidence intervals for the signal period and amplitude and computed a so-called quality factor Q, which was also used to measure the detection strength. We then selected candidates with p-values less than 5% and Q factors greater than 50% . The final step was to evaluate the coherence of the estimated parameters with the binary asteroid model adopted by estimating the minimum density of the objects and the minimum separation between the components of the binary candidates.

When updating this method for *Gaia* FPR data, we find that the approach above could be improved and optimised in several ways. That is, the use of *Gaia* FPR data at the CCD level, in contrast to *Gaia* DR3, offers the opportunity to better account for the error bars when computing data at the transit level. The reliability of these errors can be evaluated and propagated to estimate the uncertainties at the transit level. We improved the method for estimating confidence intervals (CIs), resulting in reduced false coverage. Finally, we changed the criterion for selecting the most interesting candidates (lowest p-values) from a simple threshold to a classical algorithm aimed at controlling the false discovery rate (FDR).

In Sect. 2, we discuss the new error model adopted. We present the challenges in CI estimation and the algorithm implementation for this work in Sect. 3. In Sect. 4, we show an intriguing trend-like behaviour observed in the residuals and how we explore it in the detection of binaries. We then present the new approach used for the statistical selection of objects in Sect. 5. We discuss the updates in the physical validation of the outcomes from the previous step in Sect. 6. We apply our revised approach to the astrometric data in *Gaia* FPR and present our results and discussions in Sect. 7. Finally, we draw our conclusions in Sect. 8.

2. Noise model

In this section, we present a statistical model of the data that accurately propagates the error properties from the observation (CCD level) to the transit level. This statistical model is critical to ensure reliable data exploitation, as it directly impacts the determination of p-values used for candidate detection and the construction of reliable confidence intervals (CIs) for the amplitude and period of the detected binaries. In Sect. 2.1, we recall the properties of the astrometric data and show that the random astrometric errors provided by *Gaia* FPR can be used to accurately estimate the standard deviation (hereafter std) of the CCD-level errors on the residuals in the AL direction. In Sect. 2.2 we then present the statistical data model per observation and at the transit level, along with the noise parameters and how they are computed. Throughout the paper, bold letters denote column vectors.

2.1. Properties of the Gaia astrometric data

Previous publications (David et al. 2023; Tanga et al. 2023; Gaia Collaboration 2018) have thoroughly described the main features of *Gaia* asteroid astrometry. Here, we briefly recall the main properties that directly affect their exploitation in our context. *Gaia* observes the targets' positions on the focal plane, which correspond to independent measurements on different CCDs. Each CCD position measurement is referred to as an 'observation'. The observations are grouped by 'transits', which occur over a short time span ($\approx$40 s). A maximum of $N = 9$ positions is possible per transit, although this is not often reached for minor bodies. Transits are irregularly spaced in time, with long periods without observations for a given target. Sequences of consecutive transits (minimum interval of 106 minutes) are common, but their frequency decreases with their length.

The maximum astrometric accuracy of *Gaia* is in the along-scan direction (AL), so the measurements can be considered essentially one-dimensional. The AL direction gradually changes orientation over time due to the precession of the satellite spin axis, so that any direction on the sky is scanned over time, with different AL orientations. The same applies to moving sources. The error model of each observation must therefore account for the AL direction and all dependencies of the derived positions on calibration errors of the focal plane, the reconstructed attitude of the satellite, and related effects. We distinguish two main components in the final error: a systematic component, assumed to be constant along a transit and common to all observations within the transit, and a random component, which differs for individual observations (Lindegren et al. 2021; Tanga et al. 2023).

As discussed in the next section, the systematic component (denoted μ) cannot be disentangled from the wobble signature, as both act as a common shift on observations over the same transit. However, wobbles can be detected by combining multiple transits over a window, as the signal changes over time[2]. As for the random component, estimating the variance of the random component is not straightforward. In L24, we used the empirical variance directly from the post-fit residuals in the AL direction obtained from the N observations of the transit. While this is a classical estimate, it is also significantly noisy because N is small ($N \leq 9$) and the variance of the estimator of the variance estimate decreases as $\frac{1}{N}$.

However, an estimate of the std of the random component in the right ascension and declination (denoted $\widehat{\gamma}$ below) is provided in the *Gaia* FPR data and can be projected in the AL direction. To assess whether this projected std is a good proxy for the actual std of the observation data, we selected AL residuals per observation $\widehat{\gamma}$ in a narrow range (one per panel in Fig. 1), plotted the resulting distributions, and computed the empirical std (reported as $\widetilde{\gamma}$ in the legends).

The empirical dispersion of the AL residuals (per observation) from *Gaia* FPR is marginally larger than the 'theoretical' error $\widehat{\gamma}$ in the AL direction provided by the *Gaia* FPR error

[1] We adopted these parameters in L24 as a compromise between the observation arc and the number of exploitable targets.

[2] The systematic component may also vary between transits, so that wobbles can be detected when these variations are small and/or when the systematic component is small compared to the wobble amplitude.

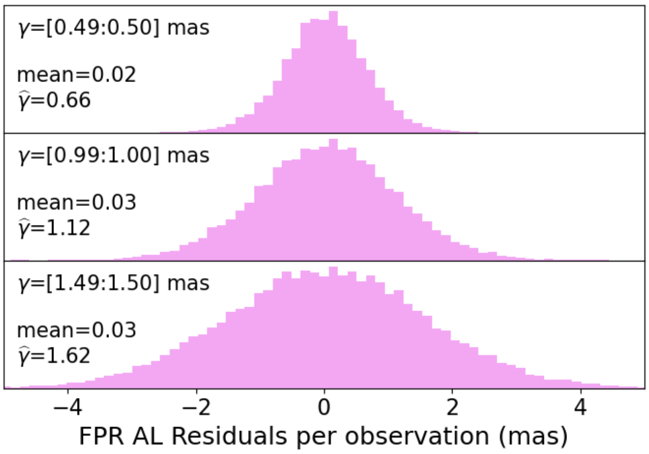

Fig. 1. Distribution of AL-projected residuals from *Gaia* FPR per observation (i.e. not averaged by transit). For each panel, we select an interval of per-observation random errors projected in AL (γ) from *Gaia* FPR. For all observations in the *Gaia* FPR catalogue with γ within this interval, we retrieve the corresponding residuals; their distribution is shown in pink, with the empirical mean and std $\widehat{\gamma}$ indicated. The distributions in the top, middle, and bottom plots contain 22k, 45k and 56k observations, respectively.

model. Several factors may explain this slight underestimation. First, the orbital-fit residuals include systematic components, which broaden the empirical distribution by acting as a random mean added to the residuals. However, this is not the only effect[3]. For example, recent analyses show that differences between the photocentre and the barycentre should be taken into account for the most accurate orbital fit, particularly for bright asteroids (Fuentes-Muñoz et al. 2024), and may therefore introduce additional scatter. In conclusion, although there is a slight disagreement between the residual dispersion and the theoretical uncertainty provided by the *Gaia* error model, our investigations (Fig. 1) indicate that the *Gaia* FPR value fairly reflects the actual dispersion and can therefore be used in subsequent processing steps.

2.2. Data model and parameter estimation

We first describe the data model per observation (i.e. at the CCD level), and then consider the transit-level data, which combine N individual observations. Let $\mathbf{t} := [t_1, \cdots, t_N]^\top$ denote the vector of observation epochs over one transit, r be the vector of residuals, and let $r_i := r(t_i)$ be the residual for a given observation obtained from the difference between the astrometric positions of the asteroid and the orbital fit to these measurements, projected in the AL direction.

These residuals are stored in the vector $\mathbf{r} := [r_1, \cdots, r_N]^\top$. As discussed in Sect. 2.1, the astrometric error comprises two components: an unknown systematic offset, denoted μ, and a random perturbation, modelled as Gaussian noise (n) with variance γ^2 provided by *Gaia* FPR (Sect. 2.1). The random perturbation for a given observation is denoted $n_i := n(t_i)$ by $n_i \sim \mathcal{N}(0, \gamma_i^2)$, where $\gamma_i^2 := \gamma^2(t_i)$. This leads to the following data model for the residuals per observation:

$$\begin{cases} \mathcal{H}_0 : r_i = \mu + n_i, \\ \mathcal{H}_1 : r_i = \mu + A \sin(2\pi f t_i + \varphi) + n_i, \end{cases} \tag{1}$$

where $\mathcal{H}_0$ denotes the hypothesis in which no wobble is present in the data, and $\mathcal{H}_1$ denotes the alternative hypothesis (wobble present). Under $\mathcal{H}_1$, the terms A, f, and φ denote the unknown amplitude, frequency, and phase of the sinusoid that models the binary wobble in the AL direction. Since the wobble period is much longer than the transit duration (≈40 s), this term can be considered constant over a transit.

From model (1), our first goal was to estimate the constant term from N measurements $\{r_i\}_{i=1,\cdots,N}$ and derive the transit-level residual (denoted y below), which contains the wobble signature under $\mathcal{H}_1$. The standard maximum likelihood estimation of the constant term in model (1) leads to

$$y = \sum_{i=1}^{N} \frac{r_i/\gamma_i^2}{1/\gamma_i^2}. \tag{2}$$

This estimate is clearly Gaussian and unbiased (its mean is equal to the constant term), with std $\overline{\gamma}$ given by

$$\overline{\gamma} = \sqrt{\frac{1}{\sum_{i=1}^{N} 1/\gamma_i^2}}. \tag{3}$$

Let k denote the index of the transit to which the N observations per transit belong, and t_k the corresponding transit epoch[4]. This leads to the following data model for the residuals per transit:

$$\begin{cases} \mathcal{H}_0 : y_k = \mu + \epsilon_k, \\ \mathcal{H}_1 : y_k = \mu + A \sin(2\pi f t_k + \varphi) + \epsilon_k, \end{cases} \tag{4}$$

where $y_k := y(t_k)$, $\epsilon_k \sim \mathcal{N}(0, \overline{\gamma}_k)$, with $\overline{\gamma}_k$ is the std given in (3) for transit k. For each target, we grouped the transit data into windows of observation (WO) containing K transit data points. This defines a time series $\mathbf{y} := [y_1, \cdots, y_K]^\top$. As discussed in L24, the detection method relies on a GLSP analysis that compares the sums of the weighted least-squares residuals obtained from fitting a constant and a constant-plus-sinusoid. We calibrated the ‘significance’ of the score reflecting this comparison via Monte Carlo (MC) simulations with a p-value.

We note that in the transit data model (4), the systematic term μ is not indexed by k. This is an approximation, as this bias may vary slightly within a window. One way to account for such possible variations is to model the bias signal as a low-order polynomial and to include it in the GLSP analysis of the K-points time series (Appendix C). While this increases the computational cost (each additional computation must be multiplied by about 10^5 candidates and 10^4 MC simulations per candidate), our investigations show that the results of such a model are often very similar to those obtained with a constant μ in model (4) (see Fig. C.1, top panel for a typical example). Consequently, we adopted this simpler model for almost all candidates, except those for which a linear trend was clearly detected in the residuals the studied WO (Sect. 4.

