Selection Effects in Wide Binaries (Postprint)

Authors: Liu Di, Zhang Peng, Tian Haijun, Liu Gaochao, Yang Xiangming, Xiong Zhuang

Date: 2025-10-10T00:00:00+00:00

Abstract

Wide binary systems consist of two widely separated celestial bodies orbiting a common center of mass, with their physical projected separation being the most direct observable quantity. Using high-purity wide binaries within 1 kpc of the solar neighborhood as a test sample, an anomalous γ-value phenomenon caused by selection effects was discovered (showing increasing deviation from the normal value $\gamma = -1.5$ across different distance shells). To address this problem, the strength of selection effects in each distance shell was quantified, the selection function was recalibrated accordingly, and three different power-law models were employed to perform parameter fitting on the intrinsic distribution of their projected separations through Bayesian statistics and Markov Chain Monte Carlo methods. The results demonstrate that after applying the improved selection function, the power-law index γ of the three mathematical models in each distance shell becomes stable and normal. These results provide a reference basis for correcting model parameters of the projected separation distribution of wide binaries affected by selection effects, which is of significant importance for understanding the formation and evolution of wide binaries, as well as for other related fields.

Full Text

Preamble

Vol. 43, No. 3

September 2025

PROGRESS IN ASTRONOMY

Study on Selection Effects of Wide Binary Stars

LIU Di[1], ZHANG Peng[1],[2], TIAN Haijun[3], LIU Gaochao[1],[2], YANG Xiangming[1], XIONG Zhuang[1]

(1. College of Science, China Three Gorges University, Yichang 443002, China;
2. Center for Astronomy and Space Sciences, China Three Gorges University, Yichang 443002, China;
3. College of Science, Hangzhou Dianzi University, Hangzhou 310018, China)

Abstract

Wide binary systems consist of two widely separated celestial bodies orbiting a common barycenter, with projected physical separation being the most direct observable quantity. Using a high-purity sample of wide binaries within 1 kpc of the solar neighborhood as a test case, we identify anomalous γ values caused by selection effects that increasingly deviate from the expected value of $\gamma = -1.5$ across different distance shells. To address this issue, we quantify the strength of selection effects in each distance shell, recalibrate the selection function accordingly, and employ three distinct power-law models to fit the intrinsic distribution of projected separations using Bayesian statistics and Markov Chain Monte Carlo methods. Our results demonstrate that after applying the improved selection function, the power-law indices γ from all three mathematical models stabilize and approach normal values in each distance shell. These findings provide a reference for correcting model parameters of projected separation distributions affected by selection effects, offering important insights for understanding the formation and evolution of wide binaries and related fields.

Keywords: binary stars; wide binary systems; selection effects; selection function

1 Introduction

Binary star systems consist of two celestial bodies born contemporaneously with identical initial chemical compositions, bound by mutual gravitational forces that cause them to orbit their common center of mass. Approximately half of Sun-like stars belong to binary systems with orbital periods ranging from decades to hundreds of millions of years. A small fraction of these binaries have small separations (s 10 AU), where mass exchange between components leads to complex evolutionary pathways, while most binaries have larger separations reaching up to 2×10^4 AU (0.1 pc), allowing components to evolve independently. Systems with widely separated components experiencing only weak mutual interactions are known as wide binaries.

The formation of binary stars remains a central topic in astronomical research, with different theoretical models favoring specific formation mechanisms depending on projected separation, though no consensus has been reached. Close binaries (s 100 AU) likely form through fragmentation of a common circumstellar disk (disk fragmentation mechanism). Wide binary formation is more complex: systems with separations below 10^3 AU may form via turbulent fragmentation of molecular cloud cores, while those with larger separations (s 10^3 AU or 10^4 AU) may originate from random pairings of dissolving cluster members, dynamical evolution of unstable triple systems, or chance encounters between neighboring stellar cores.