Figure 2 compares the distributions of the per-transit std obtained using three approaches: (1) a simple per-transit mean

[3] We estimated the std of the systematic component over all post-fit residuals (denoted $\widetilde{\mu}$) per transit $\tilde{\mu}$ as 0.27 mas, see L24. For the first panel, if this were the only cause, we should obtain $\widehat{\gamma} = \sqrt{\gamma^2 + 0.27^2} \approx$ 0.56 mas, which is smaller than the observed value of $\widehat{\gamma} =$ 0.66 mas. The mismatch is similar in the other dispersion ranges.

[4] The quantity t_k appears only in the sinusoidal term; since it is essentially constant over the transit with respect to the duration of a wobble, it can be taken as the mean of the N epochs, or by any of the t_i, with negligible impact.

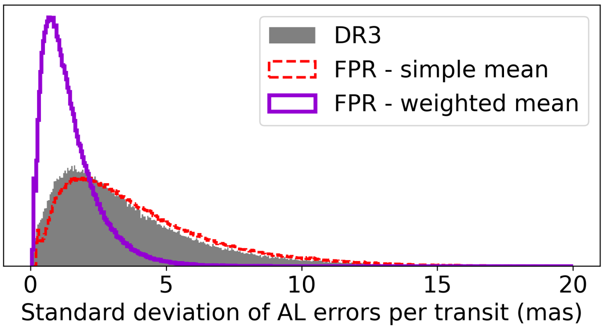

Fig. 2. Comparison between the distribution of 50 000 values of the per-transit standard deviation projected in the AL direction, obtained using different approaches for computing the transit data.

of the observation residuals, with std values provided by *Gaia* DR3 data; (2) a simple mean per transit of the observation residuals of *Gaia* FPR data, with std computed as $(\sum_{i=1}^{N}\gamma_i^2)/N$; and (3) the weighted mean of *Gaia* FPR data from Eq. (1), with std given by Eq. (3). The residuals per observation (CCD) were not available in *Gaia* DR3, which prevented the calculation of the weighted mean. This figure shows that when computing the transit data as a simple mean, the resulting data are similar for *Gaia* DR3 and *Gaia* FPR. However, *Gaia* FPR data computed using the weighted mean as in Eq. (1) exhibit substantially lower dispersion. Indeed, samples affected by larger errors are assigned lower weights, making the final estimate more accurate.

3. Confidence intervals

We denote the GLSP of the time series $\mathbf{y} := [y_1, \cdots, y_K]$ by $\mathcal{P}(f)$. We define the wobble frequency $\widehat{f}$ as that of the highest peak in the periodogram: $\widehat{f} := \arg\max_\nu \mathcal{P}(\nu)$.

We obtained the estimated amplitude $\widehat{A}$ and phase $\widehat{\varphi}$ through a standard weighted least-squares (WLS) fit of a sinusoid plus constant to the data[5]. We next address the problem of providing a reliable CI for these quantities.

3.1. Algorithm

This algorithm is inspired by bootstrap methods Efron (1979, 2010). In this section, we present numerical studies investigating the reliability of the CI using this algorithm. We provide the pseudo-code summarising the CI estimation algorithm (Algorithm 1) in Appendix A. We estimated the CI from empirical distributions of the amplitude and period estimates obtained through MC simulations.

Starting from the estimated parameter of the wobble $\widehat{f}$ and $\widehat{A}$, together with the other WO parameters, we generated M simulated time series consisting of a sinusoid with the estimated wobble amplitude, sampled at the considered epoch $\mathbf{t}$, to which we added a systematic offset $\mu^{(i)}$ and random noise $\mathbf{n}^{(i)}$. To add variability and thus robustness to the procedure, we generated a new phase (drawn from a uniform distribution) and a new systematic offset for each simulation i. This offset is consistent with *Gaia* residuals, and we modelled it as the realisation of a

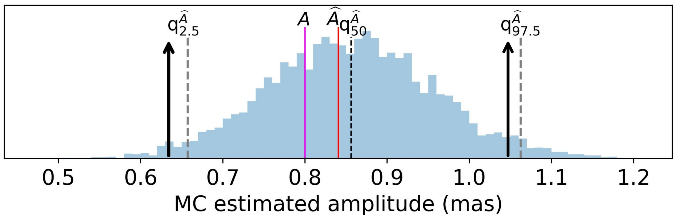

Fig. 3. Illustration of the CI computation for the amplitude estimate $\widehat{A}^{(i)}$ (blue distribution) produced by Algorithm 1 showing the true (unknown) wobble amplitude A (violet), the estimated amplitude $\widehat{A}$ (red), the median $q_{50}^{\widehat{A}}$ (dashed black), the quantiles $q_{2.5}^{\widehat{A}}$ and $q_{97.5}^{\widehat{A}}$ (dashed grey), and the final CI $\mathcal{I}_A$ for $\widehat{A}$ (black arrows).

Laplacian random variable with the parameters (step 3 of Algorithm 1; see also App. A of L24 for details). We derived the noise std at the transit level in the previous section ($\overline{\boldsymbol{\gamma}}$ in Eq. (3)). In practice, to remain conservative, the std values used in Algorithm 1 are slightly higher than those of Eq. (3) because, as shown in Sect. 2.1, those values are slightly underestimated[6].

For each MC time series, frequency, period, amplitude, and phase are re-estimated. This yields distributions of estimated frequencies (or periods) and amplitudes, from whose quantiles the CI can be computed. Figure 3 illustrates this procedure for an amplitude CI ($\mathcal{I}_A$). We show the nominal (unknown) wobble amplitude (A = 0.8 mas) and the estimated amplitude $\widehat{A} \approx 0.84$ mas. We generated the simulated data used to estimate $\widehat{A}$ as a sinusoid with amplitude A and a random (uniform) phase, sampled at the same epochs $\mathbf{t}$ as an arbitrary WO from a *Gaia* target, with values for the added offset and noise std consistent with the *Gaia* FPR data model. Algorithm 1 produces the distribution of estimates $\widehat{A}^{(i)}$, whose median $q_{50}^{\widehat{A}} \approx 0.86$ mas. We estimate the quantiles as $q_{2.5}^{\widehat{A}}$ and $q_{97.5}^{\widehat{A}}$. Here $q^{\widehat{A}\star} := \max\{q_{97.5}^{\widehat{A}} - q_{50}^{\widehat{A}}, q_{50}^{\widehat{A}} - q_{2.5}^{\widehat{A}}\} = q_{97.5}^{\widehat{A}} - q_{50}^{\widehat{A}} \approx 0.205$ mas, which gives the resulting CI $\mathcal{I}_A$. In this case, the CI contains the true amplitude value. As shown in the next section, this occurs with a probability of around 95% for most amplitudes.

3.2. Changes with respect to Liberato et al. (2024) and performance evaluation

Algorithm 1 presents several important changes compared to the previous implementation used in L24. First, the noise added in step 4 was previously uniform (in an interval ranging from 0 to the estimated std of the error in *Gaia* DR3), which caused an underestimation of the noise effect. Second, the amplitude was previously estimated as the difference between the maximum and minimum of the sinusoid fitted to the data, rather than from the WLS coefficients $\widehat{\beta}_1$ and $\widehat{\beta}_2$ mentioned above, resulting in less accurate amplitude estimates. Third, the quantiles $q_{2.5}^{A}$ and $q_{97.5}^{A}$ were previously estimated using binned histograms of $\widehat{A}^{(i)}$ and $\widehat{T}^{(i)}$, which created a slight but unnecessary dependence on the bin width. In contrast, we obtained the quantiles computed in steps 12 to 14 using a consistent estimator (David & Nagaraja 2004).

We now evaluate the actual false coverage rate (FCR) of the derived CI intervals. The FCR is the probability that $A \notin \mathcal{I}_A$,

[5] The time series can be written as $y_k = \beta_0 + \beta_1\cos(2\pi\widehat{f}t_k) + \beta_2\sin(2\pi\widehat{f}t_k) + \epsilon_k$ for $k = 1, \cdots, K$. The WLS estimation yields estimates $\widehat{\beta}_0$, $\widehat{\beta}_1$, and $\widehat{\beta}_2$ (Appendix C). The estimated parameters are then given by $\widehat{\mu} = \widehat{\beta}_0$, $\widehat{A} = \sqrt{\widehat{\beta}_1^2 + \widehat{\beta}_2^2}$, and $\widehat{\varphi} = \operatorname{atan}(-\frac{\widehat{\beta}_2}{\widehat{\beta}_1})$.

[6] Specifically, we used $\sigma := \sqrt{\tilde{\mu}^2 + \overline{\gamma}^2}$, where $\tilde{\mu}$ denotes the systematic error component in the AL direction as estimated in L24; however, the value of σ is very close to $\overline{\gamma}$ because generally $\tilde{\mu} \ll \overline{\gamma}$.

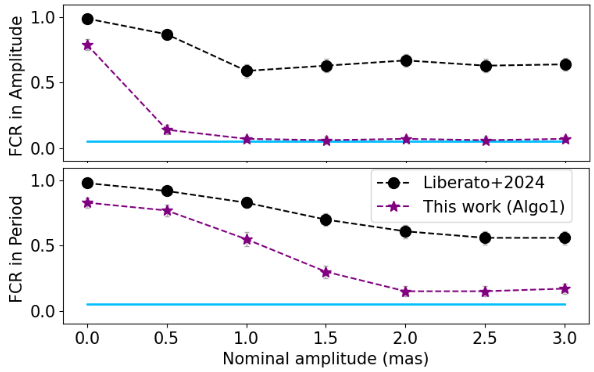

Fig. 4. False coverage rate (FCR) with error bars (in grey) for the CI of the estimated amplitude (top) and the estimated period (bottom) for the two versions of the CI algorithms. The horizontal solid blue line indicates the 95% coverage targeted.

which we estimate as the fraction of simulated cases in which the CI does not contain the initial value. Our approach is based on MC simulations. For a given wobble (sinusoidal signal) with known period T, amplitude A, and phase φ, we computed a set of 100 simulated data (time series). For each time series, we estimated the amplitude and the period, applying Algorithm 1 and verifying whether A and T belong to the corresponding CI. The epochs $\mathbf{t}$ and noise std $\boldsymbol{\sigma}$ vary for each MC simulation, being randomly drawn from those of the WO in *Gaia* FPR, while we generated the offset as a Laplacian random variable (step 5 of Algorithm 1). We repeated this experiment by varying the nominal amplitude of the signal A, between 0 and 3 mas.

In Fig. 4 we show the empirical FCR as a function of the initial (or nominal) wobble amplitude. In the top panel we observe that the CIs from Algorithm 1 reach the expected 95% confidence level for input wobble amplitudes larger than about 1 mas (see Appendix D). In contrast, the L24 algorithm (without the modifications described above) often fails to cover the true amplitude. The period estimation (bottom panel) is more challenging, although Algorithm 1 still performs better. In this case, this algorithm provides CIs for periods that are valid with a probability of 90% instead of 95% for wobble amplitudes of about 2 mas or higher.

4. Linear trend detection

For fewer than 2% of the WOs explored, the residuals show a global trend that dominates the time variation (see the two examples in Fig. 5). Such trends could be due to variation in the systematic offset discussed in Sect. 2.2, to other unknown artefacts, or to wobble periods that are much longer than the WO. Hence, these cases require dedicated processing. To address this, we set up a two-step procedure: (1) a trend-detection step, using a test on the Pearson correlation coefficient c between epochs and residuals of the WO in question (Appendix B); (2) a dedicated period search using an extension of the GLSP that includes a linear trend in its data model (Appendix C).

For step (1), we computed a p-value[7] associated with each correlation score c. If this p-value is $< 0.5\%$ then the WO was

[7] We performed 10^4 MC simulations in which we estimated c on simulated data sampled at the same epochs as the WO, using the noise std and systematic as described in the data model. The p-value of c corresponds to the proportion of this population of 10^4 correlation coefficients (obtained with zero correlation between $\mathbf{t}$ and $\mathbf{y}$) that is larger than c.

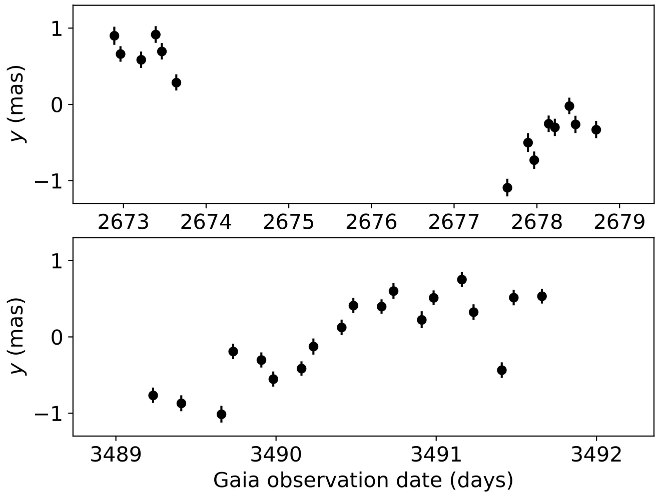

Fig. 5. Astrometric residuals in the AL direction per transit y versus *Gaia* observation epochs for the known wide binary asteroid (317) Roxane (Drummond et al. 2021) shown in two different WOs. Both are flagged as ‘trendy’.

flagged, and are refereed to as a ‘trendy’ throughout the paper. For example, for the two cases shown in Fig. 5, the correlation coefficients are $c = -0.87$ (top) and $c = 0.77$ (bottom), with both p-values below 10^{-4}. In step (2), we processed trendy WOs using the same procedure as for the non-trendy WOs, but using our implementation of the combined GLSP with a trend term instead of the conventional GLSP used in period search for the rest of the sample (Appendix C). For all WOs tested and a selection threshold of 0.5%, one expects approximately 300 false positives under $\mathcal{H}_0$. We identify 1,433 trendy WOs, representing a significant excess relative to the null expectation. This excess indicates that a substantial fraction of the sample exhibits genuine correlated residuals.

Adopting a conservative FDR estimate, we infer that the majority ($\approx 80\%$) of the selected objects likely correspond to real correlations rather than statistical fluctuations. However, 207 of these WOs correspond to the first or last observations available for such objects in *Gaia* FPR. Boundary observations inherently contribute less new information to the orbital fitting procedure, leading to larger uncertainties and potentially affecting residuals near the start or end of the observation arc (Milani & Gronchi 2010; Spoto et al. 2018). Therefore, as trend detections in these 207 WOs are more likely to be non-physical, we omitted them. Among the remaining 1226 trendy WOs, 45 objects, including the known binary (317) Roxane, have two WOs flagged as trendy. For these objects, the linear trends are less likely to be spurious since they are detected in different epochs. This makes them interesting targets for further studies.

5. Statistical selection of candidates

In L24, we performed the statistical selection of binary candidates in two steps. In the first step, we computed a p-value for each WO using the *Gaia* noise model and selected the object if this p-value was less than 5%. In the second step, we computed a quality factor Q and selected the candidate if Q was greater than 50%, meaning that the detected period value should be recovered, within a given tolerance, with a probability greater than 50% in simulated noisy sinusoids with that period.

For step one, the selection process for the p-values relies on a simple threshold rule. This approach provides an estimate of the average number of false detections when $\mathcal{H}_0$ is true for all candidates, equal to 5% of the total number of cases tested. However, a more interesting criterion is the proportion of false discoveries, that is, the FDR in the selected list.

We used the Benjamini–Hochberg (BH) procedure (Benjamini & Hochberg 1995) for this purpose. This method controls the FDR if the p-values in the sample are independent and uniform. In our case, independence of the p-values follows from the independence of the data from one transit to another. The uniform distribution depends on the accuracy of the noise model[8] and we show that empirically that this holds in this section.

To choose the target FDR for step one, we must take several points into account: (1) low p-values may arise from statistical fluctuations of noise, from genuine wobble signals, or, in rarer cases, from other sources (e.g. systematic or instrumental effects, undetected long-term trends, or locally underestimated uncertainties); (2) we wish the list to contain some known binaries, even if their signatures are weak in the data and their p-values are not particularly small; and (3) we can afford a list with a large proportion of false discoveries because the subsequent physical validation step (Sect. 6) is expected to remove a large fraction of them. For these reasons, the adopted target FDR is 75%.

Regarding step two, we noted in Sect. 4.3 of L24 that while the Q factor can provide valuable information in some circumstances, it can also be misleading, as, conversely, large values of Q may arise from pure noise fluctuations, whereas true but weak detections may yield low values of Q. Hence, we did not use the Q factor in the selection, but it remained encapsulated in the provided CI interval.

In the rest of this section, we first present a new method for statistically combining the p-values of objects with more than one WO, and then apply the whole statistical detection pipeline to *Gaia* FPR data. Finally, we present numerical tests aimed at verifying the validity of the approach.

5.1. The method applied to Gaia FPR

Of the approximately 157,000 objects in *Gaia* FPR, 47,896 objects contain at least one WO within the 66 month of *Gaia* FPR data. This corresponds to 60,152 WOs to be analysed. For this data set, we computed the transit-averaged residuals from the orbital fit as in Eq. (1) and the corresponding errors as in Eq. (3). For each WO, we applied a period analysis using GLSP, as briefly explained in Sect. 1 (Appendix C), using the nifty-ls implementation in Astropy (VanderPlas 2018; Garrison et al. 2024), and estimated the corresponding empirical p-values through 10 000 MC simulations.

With the full sample of 60,152 p-values from all of the WOs, we first separate the objects with a single WO in *Gaia* FPR from those with multiple WOs, and perform the BH-based selection independently in each group at a target FDR of 75% (Table 1).

[8] By definition, if S denotes the random variable corresponding to the GLSP score (i.e. the value of the highest peak in the periodogram), the p-value p for a particular value s (a realisation of S) is $p := \Pr(S > s \mid \mathcal{H}_0)$. Hence, the probability that the random variable P is less than p is $\Pr(P < p \mid \mathcal{H}_0) = \Pr(S > s \mid \mathcal{H}_0) = s$ (since any score larger than s will have a p-value less than p) showing that P is uniform. This also shows that uniformity of the p-values holds as long as $\Pr(S > s)$ is accurately estimated. If this calibration is not accurate, small p-values may be more likely than expected, leading to an increased and worse uncontrolled false alarm rate.

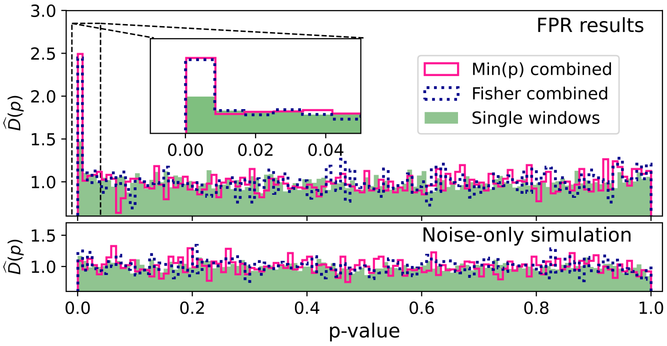

Fig. 6. Estimated density distribution $\widehat{D}(p)$ of p-values for single-window objects (green), and for multiple-window objects with p-values combined using the Fisher method (magenta) and combined with the min(p) method (blue). Top panel: p-value distributions from the FPR data. The inset shows a zoom of the area delimited by the dashed line. Bottom panel: Results from one noise-only simulation run.

For the 37,354 objects with only one window, each object is associated with a single p-value, which is treated independently. As a result, we obtain 713 selected objects, of which about 25% (≈178) are expected to be true detections.

For the 10 542 objects with multiple windows, we computed an independent p-value for each window. Combining both windows as a single data set requires a different approach. Such a method would need to account for variations in the observing geometry, which is beyond the scope of this work. However, individual windows may yield marginal or no detections, while their combined behaviour may increase the detection strength when considered jointly. Hence, our goal was to combine the multiple p-values associated with a given object to evaluate the global evidence against the null hypothesis. For this purpose, we applied two complementary p-value combination methods to objects with multiple windows.

The first is Fisher's method (Fisher 1948), which accumulates evidence from the combined windows and is most powerful when many p-values are moderately small. It is therefore sensitive to weak but consistent signals spread over several WOs of an object. This method combines n independent p-values, $p_1, p_2, \ldots, p_n$, by calculating

$$k_F := -2 \sum_{i=1}^{n} \ln(p_i). \tag{5}$$

If the null hypothesis holds for all tests, k_F follows a χ^2 distribution with $2n$ degrees of freedom. From k_F, the Fisher combined score can be written as

$$p_{\text{Fisher}} := 1 - \Phi_{\chi^2_{2n}}(k_F), \tag{6}$$

where $\Phi_{\chi^2_{2n}}$ is the cumulative distribution function (CDF) of a χ^2_{2n} random variable. This quantity is also a p-value if the initial p-value sample is independent and uniform (Fig. 6), which is the case here.

The second is the min(p) method (Tippett 1931), which is sensitive to the presence of at least one strong signal. This method is powerful in scenarios where a strong detection appears in only one of the WOs of an object. The min(p) method takes the smallest value among them,

$$k_m := \min\{p_1, p_2, \ldots, p_n\}, \tag{7}$$

Table 1. Distribution of WOs, objects, and WOs per object in each group, together with the total number of WOs and objects in the analysed sample.

# of WO per object	# of WO	# of objects	# of objects selected with BH
1	37 354	37 354	713
>1	22 798	10 542	735
2	18 020	9 010	596
3	4 107	1 369	129
4	584	146	8
5	75	15	2
6	12	2	0
total	60 152	47 896	1 448

and adjusts it to account for the number of values in the set. The p-value associated with k_m can be easily computed as

$$p_{\min(p)} := 1 - (1 - k_m)^n. \tag{8}$$

Using min(p) and Fisher, we probed the two limiting and physically relevant cases—dominance by a single window versus coherent evidence from many windows—while relying on methods with simple, well-defined null distributions and a straightforward interpretation, without requiring assumptions about the distribution of true detections or prior knowledge of it (Heard & Rubin-Delanchy 2018; Vovk & Wang 2020).