Wide binaries represent one of the simplest, smallest, and most fragile celestial systems. Studying their projected separation distribution enables probing variations in the Galactic gravitational potential. With extremely low binding energy, wide binary orbits are highly vulnerable to disruption by external gravitational perturbations from other stars, giant molecular clouds, and Galactic tides, as well as internal evolutionary processes. Since disruption probability increases with separation, the separation distribution of an evolved wide binary population should exhibit a break (characterized by the power-law model parameter s_{br}, the critical break value of projected separation). Conversely, the location of this break can be used to assess differences in gravitational potential across regions of the host galaxy. Thus, wide binaries serve as powerful probes of Galactic gravitational potential on small scales. Analysis of halo wide binary separation distributions can address questions regarding the existence and properties of massive compact halo objects (MACHOs). For instance, Yoo et al. predicted a break in the angular separation power-law distribution for halo wide binaries in the presence of MACHOs, with the break location determined by MACHO mass and density. Tian et al. used total tangential velocity (V_ ,tot) as a tracer of age and metallicity to divide wide binary samples into young disk-like, old halo-like, and intermediate subsamples, revealing distinct characteristics in their projected separation breaks. In the Milky Way, wide binary systems cannot maintain stable gravitational binding once their separation exceeds the threshold sensitive to Galactic tidal fields. As extremely weakly bound systems, wide binaries also provide opportunities for testing modified gravity theories in weak-field environments and offer an ideal probe for investigating dark matter mass distribution in ultra-faint dwarf galaxies dominated by dark matter, enabling tests of various dark matter models on small galactic scales. Additionally, wide binary disruption may stem from internal stellar evolution, where the "kick" imparted during supernova explosions forming neutron stars or black holes can destabilize these fragile systems.

Astronomers typically study binary projected separation distributions using the power-law model p(s) s^γ with γ -1.5. Since projected separation (s) relates linearly to orbital semi-major axis (a) as s 0.978a for wide binaries, they share similar statistical properties. Lépine and Bongiorno found that wide binary projected separation distributions follow f(s)ds s^{-} ($l = 1.6^{+0\cdot 1}{}_{0\cdot 1}$, s 4000 AU) based on Hipparcos data. Chanamé and Gould fitted angular separation distributions of disk-like and halo-like wide binaries, obtaining best-fit power-law indices of $-1.67^{+0\cdot 07}{}_{0\cdot 10}$ and $-1.54^{+0\cdot 09}{}_{0\cdot 10}$ in the range $3.5 < \Delta < 900$. El-Badry and Rix characterized main sequence-main sequence (MS-MS) wide binary projected separation distributions with γ -1.6. Tian et al. employed a smoothly broken power-law model to fit intrinsic projected separation distributions of disk-like, halo-like, and intermediate wide binary samples, obtaining power-law indices γ of $-1.51^{+0\cdot 03}{}_{0\cdot 07}$, $-1.55^{+0\cdot 10}{}_{0\cdot 03}$, $-1.56^{+0\cdot 03}{}_{0\cdot 04}$, and $-1.55^{+0\cdot 05}{}_{0\cdot 06}$.

Due to limitations in Gaia' s angular resolution and binary search algorithms, our wide binary catalog is incomplete. El-Badry and Rix demonstrated that incompleteness primarily arises from blending of the two stellar components, particularly severe when brightness differences are large or angular separations are small. Using a high-purity wide binary sample constructed from the El-Badry 2021 binary catalog, we fit the intrinsic projected separation distribution of subsamples in different distance shells using three mathematical models: single power law (SPL), double power law (DPL), and smoothly broken power law (SBPL). We identified anomalous power-law index γ values caused by selection effects, which bias the inferred intrinsic distribution away from reality. To address this, we recalibrated the selection function in each distance shell and applied it to the fitting process to obtain more accurate intrinsic distributions.

Section 2 describes the additional selection criteria used to construct the high-purity wide binary sample. Section 3 discusses the selection effects affecting the binary sample. Section 4 explains how we improved the selection function to infer the intrinsic separation distribution. Section 5 presents fitting results from the three power-law models and compares power-law index γ (or γ_1) values before and after selection function calibration. Finally, we summarize our findings and outline future research directions.

2 Data and Selection Process

Our dataset utilizes the catalog of wide binaries identified in Gaia Early Data Release 3 (Gaia EDR3) from the Gaia Space Telescope. El-Badry et al. first used this catalog to construct an initial sample containing 1,817,594 wide binary candidates within approximately 1 kpc of the Sun, with physical projected separations ranging from tens of AU to 1 pc.

The wide binary catalog contains numerous false binaries—pairs artificially created by chance alignments during the binary identification process. Merged galaxies from past interactions can maintain coherent motions for several Galactic dynamical timescales, and many sources from the initial binary sample can be cross-identified with known open clusters and moving groups, or have companions selected from nearby cluster members. Although these pairs formally satisfy binary selection criteria, most are not genuine binaries, resulting in high contamination rates. Therefore, we adopted methods similar to El-Badry et al. and Tian et al. to exclude these contaminants and performed additional selection procedures to obtain a purer wide binary sample. We summarize the selection steps below.

2.1 Selection Criteria

First, we applied stricter constraints on astrometric and photometric parameters for both components, eliminating numerous false sources in non-physical regions of the color-magnitude diagram along with a few genuine sources. (1) Both stars in candidate pairs must have five-parameter astrometric solutions and successful G_{BP}−G_{RP} color index measurements, with low astrometric excess noise satisfying ^{2}/(−5) < 1.2 × max(1, exp(−0.2(G−19.5))), where ^{2} and correspond to parameters astrometric_{{{chi2}}*{{al}}} and astrometric_n*{{{good}}*{{obs}}}*{al} in the Gaia archive. (2) Since G_{BP} and G_{RP} magnitudes are derived from low-resolution spectral integration with broader point spread functions than the G band, we assessed contamination from nearby sources by comparing BP/RP total fluxes (G_{BP}, G_{RP}) and excess factors (phot_{{{bp}}*{{rp}}}*{{{excess}}*{{factor}}}). To ensure minimal photometric contamination, both stars must satisfy 1.0 + 0.015(G*{BP} − G_{RP})^2 < phot_{{{bp}}*{{rp}}}*{{{excess}}*{{factor}}}* < 1.3 + 0.06(G*{BP} − G_{RP})^2. (3) Both components must have high signal-to-noise photometry: uncertainties in mean G-band flux less than 2% (phot_g_{{{mean}}*{{flux}}}*{{{over}}*{{error}}} > 50) for both stars, with uncertainties in mean BP and RP fluxes less than 5% and 10% respectively (phot*{{{bp}}*{{mean}}}*{{{flux}}*{{over}}}*{error} > 20 (>10) and phot_{{{rp}}*{{mean}}}*{{{flux}}*{{over}}}*{error} > 20 (>10) for primary (secondary) stars.

Second, we defined nearby binaries as systems within 1° on the sky, $±3*mas* · *yr*^{-1}$ in proper motion coordinates, and $±$5 pc in parallax (1/). We calculated the number of nearby binaries (N) for each candidate in position-parallax-proper motion space, accepting only candidates with N < 2. This process removed 565,668 systems associated with clusters, moving groups, and higher-order multiples. For further purity, we imposed Δ ≤ Δ _{or} + 1.0σ_Δ and *σΔ < 0.12, where Δ is the proper motion difference, Δ* {or} is the maximum proper motion difference sustainable for a circular orbit of a 5 M binary, and σ_Δ is the uncertainty in proper motion difference.

Finally, we defined "main sequence (MS) stars" as members satisfying M_G < 2.75(G_{BP} − G_{RP}) + 5.75 and "white dwarfs (WD)"as objects satisfying M_G < 3.25(G_{BP} − G_{RP}) + 9.63, where absolute magnitude M_G = G + 5 log(/mas) − 10. To eliminate effects of internal orbital evolution, we further excluded WD-containing binaries (WD-MS and WD-WD) from the high-purity sample, retaining only longer-lived, more stable MS-MS binaries for model fitting.

These selection criteria yielded a high-purity catalog of 144,458 wide binary candidates, free of any resolved higher-order multiples.

2.2 Overview and Characteristics of the Wide Binary Catalog

A total of 144,458 wide binary candidates passed our selection criteria: 142,995 MS-MS pairs, 1,412 WD-MS pairs, and 51 WD-WD pairs. We note that these classifications are not strictly defined, particularly for WD-MS pairs where some boundary cases may be misclassified. The MS category primarily serves to exclude WDs, though it includes small fractions of giants, subgiants, and pre-main-sequence stars. The BP/RP spectral window measures 2.1 × 3.5 , so most sources with companions within 2 and relatively bright secondaries lack BP/RP measurements and cannot be classified as WD or MS using Gaia data alone; we excluded such candidates. Assuming equal distances for both components, we calculated absolute magnitudes M_G using the brighter primary’ s parallax (typically more precise). For extinction, we assigned values to each component using the Galactic extinction catalog constructed by Leike and Enßlin.