After applying both combination methods and using the BH selection with a target FDR of 75% on the two sets of combined p-values, we obtain 605 objects selected using the min(p) method and 588 selected using the Fisher method. The union of the two multi-window selected object samples yields a total of 735 objects selected, as shown in Table 1. Finally, taking the union[9] of the single-window objects selected with the multi-window selected objects, we obtain a final sample of 1,448 statistically selected binary candidates from the *Gaia* FPR astrometric data.

For objects with multiple windows, we used all WOs associated with a selected object in the subsequent selection steps (discussed in the next sections), even if only one of the WOs successfully passed the BH selection threshold (because a WO that does not present a signal strong enough to be selected with a low p-value can still be useful to confirm the period detected in the main WO).

5.2. Performance evaluation using control simulations

To assess the reliability of the statistical selection procedure and quantify the level of spurious detections expected in the absence of any true astrometric signal, we performed simulations on signal-free data. We verified that the adopted period search, p-value estimation, combination procedures, and BH selection behave as expected under the null hypothesis, providing a reference against which the results obtained in real data *Gaia* FPR can be interpreted.

1) *FDR control*. We performed 10 000 simulations on sets of 60 000 random p-values mimicking the single- and multi-window samples, applying the same p-value combination procedures. We obtained an average final FDR of 74.5% for the non-combined sample, 74.9% for the Fisher-combined dataset, and 74.6% for the min(p)-combined p-values, confirming that our procedures guarantee the expected FDR in all cases.

2) *Full statistical detection pipeline*. To compare our results on the *Gaia* FPR dataset with those obtained from mirror but signal-free data sets, we performed ten independent simulation runs. For each simulation run, we used all the ≈ 48 000 objects explored in the candidates search, but replacing the residuals with noise plus systematic offset as described in the data model (Sect. 2). We then submitted each simulated dataset to the same pipeline used for the *Gaia* FPR exploration. This approach ensured that the only meaningful difference between the simulations and the *Gaia* FPR data is the certainty that no wobble is present in the simulations.

An interesting first result concerns trend detection. We detect about 300 trendy WOs per simulation run, as expected from the 0.5% threshold adopted. This suggests that it is unlikely that most of the 1226 WOs in which we detect trends (Sect. 4) are caused by noise fluctuations, especially for the 45 objects with two trendy WOs. Therefore, they warrant further investigation. A second result concerns the p-value distribution. Under $\mathcal{H}_0$, p-values are expected to follow a uniform distribution on [0,1], since they are derived from the quantiles of the test statistic distribution under noise-only simulations. As discussed above, this uniformity depends on the accuracy of the adopted noise model and its parameters: any mismatch leads to an incorrect calibration of the quantiles, and hence to departures from uniformity, for instance, in the presence of genuine signals or systematic effects.

Figure 6 illustrates this behaviour. The bottom panel shows the empirical p-value distributions (for the single, Fisher, and min(p) statistics) obtained from one of the ten control simulation runs. Those are consistent with a uniform distribution, as expected, because in the simulations the noise is generated according to the model, so the p-values are well calibrated. The top panel shows the corresponding distributions for the *Gaia* FPR data. Two features are apparent: first, a broad plateau, indicating that for most objects the p-values are consistent with uniformity and that the noise model used to calibrate the GLSP scores (Sect. 2) is appropriate; second, a clear probability excess at the smallest p-values. This excess indicates a subset of objects whose behaviour is inconsistent with $\mathcal{H}_0$, likely due to binary systems and other effects that are not consistent with the noise model.

Table 2 summarises the results from the control simulations and from the *Gaia* FPR data analysis. The number of WOs selected by our procedure in the control simulations is one order of magnitude smaller than those obtained in the *Gaia* FPR period search. These results support the idea that although some detections may be spurious (due to unknown artefacts or modelling errors), a significant number of the period detections in the *Gaia* FPR residual exploration are likely to be real. In fact, in the absence of any such effect, roughly 25% (≈350) should be true detections.

6. Physical validation of candidates

After assessing the statistical relevance of the periods detected in the astrometric residuals, we verified whether the signal detected in the astrometric data was consistent with plausible physical parameters for a binary system.

[9] We highlight that while the FDR is controlled at the target level for each set (single, min(p), and Fisher), this is not guaranteed theoretically for the union data set.

Table 2. Benjamini–Hochberg (BH) selection counts from simulations and *Gaia* FPR asteroid data.

Run	Single	Fisher	min(p)	Union	Total
sim 1	53	11	87	89	142
sim 2	26	11	70	71	97
sim 3	11	9	6	11	22
sim 4	14	56	48	76	90
sim 5	14	62	79	100	114
sim 6	74	95	41	98	172
sim 7	54	24	30	41	95
sim 8	31	24	12	24	55
sim 9	94	12	2	13	107
sim 10	140	35	9	35	175
sim mean	51.1	33.9	38.4	55.8	106.9
Gaia FPR	713	588	605	735	1448

Notes. The second column indicates the number of objects selected from the single-window list; the third and fourth columns show the number of objects selected from multi-windows objects with p-values combined using the Fisher and min(p) methods, respectively; the fifth column shows the number of objects resulting from the union of Fisher and min(p) selected objects; and the last column shows the total number of objects selected in each run.

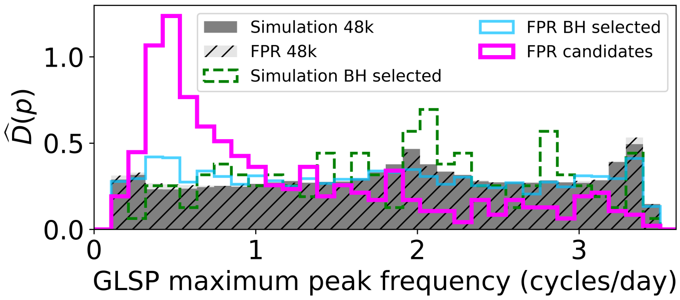

Fig. 7. Density estimation on the distribution of GLSP maximum peak frequencies for one of the simulations described in Sect. 5 (solid grey bars); objects selected by the BH method from the simulation (hatched bars); all 48k objects from *Gaia* FPR (dashed green); objects from *Gaia* FPR selected statistically using the BH method (thin blue); and the final list of *Gaia* FPR-selected candidates (thick magenta).

As explained in L24, we derived the minimum bulk density profile as a function of the mass ratio by combining the simple binary wobble model from Hestroffer et al. (2010) with Kepler's third law. The resulting expression combines the measured parameters (period and amplitude of the signal) with literature data (diameter of the equivalent sphere) obtained from the SsODNet database (Berthier et al. 2023). Setting thresholds for plausible densities (Carry 2012; Scheeres et al. 2015) allowed us to determine possible ranges of size ratio and separations. By adopting a density range 0.8–5.5 g/cm^3, we obtain a list of 988 WOs (729 objects) whose minimum density estimates fall within the chosen range.

Whenever necessary, we re-constrained (trimmed) the intervals of possible separations in one or both extremes. At minimum, we adopted the fluid Roche limit of the system[10]. For maximum separation, we set a limit of 20% of the Hill radius[11], assuming that the density could be as high as 5.5 g/cm^3.

We rejected candidates whose separation had to be re-constrained at both the minimum and maximum, as their properties can be considered to be very weakly constrained and thus less reliable. Additionally, when this procedure left only one WO for multiple-window candidates, we retained it only if its p-value was lower than 4.2%, the highest selected by the BH method in the multiple-window approach. After applying these criteria, 353 binary candidates (and 421 WOs) remain.

We recall that, due to the single-dimensional nature of our approach, the measured wobble signature is a projection of the real photocentre offset and is thus taken as a minimum value. Therefore, the derived binary parameters (density, separation, and mass ratio) are based on a minimum wobble amplitude and correspond to lower limits (or interval estimates) of the true values. Our estimates are limited by the signal measured in the astrometric residuals and may differ from the values obtained with other observational techniques. Additionally, phase and shape effects (not accounted for in our simplified point-source model) can introduce photocentre offsets comparable to or greater than the expected binary-induced signal in some configurations (Pravec & Scheirich 2012). These effects may bias the inferred parameters or reduce the satellite wobble detectability, but modelling them requires prior knowledge of the system or a probabilistic approach, which is beyond the scope of this work.

7. Results and discussion

Gaia is a complex system whose scanning law, combined with the motion of the asteroids and the satellite, can potentially inject spurious frequencies into the data. It is therefore interesting to compare the distribution of periods in our candidate sample with those obtained from simulations with random noise.

Figure 7 shows that the distributions from the full sample with all of the 48 000 objects in the simulations and in the *Gaia* FPR data are very similar, indicating that the simulations successfully reproduce the dominant noise-driven behaviour of *Gaia* astrometric residuals. Additionally, we observe two bumps at frequencies around 3.6 and 2 cycles per day in both full-sample distributions, which are likely aliases of the frequencies associated with the motions of the *Gaia* satellite (Cellino et al. 2024).

The WOs statistically selected by the BH procedure, both in *Gaia* FPR and in the simulations, closely follow the same distribution as the full samples. This is consistent with the fact that we adopted an FDR of 75%, implying that three-quarters of the selected objects at this stage are likely false detections. However, our final *Gaia* FPR selected sample shows a very different frequency distribution, suggesting that the physical validation step has likely eliminated the majority of spurious detections.

Figure 8 shows the distribution of the amplitudes and periods of the selected WOs. We see that shorter periods (<24h) appear to be favoured by our method, as well as wobbles amplitudes of 0.3–1 mas, consistent with those obtained in L24 (Sect. 7.3). The updates to our selection procedure do not strongly affect the overall distribution. However, these low amplitude values tend to be poorly estimated, as we show in Fig. 3.

Among the final candidates, 27 objects exhibit at least one trendy WO, demonstrating that the de-trending procedure (Sect. 4) effectively recovers signals for these cases. In addition,

[10] Asteroids in the size range of our targets are likely rubble piles; therefore, adopting the fluid Roche limit is a more conservative approach.

[11] Asteroid satellites are typically in compact configurations with separations ~1% of the Hill radius; thus, assuming separations for our candidates up to 20% is generous but still physically realistic.

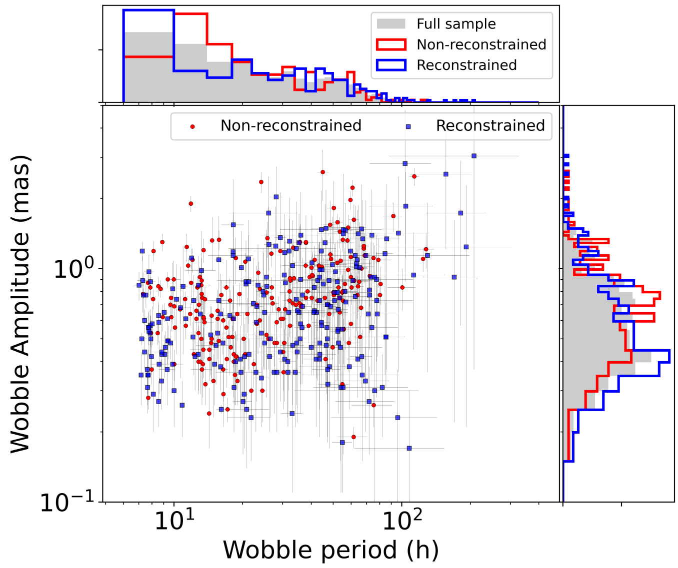

Fig. 8. Amplitudes of the selected sample as a function of period. The pink dots correspond to separation intervals that did not require re-constraining, while the blue squares were re-constrained, as explained in Sect. 6. The histograms show the distribution of the entire sample (solid grey), while the coloured lines represent the same two categories.

45 objects show two WOs dominated by trends rather than fluctuations. A notable example is (317) Roxane (Fig. 5), a known wide binary (Drummond et al. 2021) with a secondary orbital period (~12 days) much longer than the maximum WO span allowed by our selection. This indicates that, even when the wobble period cannot be directly measured, linear trends in astrometric residuals may signal the presence of wide binaries. For this reason, we select these 45 objects (List available in Liberato et al. 2026) as targets deserving deeper investigations.

7.1. Selected known binaries

Cross-matching our candidates with the Johnston's Archive list of asteroids with satellites (Johnston 2025), we find that among the 353 objects, nine are known (or highly suspected) binaries that satisfy all the selection parameters adopted.

- (720) Bohlinia was identified in L24 and later confirmed by photometry (Gorshanov et al. 2025). The photometric period $T =$ 17.418±0.006 h (or double) is consistent with our estimates $\widehat{T} =$ 17.489±0.130 h (L24) and $\widehat{T} =$ 17.748±0.160 h (this work). The separation estimate from photometry *sep*= 73.47±0.01 km is close to our findings of 77.04±8.55 km in L24 and to the lower end of 87.18 km (this work). Differences are likely due to the strong dependence on assumed physical properties, and especially to the new data model adopted, which led to slightly different wobble amplitude measurements.
- (1509) Esclangona is a known wide binary with a size ratio $k \sim 0.3$, *sep* ~140 km, and $T \sim 23$ d (Merline et al. 2003), well beyond the maximum observing window. We detect a trend with p-value$_{corr}$ = 0.541%, slightly above our threshold. The period is close to the upper limit of the searched range. This case illustrates the limitation of our approach for wide systems, while still hinting at binarity through the trend in the residuals.
- (1770) Schlesinger is a suspected binary based on reported mutual events, without reliable parameters. We selected two WOs: one weak, and one strong, with $\widehat{T} =$ 53.37±0.07 h and p-value < 0.001%. Combined with previous suspicions, this makes Schlesinger a highly probable binary.
- (1879) Broederstroom hosts a satellite with a reported $T =$ 47.83±0.02 h and $k = 0.34 \pm 0.02$ (Benishek et al. 2024). We correctly detect $\widehat{T} =$ 50.04±2.83 h and estimate a smaller size ratio $0.12 \leq k_1 \leq 0.24$. This discrepancy is most likely due to the observation geometry, as explained in Sect. 6, where the projection of the wobble in the AL direction leads to a smaller amplitude and, consequently, mismatched mass ratio and separation estimations, similar to the case of (4337) Arecibo (see below).
- (1967) Menzel was identified as a binary candidate in L24 and confirmed by photometry (Monteiro et al. 2024). Photometry yields $T =$ 63 h. We estimate $\widehat{T} =$ 32.4308±1.01 h, which is about half of the photometric period, likely due to the limited length of the WO. The detection is very clear in the astrometry, both in *Gaia* DR3 and in *Gaia* FPR.
- (2871) Schober is a known binary, with $T =$ 42.47±0.02 h and $k > 0.28$ (Benishek et al. 2023). We obtain $\widehat{T} =$ 50.65±21.65 h and size ratio intervals of $0.1 \leq k_1 \leq 0.3$ and $0.76 \leq k_2 \leq 1$, consistent with the published parameters within uncertainties.
- (4337) Arecibo is a synchronous binary with $k \sim 0.19$ and $T \sim$ 32.97 h (Gault et al. 2022; Tanga et al. 2023; Liu et al. 2024). We measure $\widehat{T} =$ 35.4± 4.0 h, consistently, but estimate a maximum k of ~0.1 3. This underestimation is due to unfavourable observation geometry (the AL-projected wobble amplitude reaching only ~8.5% of the system separation), or to the flattening of the components, as mentioned in Tanga et al. (2023).
- (31450) Stevepreston has a satellite with $k > 0.22$ and $T =$ 53.47±0.07 h (Pray et al. 2015). We find $\widehat{T} =$ 61.7±3.1 h, $0.1 \leq k_1 \leq 0.126$ and $0.96 \leq k_2 \leq 0.98$. The period agreement is poor, but we still obtain a strong and clear detection with p-value < 0.001. A possible explanation is the detection of a second, further (and maybe smaller) undetected satellite.
- (55637) Uni is a ~660 km Kuiper belt object (KBO) with a satellite at 4770 ± 40 km and $T =$ 199.42 h (Brown & Suer 2007; Brown 2013), far longer than the *Gaia* consecutive observations span. We detect $\widehat{T} =$ 45.02±1.23 h, close to three times the primary rotation period (~14.4 h), which could trace the photocentre shifts due to rotation of the primary. However, a more likely explanation would be the detection of a second satellite. A weak trend is still visible in the data, consistent with the presence of the known satellite.

7.2. Candidate's binarity likelihood

The motion of the photocentre with respect to the centre of mass of the system, visible in the residuals of *Gaia* astrometry, can originate not only from satellites, but also from irregular shapes of single objects, provided that its amplitude is large enough to be detected (Kaasalainen & Tanga 2004; Dell'Oro & Cellino 2012). (21) Lutetia provides an example of this ambiguity, as illustrated in Tanga et al. (2023). This raises the question whether all our current *Gaia* FPR candidates are likely binaries.

Figure 10 compares the wobble amplitude with the average apparent size (computed for each WO). It is divided into three regions. Region (1) contains objects with apparent sizes up to about 12 mas, corresponding to typical diameters smaller than 12–18 km in the main belt. Most of the known binaries selected

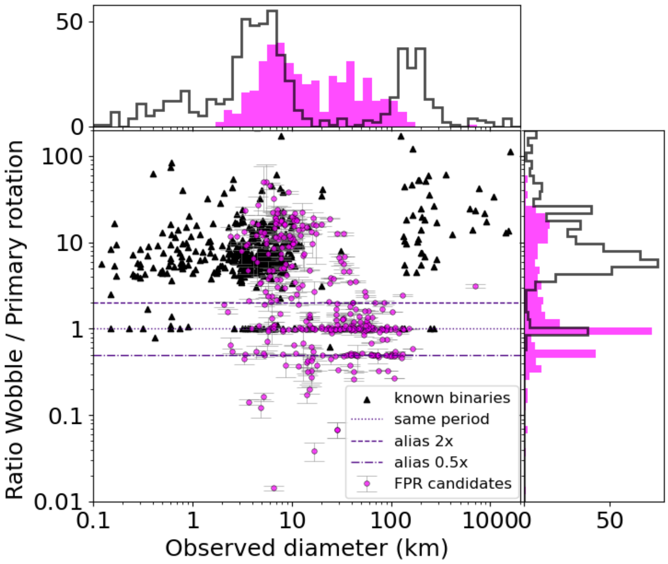

Fig. 9. Population of known binary asteroids (black triangles) compared with our *Gaia* FPR binary candidates (pink circles), in the plane defined by diameter and wobble period normalised to the rotation period of the primary (as usually derived from photometry). A few peculiar objects with a very low period ratio appear at the bottom of the plot. These correspond to very slow rotators that may also have incorrect photometric periods.

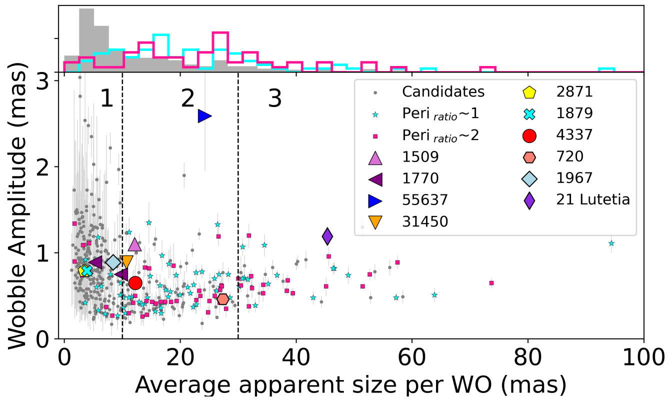

Fig. 10. Distribution of *Gaia* FPR candidate WOs in apparent size during the observation versus the measured wobble amplitude. The grey dots represent all *Gaia* FPR WOs selected in the candidate sample. The green stars represent WOs in which the estimated wobble period has a period ratio of about ∼1 with respect to the photometric rotation period of the object, while the pink squares indicate WOs with a period ratio of ∼2.

in our sample lie in this region. They span a wide range of wobble amplitudes, mostly below 1.5 mas with larger uncertainties, consistent with lower astrometric accuracy for fainter objects. The known binaries in this region were discovered by photometry. This suggests that many candidates in region (1) may also be appropriate targets for this technique, as demonstrated for (1967) Menzel (Monteiro et al. 2024).

Region (3) contains the largest apparent sizes, exceeding 30 mas, with typical diameters larger than 40–50 km. These objects exhibit a different behaviour, with small wobble amplitudes despite their larger size compared to those in (1). The slightly increasing trend is suggestive of a wobble proportional to size, as expected for large non-binary bodies. This size range hosts the largest fraction of wobble periods coinciding with the rotation of the primary (sometimes with an alias of double or half the value) as shown in Fig. 9.

The presence of (21) Lutetia in this sample is illustrative: we estimate a period of (8.12 ± 0.03 h), closely matching its rotation (8.168 h; Carry et al. 2010; Sierks et al. 2011). Moreover, the Rosetta mission excluded the presence of satellites capable of producing the observed wobble (Bertini et al. 2012). As the evidence is clear for this specific object, we excluded it from our candidate list. We retain the other candidates and flag them in the list to allow observers to perform further verifications.

Finally, in the intermediate region (2), roughly corresponding to objects of 15–50 km, we find several confirmed binaries in our candidate sample. Wobble amplitudes here have moderate uncertainties, indicating stronger and cleaner signals. Most suspected synchronous binaries fall in this region, as shown by the histogram on the top of Fig. 10, which corresponds to the underrepresented population of intermediate–size binaries consistent with formation via moderately catastrophic impacts (Durda et al. 2004). Therefore, the fact that several known binaries detected lie within these limits, and that most objects in this region are suspected synchronous binaries within the expected size range for such a formation mechanism, provides additional evidence that this is likely a binary–rich region.