[Figure 1: see original paper] shows the sky distribution of our high-purity sample in Galactic longitude-latitude coordinates (l—b). Gaia’ s scanning pattern leaves a clear imprint, as sources with well-constrained astrometric and photometric parameters are preferentially located in frequently visited sky regions, resulting in non-uniform distribution in the l—b plane. Binary candidate density is also higher near the Galactic plane, a phenomenon more pronounced before removing sources with numerous phase-space neighbors.

[Figure 2: see original paper] presents color-magnitude diagrams (CMDs) for primary and secondary stars in the high-purity sample. Both CMDs clearly show secondary sequences above the main sequence, indicating the presence of hierarchical triple systems containing unresolved close binaries.

[Figure 3: see original paper] displays the distributions of two fundamental parameters: primary apparent magnitude and magnitude difference. The median G magnitudes are 13.9 mag for MS-MS binaries and 15.2 mag for MS stars, with 67.4% and 67.8% of the distributions falling in ranges 11.5-15.9 mag and 12.5-16.9 mag, respectively. This reflects our stringent requirements on parallax, proper motion, and photometric measurements, which favor relatively bright stars. Apparent magnitude distributions vary significantly across distance shell subsamples (Figure 3a).

Figure 3b shows the magnitude difference distribution ($\Delta G = |G_1 - G_2|$). The median ΔG is 1.71 mag, with 67.4% of binaries distributed between 11.5-15.9 mag. Substantial differences exist among distance shell subsamples, which is crucial for inferring intrinsic separation distributions because Gaia’ s sensitivity to detect companions at a given angular separation varies with magnitude difference. This demonstrates that subsamples in different distance shells experience varying degrees of selection effects at small separations, motivating our distance-dependent calibration approach in Section 4.2.

[Figure 4: see original paper] presents the number density distribution of wide binaries as a function of distance. The high-purity sample (Figure 4b), derived from the initial catalog (Figure 4a) through our selection criteria, has a median distance of $d = 316.3$ pc, with 67.8% of binaries distributed between 151.6 d 599.7 pc and approximately 83.8% within 600 pc. We binned both initial and high-purity samples into 10 distance shells with 100 pc width to examine distance-number density variations, while further subdividing each shell subsample with 10 pc bins.

Due to selection effects, both distributions show decreasing number density with increasing distance, with the high-purity sample exhibiting a steeper decline because of our stricter selection criteria. Figures 4c–4l show distance-number density distributions for high-purity subsamples in each shell, which remain nearly constant and approximate uniform distributions in log space across each shell. The $0 < d < 100$ pc subsample shows the highest values and largest decrease in distance-number density.

2.3 False Binaries and Contamination Rate

Before the Gaia mission, precise astrometric data were unavailable for constructing pure binary catalogs. The first systematic binary catalog was built under the assumption that close pairs were false binaries—a hypothesis proven incorrect two decades later. Binary search methods evolved from low-dimensional phase-space searches using position or proper motion to high-dimensional searches incorporating parallax and radial velocity. Gaia data releases have enabled construction of high-purity, large-sample binary catalogs. The contamination rate from false binaries depends on multiple factors; for wide binaries with small parallaxes and proper motions, false binaries cannot be completely eliminated. Contamination increases with projected separation, particularly dominating when s > 30,000 AU. El-Badry et al. constructed a seven-dimensional parameter space including angular separation, distance, parallax difference uncertainty, local sky density, tangential velocity, and signal-to-noise ratios of parallax and proper motion differences, using kernel density estimation to compute contamination rate R. Our high-purity sample is limited to $R < 0.1$.

We also quantified contamination indirectly by comparing absolute radial velocity differences (ΔRV) between components. Among 5,579 pairs with accurate radial velocity measurements ($\sigma\{RV\} < 3\ km \cdot s^{-1}$), *we identified potential contaminants as those satisfying* $\Delta RV/\sigma\{\mathrm{RV}\} > 5$ and $\Delta\mathrm{RV} > 10\ \mathrm{km \cdot s^{-1}}$. [Figure 5: see original paper] shows the primary-secondary RV comparison and ΔRV versus projected separation. Most binaries with large ΔRV also have above-average σ_ΔRV, with only 16 pairs meeting false binary criteria, implying a contamination rate of ~0.15%. Most contaminants are distributed in the range 3.0   log(s/AU)   4.2. However, the true contamination rate for the full sample is higher. Radial velocity provides contamination estimates only for systems where both components are relatively bright.