Visual inspection of several shape models available on the DAMIT database (Durech et al. 2010) reveals that confirmed binaries often have problematic single-body shape solutions. For example, (4337) Arecibo has sharp edges, while (720) Bohlinia appears unrealistically elongated; in both cases, this is clearly a result of binarity (Ďurech & Kaasalainen 2003). Among our candidates with diameters larger than 20 km, some shapes are smooth and spheroidal, but others present similar features. In region (2), (1105) Fragaria exhibits a sharp edge similar to Arecibo; (1127) Mimi (see also binary features in Lallemand et al. 2026), (519) Sylvania, (542) Susanna, and (6475) Refugium display significantly large elongation. In region (3), (303) Josephina shows similarities to Arecibo, while objects such as (103) Hera, (538) Friederike, (605) Juvisia, (977) Philippa, (2906) Caltech, and (625) Xenia, exhibit large flat surfaces that may reflect poorly modelled concavities rather than companions.

In summary, candidates in regions (1) and (2) are the most robust in the sample, but we cannot exclude that at least some of the objects in region (3) have satellites, possibly with properties different from those of smaller objects. We therefore decided not to discard them, as they can be interesting targets for other techniques. Users of our list should, however, keep in mind the possible ambiguity in their cases.

7.3. Comparison with Gaia DR3 results

Gaia DR3 and *Gaia* FPR provide astrometry for the same number of asteroids (about 150 000), obtained over 34 and 66 months, respectively. Consequently, the number of WOs that we extracted increased from 30 030 to 47 896, an increase of about 60%. However, the lower number of binary candidates in this work (343, compared to 358 in L24) shows that our implemented improvements lead to a more conservative approach. In particular, instead of a simple threshold in *Gaia* DR3, in this work we used an FDR-based selection rule, tuned so that on average roughly one fourth of the selected objects (≈ 1448/4 = 362) are not spurious detections. Interestingly, 343 candidates remain after the physical validation steps. Moreover, we find an overlap

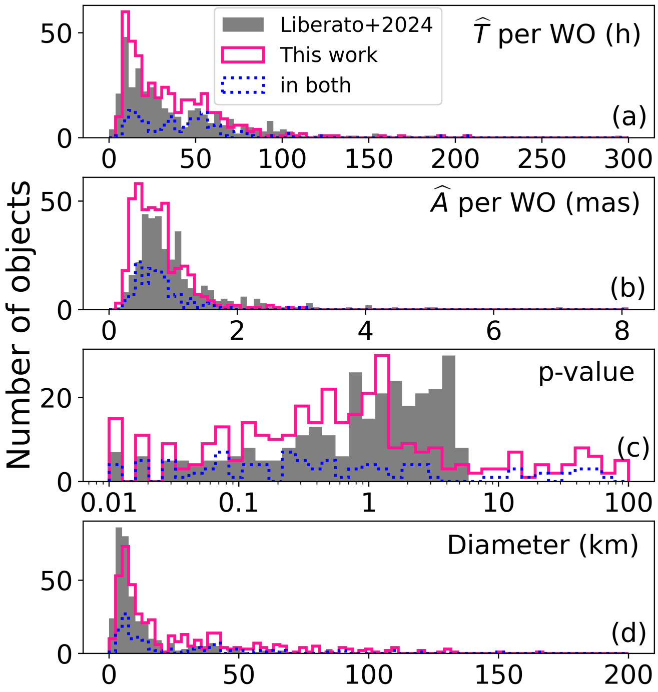

Fig. 11. Distribution of period (a), amplitude (b), and p-value in % (c) estimates per WO, and of diameters from the literature (d), for the *Gaia* DR3 candidates in L24 (grey bars), for the *Gaia* FPR candidates in this work (solid pink line), and for the results of the present work for the candidates in common between the two list (dotted blue line).

of 99 candidates (∼28%) selected in both *Gaia* DR3 and *Gaia* FPR.

Among the objects selected in L24 but not recovered here are the known binaries (3220) Murayama, (5817) Robertfrazer, and (18301) Konyukhov. For Murayama and Robertfrazer, although the detected periods and amplitudes remain consistent between FPR and DR3, the modified data set led to higher p-values in this work, which reduced the statistical significance of the detections. As a result, the BH procedure does not select them. In the case of Konyukhov, the derived physical parameters do not satisfy the updated physical constraints, and therefore prevented the object from being selected in the final sample. This illustrates that some candidates lie near the detection limits and different approaches can affect their detection, although the approximate 30% overlap between both lists also demonstrates some robustness for stronger detections.

In Fig. 11, we compare the results of the previous and current binary search analyses. We observe overall qualitative agreement, although small periods and amplitudes are slightly more abundant. Panel (b) shows a larger selection of smaller amplitudes due to the new noise model (Sect. 2). Small wobbles may arise either from smaller sizes or from binaries with similar size ratios.

Panel (c) shows that the current p-values distribution is much more concentrated around small values than in L24, where we adopted a simple threshold at 5%. This evidence supports the idea that our new noise model allows for clearer detections, together with the FDR control that selects smaller p-values (Sect. 5). Moreover, the majority of the objects present in both lists also have small p-values, as seen from the dotted blue distribution in Fig. 11 (c). Such objects can be considered to have the strongest wobble signal.

8. Conclusions

In this work, we present the methodological approach and results of a comprehensive search for astrometric binary asteroids in the *Gaia* FPR catalogue. The current method is a substantially improved version of that presented in L24. The main improvements include a dedicated noise model for post-fit residuals consistent with the *Gaia* error model, the identification and de-trending of linear systematics in residuals prior to period searches, and a statistically robust selection framework explicitly controlling the false discovery rate (FDR).

The consistency between our FDR threshold and the fraction of objects that are further selected by physical criteria supports the reliability of our detections relative to L24. The 99 objects common to both methods likely correspond to the strongest candidates identified in L24. The other asteroids selected in L24 but not recovered here should be considered weaker with respect to our current, updated sample.

From *Gaia* FPR we obtain a list of 410 WOs corresponding to 343 binary asteroid candidates. These represent ∼24% of the 1448 initially statistically selected objects and are consistent with the expected fraction of true detections for a 75% FDR threshold. We are thus confident that the detection of a periodic signal in the residuals is very solid for a significant fraction of these asteroids. In some cases, especially for the largest objects in our sample, either a single or a binary object could be compatible with this signal (as discussed in Sect. 7.2).

The recovery of the known binaries illustrates both the strengths and limitations of this method, which is strongly dependent on the data quality, observational geometry, and system configuration. In addition, 45 objects show clear trends in multiple WOs, potentially indicating periods longer than those searched for by our approach.

A significant fraction of our candidates could also be synchronous, as the wobble period is equivalent to the rotation period of the potential primary. They tend to be more frequent above ∼10 km in size, extending into the 'binary desert' found by other techniques. We now have stronger evidence that *Gaia* is capable of extending binary detection to an unexplored domain.

Some of the candidate binaries revealed by *Gaia* overlap with binaries discovered from mutual events observed in light curves, so they are good candidates for photometry. Systematic surveys in the coming years should provide an unprecedented amount of high-quality observations, such as those from LSST (Kurlander et al. 2025; Greenstreet et al. 2026). All of our candidates are also excellent targets for stellar occultations, with coordinated campaigns[12](Lallemand et al. 2026).

Besides these techniques, few studies can be compared with our results. We highlight in particular Ou et al. (2022), who analyse the point spread function of Pan-STARRS1 asteroid observations. From their 2930 suspected binaries, 677 objects have at least one WO in our *Gaia* FPR sample. We select 25 of them (∼3.7%) as binary candidates, including the recently confirmed (720) Bohlinia.

In the future, we intend to apply our revised approach to the fourth Gaia data release, covering double the number of asteroids. The revised astrometry in *Gaia* DR4 should allow us to consolidate and potentially expand our candidate list.

[12] Stellar occultation predictions for our candidates are available in https://gaiamoons.imcce.fr/

Data availability

All the lists presented in this work can be downloaded from the Zenodo repository: https://doi.org/10.5281/zenodo.18675577

Acknowledgements. This work presents results from the European Space Agency (ESA) space mission *Gaia*. *Gaia* data are being processed by the *Gaia* Data Processing and Analysis Consortium (DPAC). Funding for the DPAC is provided by national institutions, in particular, the institutions participating in the *Gaia* Multilateral Agreement (MLA). The *Gaia* mission website is https://www.cosmos.esa.int/Gaia. The *Gaia* archive website is https://archives.esac.esa.int/Gaia. This work was supported by the project GaiaMoons of the Agence Nationale de Recherche (France), grant ANR-22-CE49-0002. It was financed in part by the French Programme National de Planetologie, and by the BQR program of Observatoire de la Côte d'Azur. The authors acknowledge the support by the French National program SUN, project TENET. We made use of the software products: SsODNet VO service of LTE, Observatoire de Paris (Berthier et al. 2023); Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration 2018, 2022); Matplotlib (Hunter 2007); Multiprocess package (McKerns & Aivazis 2010; McKerns et al. 2012). The authors also thank the valuable contributions of Federica Spoto, Dagmara Oszkiewicz and Aurelie Duchamps.

References

Astropy Collaboration (Price-Whelan, A. M. P., et al.) 2018, AJ, 156, 123
Astropy Collaboration (Price-Whelan, A. M., et al.) 2022, ApJ, 935, 167
Benishek, V., Pravec, P., & Pilcher, F. 2023, Central Bureau Electronic Telegrams (CBET) No. 5215
Benishek, V., Pravec, P., Marchini, A., et al. 2024, Central Bureau Electronic Telegrams (CBET) No. 5373
Benishek, V., Pravec, P., Kusnirak, P., et al. 2025, Central Bureau Electronic Telegrams (CBET) No. 5507
Benjamini, Y., & Hochberg, Y. 1995, J. Roy. Statist. Soc. B (Methodological), 57, 289
Berthier, J., Carry, B., Mahlke, M., & Normand, J. 2023, A&A, 671, A151
Bertini, I., Sabolo, W., Gutierrez, P. J., et al. 2012, Planet. Space Sci., 66, 64
Brown, M. E. 2013, ApJ, 778, L34
Brown, M. E., & Suer, T.-A. 2007, IAU Circ., 8812
Carry, B. 2012, Planet. Space Sci., 73, 98
Carry, B., Kaasalainen, M., Leyrat, C., et al. 2010, A&A, 523, A94
Cellino, A., Tanga, P., Muinonen, K., & Mignard, F. 2024, A&A, 687, A277
David, H. A., & Nagaraja, H. N. 2004, Order Statistics (John Wiley & Sons)
David, P., Mignard, F., Hestroffer, D., et al. 2023, A&A, 680, A37
Dell'Oro, A., & Cellino, A. 2012, Planet. Space Sci., 73, 10
DeMeo, F. E., & Carry, B. 2014, Nature, 505, 629
DeMeo, F., Alexander, C., Walsh, K., et al. 2015, Asteroids IV, 1, 13
Drummond, J. D., Merline, W., Carry, B., et al. 2021, Icarus, 358, 114275
Durda, D. D., Bottke Jr, W. F., Enke, B. L., et al. 2004, Icarus, 167, 382
Ďurech, J., & Kaasalainen, M. 2003, A&A, 404, 709
Durech, J., Sidorin, V., & Kaasalainen, M. 2010, A&A, 513, A46
Efron, B. 1979, Ann. Statist., 7, 1
Efron, B. 2010, Significance Testing Algorithms, Institute of Mathematical Statistics Monographs (Cambridge University Press), 30
Fisher, R. A. 1948, Am. Statist., 2, 30
Fuentes-Muñoz, O., Farnocchia, D., Naidu, S. P., & Park, R. S. 2024, AJ, 167, 290
Gaia Collaboration (Spoto, F., et al.) 2018, A&A, 616, A13
Garrison, L. H., Foreman-Mackey, D., hsuan Shih, Y., & Barnett, A. 2024, RNAAS, 8, 250
Gault, D., Nosworthy, P., Nolthenius, R., Bender, K., & Herald, D. 2022, Minor Planet Bull., 49, 3
Gorshanov, D. L., Sokova, I. A., Petrova, S. N., Naumov, K. N., & Aliev, A. K. 2025, Planet. Space Sci., 106217
Greenstreet, S., Li, Z. C., Vavilov, D. E., et al. 2026, ApJ, 996, L33
Heard, N. A., & Rubin-Delanchy, P. 2018, Biometrika, 105, 239
Hestroffer, D., Cellino, A., Tanga, P., et al. 2010, in Dynamics of Small Solar System Bodies and Exoplanets (Springer), 251
Hunter, J. D. 2007, Comput. Sci. Eng., 9, 90
Izidoro, A., Raymond, S. N., Morbidelli, A., & Winter, O. C. 2015, MNRAS, 453, 3619
Johnston, W. R. 2025, Asteroids with Satellites, www.johnstonsarchive.net/astro/asteroidmoons.html, accessed: September, 2025
Kaasalainen, M., & Tanga, P. 2004, A&A, 416, 367
Kleine, T., Münker, C., Mezger, K., & Palme, H. 2002, Nature, 418, 952
Kurlander, J. A., Bernardinelli, P. H., Schwamb, M. E., et al. 2025, AJ, 170, 99
Lallemand, R., Desmars, J., Sicardy, B., et al. 2026, A&A, in press [arXiv:2606.13353]
Liberato, L., Tanga, P., Mary, D., et al. 2024, A&A, 688, A50
Liberato, L., Tanga, P., Mary, D., et al. 2026, Zenodo repository, https://doi.org/10.5281/zenodo.18675577
Lindegren, L., Klioner, S., Hernández, J., et al. 2021, A&A, 649, A2
Liu, Z., Hestroffer, D., Desmars, J., & David, P. 2024, A&A, 688, L23
Lomb, N. R. 1976, Astrophys. Space Sci., 39, 447
Margot, J.-L., Pravec, P., Taylor, P., Carry, B., & Jacobson, S. 2015, Asteroids IV, 355, 373
McKerns, M., & Aivazis, M. 2010, http://uqfoundation.github.io/project/pathos
McKerns, M. M., Strand, L., Sullivan, T., Fang, A., & Aivazis, M. A. 2012, arXiv e-prints [arXiv:1202.1056]
Merline, W. J., Close, L. M., Tamblyn, P. M., et al. 2003, IAU Circ., 8075
Milani, A., & Gronchi, G. 2010, Theory of Orbit Determination (Cambridge University Press)
Monteiro, F., Oey, J., Pereira, W., et al. 2024, in European Planetary Science Congress, EPSC2024-156
Morbidelli, A., Walsh, K. J., O'Brien, D. P., Minton, D. A., & Bottke, W. F. 2015, in Asteroids IV (University of Arizona Press)
Ou, J., Baranec, C., & Bus, S. J. 2022, Planet. Sci. J., 3, 169
Pozzi, F., Di Matteo, T., & Aste, T. 2012, Eur. Phys. J. B, 85, 175
Pravec, P., & Harris, A. W. 2007, Icarus, 190, 250
Pravec, P., & Scheirich, P. 2012, Planet. Space Sci., 73, 56
Pray, D., Pravec, P., Hornoch, K., et al. 2015, Central Bureau Electronic Telegrams (CBET) No. 4157
Sato, I. 2013, Spaceguard Res., 5, 23
Scargle, J. D. 1982, ApJ, 263, 835
Scheeres, D., Britt, D., Carry, B., & Holsapple, K. 2015, Asteroids IV, 745766, 745
Sierks, H., Lamy, P., Barbieri, C., et al. 2011, Science, 334, 487
Spoto, F., Tanga, P., Mignard, F., et al. 2018, A&A, 616, A13
Tanga, P., Pauwels, T., Mignard, F., et al. 2023, A&A, 674, A12
Tippett, L. 1931, The Methods of STATISTICS (London, Williams and Norgate), 1952
VanderPlas, J. T. 2018, ApJSS, 236, 16
Vovk, V., & Wang, R. 2020, Biometrika, 107, 791
Zechmeister, M., & Kürster, M. 2009, A&A, 496, 577

Appendix A: Algorithm for confidence interval estimation

We present here the pseudo-code that summarises the algorithm used in this work to estimate the confidence intervals for the period ($\mathcal{I}_T$) and amplitude ($\mathcal{I}_A$) measured from the wobble detected in each WO, as described in more detail in the text of Section 3.1.

The estimated parameter of the wobble $\widehat{f}$ and $\widehat{A}$, along with the residuals' sample, are the inputs. The variable M indicates the number of Monte-Carlo simulations, while $\mu^{(i)}$ is the systematic offset and $\mathbf{n}^{(i)}$ is the random noise, added to each of the M time series simulated (steps 2 to 5).

In steps 12 to 14, the notation $X_{(m)}$ denotes the mth order statistics of the data distribution X (i.e. the mth quantile of X), and ($\lfloor x \rfloor$) the nearest integer smaller than or equal to x. To add robustness to the CI, we compute the distances between two quantiles and the median (step 15) and use the largest one to compute a symmetric interval around $\widehat{A}$.

Algorithm 1: Estimation of confidence intervals.

Inputs : $\boldsymbol{t} := [t_1, \cdots, t_K]^\top$: residuals' epochs
$\boldsymbol{\sigma} := [\sigma_1, \cdots, \sigma_K]^\top$: K std of random error on the residuals
$\widehat{A}, \widehat{f}$ (and $\widehat{T} = 1/\widehat{f}$) : parameters of the estimated wobble
$\nu := [\nu_{\min}, \cdots, \nu_{\max}]^\top$: vector of frequency search for GLSP
M: number of MC realisations
Output : $\mathcal{I}_A$ and $\mathcal{I}_T$: the 95% CI for A and T

1 **for** $i = 1, \cdots, M$ **do**
2 Draw random phase: $\varphi^{(i)} \sim \mathcal{U}_{[0,2\pi]}$;
3 Draw random offset: $\mu^{(i)} \sim \mathcal{L}(0.02, 0.19)$ and constant offset vector $\boldsymbol{\mu}^{(i)} := [\mu^{(i)}, \cdots, \mu^{(i)}]^\top$;
4 Draw random noise vector: $\mathbf{n}^{(i)} \sim \mathcal{N}(\mathbf{0}, \mathrm{diag}(\boldsymbol{\sigma}))$;
5 Generate time series: $\mathbf{y}^{(i)} = \mu^{(i)} + \widehat{A}\sin(2\pi\widehat{f}\,\mathbf{t} + \varphi^{(i)}) + \mathbf{n}^{(i)}$
6 Compute GLSP: $\mathcal{P}^{(i)}(\nu; \mathbf{y}^{(i)}, \boldsymbol{\sigma}, \mathbf{t})$;
7 $\widehat{f}^{(i)} \leftarrow \arg\max_\nu \mathcal{P}^{(i)}(\nu)$;
8 $\widehat{T}^{(i)} \leftarrow 1/\widehat{f}^{(i)}$;
9 Compute $\widehat{A}^{(i)}$ and $\widehat{\varphi}^{(i)}$ from WLS fit of a sinusoid with frequency $\widehat{f}^{(i)}$ to $\mathbf{y}^{(i)}$ (Eq. (C.6)).
10 **end**
11 Sort $\{\widehat{A}^{(i)}\}_{i=1}^M$ and $\{\widehat{T}^{(i)}\}_{i=1}^M$ in increasing order;
12 $q^A_{2.5} \leftarrow \widehat{A}^{(i)}_{(\lfloor 0.025\times M\rfloor)}$;
13 $q^A_{50} \leftarrow \widehat{A}^{(i)}_{(\lfloor 0.5\times M\rfloor)}$;
14 $q^A_{97.5} \leftarrow \widehat{A}^{(i)}_{(\lfloor 0.975\times M\rfloor)}$;
15 $q^{A\star} := \max\{q^A_{97.5} - q^A_{50}, q^A_{50} - q^A_{2.5}\}$;
16 Compute the 95% CI for $\widehat{A}$: $\mathcal{I}_A \leftarrow [\widehat{A} - q^{A\star}, \widehat{A} + q^{A\star}]$;
17 Repeat steps 12–16 applied to $\{\widehat{T}^{(i)}\}_{i=1}^M$ to produce $\mathcal{I}_T$.

Appendix B: Pearson's correlation coefficient

One of the most used methods to quantify the strength and direction of a linear relationship between two variables is Pearson's correlation coefficient. If different realisations of the variables have different reliabilities or uncertainties, a weighted version is more useful since it allows more precise measurements to contribute more strongly, while still preventing noisy points from dominating the correlation.

The conventional weighted Pearson's correlation coefficient (see Sec. 3.1 of Pozzi et al. 2012) can be obtained using the following equation:

$$c(\mathbf{t}, \mathbf{y}) = \frac{\sum_{k=1}^{K} w_k\,(t_k - m_t)(y_k - m_y)}{\sqrt{\left(\sum_{k=1}^{K} w_k (t_k - m_t)^2\right)\left(\sum_{k=1}^{K} w_k (y_k - m_y)^2\right)}} \tag{B.1}$$

with:

$$w_k := \frac{1}{\sigma_k^2}, \qquad m_t := \frac{\sum_{k=1}^K w_k t_k}{\sum_{k=1}^K w_k}, \qquad m_y := \frac{\sum_{k=1}^K w_k y_k}{\sum_{k=1}^K w_k}, \tag{B.2}$$

where the weights w_k are the inverse noise variances (see Sect. 2), $\mathbf{t} = [t_1, \cdots, t_K]^\top$ are the transit averaged epochs in one WO and $\mathbf{y} = [y_1, \cdots, y_K]^\top$ the corresponding residuals.