2.4 Observable Separation Distribution

[Figure 6: see original paper] shows the distributions of angular separation   and physical projected separation s. Due to Gaia' s angular resolution limit, no binaries appear when   1.5 . With relaxed selection criteria, the separation distribution shows a bimodal pattern with peaks at log( )   0.5 and log( )   2.5 in angular separation, corresponding to log(s)   3 and log(s)   5 in projected separation (black histogram). However, the large-separation peak arises entirely from false binaries rather than genuine pairs, shifting to larger separations and eventually disappearing under stricter criteria. Dhital et al. and Oelkers et al. also found bimodal projected separation distributions, suggesting the wide-separation population contains binaries formed through different mechanisms.

For the initial sample, contaminants dominate at large separations (4.2 , s 0.05 pc), with the separation peak caused entirely by false binaries. Consistent with El-Badry & Rix and Tian et al., MS-MS binaries do not follow Öpik' s law (uniform distribution in log projected separation) at any s 500 AU. We divided the high-purity sample into 10 distance shell subsamples. Notably, angular separation distributions for these subsamples show different truncation points at various (decreasing with distance shell), while projected separation distributions exhibit corresponding truncation at s 10^3 AU (increasing with distance shell) and a clear cutoff at s 10^5 AU. This pronounced bimodality is induced by randomly paired false binaries.

3 Selection Effects

Selection effects represent a pervasive challenge in astronomical research, arising from instrumental limitations, data processing methods, and detection techniques that cause certain objects to be more or less easily discovered or studied, potentially biasing results and distorting our understanding of celestial distributions, properties, and evolution. Therefore, considering and calibrating selection effects is crucial.

Selection effects remain widespread and complex in binary star research. Bright stars are more easily detected, while faint companions may be missed, creating a bias toward luminous systems. Wide angular separations are more easily resolved, while small separations can lead to blending, making the sample preferentially select wide binaries. When angular separation is sufficiently large and both stars are bright enough for independent detection, a complete binary system can be identified. However, at small separations, the presence of a bright star makes detecting a faint companion difficult. Thus, sample incompleteness results from combined effects of apparent magnitude, magnitude difference, and angular separation.

Due to Gaia observational limitations and our binary selection criteria, our wide binary catalog is incomplete: it lacks unresolved close binaries and pairs where one component is too faint to pass selection criteria or be detected initially. The absence of binaries with small physical projected separations primarily stems from incompleteness at small angular separations (see [Figure 7: see original paper]).

Selection effect strength varies across distance shells, with incompleteness differing among subsamples (see Figures 3 and 6). Relatively, samples in smaller distance shells have more observables in apparent magnitude G, magnitude difference ΔG, angular separation , and projected separation s than those in larger shells. Additionally, different distance shell subsamples show varying degrees of truncation at small separations (s 10^3 AU) that increase with distance. In summary, selection effects cause sample incompleteness and affect wide binary property distributions in different ways.

[Figure 7: see original paper] shows the distribution of wide binaries in distance-projected separation space ((1/)-s). The typical resolution limit (2) is clearly visible, preventing detection of small-projected-separation binaries at large distances. Consequently, distant samples are dominated by wide-separation pairs. Interestingly, when uniformly dividing the sample into 10 distance shells and fitting their projected separation distributions with the three power-law models (SPL, DPL, SBPL), we find that all power-law indices γ (or γ_1) deviate from normal values and vary systematically with distance shell (see). This anomaly caused by sample incompleteness leads to severe biases in the inferred intrinsic projected separation distribution, making calibration of incompleteness due to missing unresolved close binaries and large brightness differences essential.

4 Improved Method for Correcting Power-Law Indices Affected by Selection Effects

Section 3 established that completeness is severely deficient at small angular separations ($<$ 10), particularly for even smaller . Therefore, before modeling the intrinsic distribution of physical projected separations, we must account for selection effects to compensate for incompleteness at small separations and correct model power-law indices (γ or γ_1) affected by distance. In this section, we adopt the same functional form as the empirical selection function constructed by El-Badry & Rix, but improve it by incorporating distance dependence. We uniformly divide the full distance range 0–1 kpc into 10 shells: 0–100 pc, 100–200 pc, 200–300 pc, 300–400 pc, 400–500 pc, 500–600 pc, 600–700 pc, 700–800 pc, 800–900 pc, and 900–1000 pc. In each shell, we statistically analyze the relationship between angular separation and completeness for different magnitude differences ΔG, then apply these relationships to the likelihood function for fitting projected separation distributions. Finally, we model the intrinsic distribution of physical projected separations using the three power-law models.