Appendix C: GLSP with general linear model

Consider a general linear model for the time series of the residuals per transit :

$$\mathbf{y} = \mathbf{M}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{C.1}$$

with $\mathbf{M}$ a model matrix, $\boldsymbol{\beta}$ the coefficients associated with each column of $\mathbf{M}$, and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ the covariance matrix of the noise. For instance, for a "sinusoid + constant model", the matrix is:

$$\mathbf{M}(\nu) := [\mathbf{1},\ \mathbf{c}(\nu),\ \mathbf{s}(\nu)] \tag{C.2}$$

with

$$\mathbf{1} := [1, \cdots, 1]^\top, \tag{C.3}$$

$$\mathbf{c}(\nu) := [\cos(2\pi\nu t_1), \cdots, \cos(2\pi\nu t_K)]^\top, \tag{C.4}$$

$$\mathbf{s}(\nu) := [\sin(2\pi\nu t_1), \cdots, \sin(2\pi\nu t_K)]^\top. \tag{C.5}$$

In case the noise is assumed uncorrelated (as this is the case for the K *Gaia* residuals within a WO), but with a different variance σ_i^2 on each sample, then the matrix $\boldsymbol{\Sigma}$ is diagonal with diagonal diag; $\boldsymbol{\Sigma} = [\sigma_1^2, \cdots, \sigma_N^2]^\top$. The Maximum Likelihood Estimate of $\boldsymbol{\beta}$ for model (C.1) is also the solution of the weighted least squares problem:

$$\widehat{\boldsymbol{\beta}} := \arg\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{M}\boldsymbol{\beta}\|^2_{\boldsymbol{\Sigma}} = (\mathbf{M}^\top\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1}\mathbf{M}^\top\boldsymbol{\Sigma}^{-1}\mathbf{y} \tag{C.6}$$

so that the fitted model is

$$\widehat{\mathbf{y}} := \mathbf{M}\widehat{\boldsymbol{\beta}} = \mathbf{M}(\mathbf{M}^\top\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1}\mathbf{M}^\top\boldsymbol{\Sigma}^{-1}\mathbf{y}. \tag{C.7}$$

and the error is

$$\widehat{\mathbf{e}} := \mathbf{y} - \widehat{\mathbf{y}} = \mathbf{M}(\mathbf{M}^\top\boldsymbol{\Sigma}^{-1}\mathbf{M})^{-1}\mathbf{M}^\top\boldsymbol{\Sigma}^{-1}\mathbf{y}. \tag{C.8}$$

To compare two models, say $\mathbf{M}_1$ and $\mathbf{M}_2$, a standard approach is to compare the corresponding residual sum of squares $\|\widehat{\mathbf{e}}_1\|^2$ and $\|\widehat{\mathbf{e}}_2\|^2$, where $\widehat{\mathbf{e}}_1$ (resp. $\widehat{\mathbf{e}}_2$) are computed by plugging

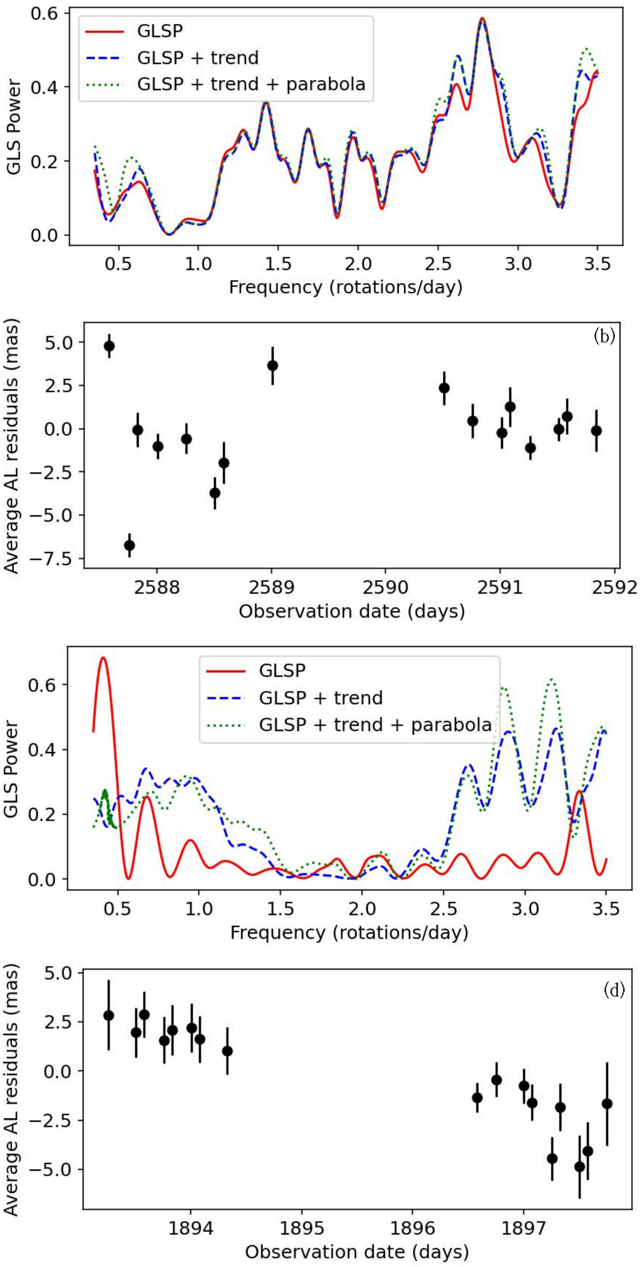

Fig. C.1. Comparison between classical GLSP (red solid) and GLSP including constant + linear dashed blue) and constant + linear + quadratic (dotted green) for two different samples of residuals explored. Plots (a) and (c) show the periodograms for samples (b) and (d), respectively.

$\mathbf{M}_1$ (resp. $\mathbf{M}_2$) in place of $\mathbf{M}$ in Eq. (C.8). A score s can then be computed as:

$$s := \frac{\|\widehat{\mathbf{e}}_1\|^2 - \|\widehat{\mathbf{e}}_2\|^2}{\|\widehat{\mathbf{e}}_1\|^2}. \tag{C.9}$$

When $\mathbf{M}_1 = \mathbf{1}$ and $\mathbf{M}_2 = \mathbf{M}_2(\nu) = [\mathbf{1},\ \mathbf{c}(\nu),\ \mathbf{s}(\nu)]$ as in Eq. C.2, the resulting score $s = s(\nu)$ is the classical GLSP $\mathcal{P}(\nu)$ (q.(4) of Zechmeister & Kürster 2009). With this description, it is straightforward to generalise this approach by including, for instance, a linear trend in the model,

in which case

$$\mathbf{M}_1 = [\mathbf{1},\ \mathbf{t}] \tag{C.10}$$

$$\mathbf{M}_2 = \mathbf{M}_2(\nu) = [\mathbf{1},\ \mathbf{t},\ \mathbf{c}(\nu),\ \mathbf{s}(\nu)] \tag{C.11}$$

or to account also for a quadratic trend, in which case

$$\mathbf{M}_1 = [\mathbf{1},\ \mathbf{t},\ \mathbf{t}^2] \tag{C.12}$$

$$\mathbf{M}_2 = \mathbf{M}_2(\nu) = [\mathbf{1},\ \mathbf{t},\ \mathbf{t}^2,\ \mathbf{c}(\nu),\ \mathbf{s}(\nu)] \tag{C.13}$$

where $\mathbf{t}^2 := [t_1^2, \cdots, t_N^2]^\top$. The columns can be normalised to improve numerical stability.

In Fig. C.1, we can see the results applied to the cases with and without a trend in the residuals. The periodograms shown in panel (a), obtained from the sample in panel (b), represent a typical example: in most cases, the frequency search results are comparable for the three models of the periodograms tested: a simple constant model, a model including a linear trend, and a model including linear and quadratic trends.

However, when the residuals present some trendy features, as in panel (d), we notice that the different periodograms provide different results, as shown in panel (c). Here, the largest peak that occurs at a low frequency for the classical GLSP is caused by the decreasing trend visible in panel (d). This is not the case for the two other periodograms, which are trend-insensitive. This makes it possible to detect the presence of potential oscillations at higher frequencies.

Appendix D: Estimation performances at small amplitudes

Figure D.1 illustrates the performances when estimating wobbles of amplitudes smaller than 1.3 mas. For each input amplitude value we obtain an estimated amplitude (pink dots) from the 50 quantile ($q_{50}^{\widehat{A}}$) and a corresponding confidence interval $\mathcal{I}_A$= [$q_{2.5}^{\widehat{A}}$,$q_{97.5}^{\widehat{A}}$], shown as the grey error bars.

For a noisy sinusoidal signal, compatible with the *Gaia* FPR data set, where the nominal amplitude is A= 0.8 mas is shown by the blue solid line, the estimated amplitude $\widehat{A}$ can have any value in a confidence interval between $q_{2.5}^{\widehat{A}}$ ~0.61 and $q_{97.5}^{\widehat{A}}$ ~1.02 mas, shown as the light blue error bar.

We do not know the true amplitude of the wobble signatures in the *Gaia* astrometric data. So, in this example, we consider that the extreme cases of the confidence interval associated with $\widehat{A}$ are the input signal amplitude. If the signal is detected with an amplitude $\widehat{A}_{min}$=$q_{2.5}^{\widehat{A}}$ represented by the green square, the corresponding confidence interval is delimited as shown by the green horizontal dashed lines.

The same observation can be made for the upper limit of the blue confidence interval where $\widehat{A}_{max}$= $q_{97.5}^{\widehat{A}}$, shown as the orange square, and the estimated confidence interval is delimited by the orange dotted lines. The true amplitude value A is at the edges of the green and orange confidence intervals but still within both of them, showing that even when the true amplitude is unknown the method was capable of estimating confidence intervals that contain the true value.

However, for amplitudes lower than 0.5 mas, the confidence intervals tend to be smaller and more asymmetrical, the 50% quantiles tend to overestimate the amplitudes, and the confidence interval does not necessarily contain the true amplitude, explaining the larger false coverage rate obtained for amplitudes smaller than 1 mas, as observed in Fig. 4.

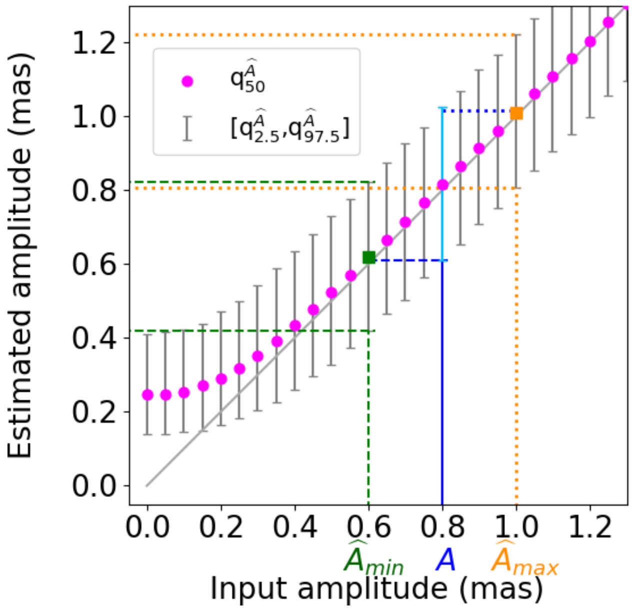

Fig. D.1. Wobble amplitude confidence interval estimation for a range of small amplitudes. The pink dots represent the quantiles $q_{50}^{\widehat{A}}$ estimated from the respective input amplitudes, while the error bars indicate the estimated confidence intervals $\mathcal{I}_A = [q_{2.5}^{\widehat{A}}, q_{97.5}^{\widehat{A}}]$. The coloured dotted and dashed lines show one example of a signal with true amplitude $A = 0.8$ mas, where $\mathcal{I}_A = [0.61, 1.02]$ mas, and where these limits are considered to be input values ($\widehat{A}_{min}$ and $\widehat{A}_{max}$) from a signal with unknown amplitude (see text).