4.1 Mathematical Models for Fitting Projected Separation Distributions

Wide binary projected separation distributions often follow specific power-law patterns. In this section, we model these distributions using three mathematical models: SPL, DPL, and SBPL.

4.1.1 SPL This model corresponds to a one-dimensional free parameter space $\mathbf{m} = (\gamma)$ with a single free parameter γ. Despite its limitations, this model can determine the best-fit power-law index γ over a given projected separation range (full or partial), expressed as:

$$\phi(s) = \phi_0 s^{\gamma}$$

4.1.2 DPL This model corresponds to a three-dimensional free parameter space $\mathbf{m} = (\gamma_1, \gamma_2, \log(\text{s\_\{br\}/AU}))$, adding new power-law index γ_2 and break point s_{br}. Here γ_1 and γ_2 represent power-law indices for separations s s_{br} and s s_{br}, respectively:

$$\phi(s) = \phi_0 \begin{cases} s^{\gamma_1}, & s \leq s_{br} \\ s^{\gamma_2}, & s > s_{br} \end{cases}$$

4.1.3 SBPL This model corresponds to a four-dimensional free parameter space $\mathbf{m} = (\gamma_1, \gamma_2, \log(\text{s\_\{br\}/AU}), \Lambda)$. Building upon DPL, it adds a smoothing factor Λ to quantify the steepness or smoothness of the transition between the two power laws:

$$\phi(s) = \phi_0 \left[1 + \left(\frac{s}{s_{br}} \right)^{\Lambda} \right]^{(\gamma_2 - \gamma_1)/\Lambda} s^{\gamma_1}$$

In equations (1)-(3), the coefficient ϕ_0 represents normalization constants.

4.2 Selection Function

Wide binary identification requires both stars to be spatially resolved and satisfy astrometric and photometric selection criteria (Section 2.1). Detection therefore depends on angular separation and flux ratio between components. Large flux ratios increase the likelihood that the secondary is masked or contaminated by the primary's light at fixed angular separation. This effect must be accounted for when inferring intrinsic projected separation distributions. Additionally, detection depends on apparent magnitudes: if either star is too faint, the system may be missed. If undetected binaries share the same intrinsic separation distribution as detected ones, this would not affect inference—requiring the separation distribution to be independent of distance and absolute magnitude. While El-Badry (2018) found no significant distance dependence out to 200 pc, our high-purity sample extends to nearly 1 kpc, necessitating consideration of distance-induced variations in power-law parameters, particularly γ (or γ_1). As established in Section 3, sample incompleteness arises from selection effects related to magnitude difference ΔG and angular separation . To quantitatively study these effects, we adopt the selection function f_ΔG( ) from El-Badry & Rix describing the probability of detecting a companion at given angular separation :

$$f_{\Delta G}(\theta) = \frac{1}{1 + (\theta/\theta_0)^{-\beta}}$$

The selection function depends on angular separation and magnitude difference ΔG (absolute G-band magnitude difference), where $_0$ is the angular separation at which sensitivity drops below unity and β determines the rate of sensitivity decline for $_0$. We first perform statistical analysis separately for the 10 distance shells (see [Figure 8: see original paper]), then estimate optimal $_0$ and β values for different ΔG through interpolation of discrete ΔG values. Note that f_ΔG( ) represents a relative selection function—the ratio of detected binaries at angular separation   to those detected at arbitrarily large separations.

[Figure 8: see original paper] shows completeness versus angular separation   for different magnitude differences ΔG across distance shells. Using a method similar to Arenou et al., we assess Gaia' s photometric sensitivity to companions at given angular separation. For each ΔG interval, we compute the relative fraction of binaries passing selection criteria (Section 2.1) in angular bins of ~0.66 width. Linear interpolation across discrete ΔG values (using interval midpoints) estimates optimal $_0$ and β for the selection function. Generally, smaller ΔG and larger   yield higher completeness.

4.3 Likelihood Function for Fitting Projected Separation Distributions

Consider an intrinsic binary projected separation distribution of the form ϕ(s|**m**) = dP/ds, where **m** represents the model free parameters (the free parameter spaces from Section 4.1). For a set of binaries with projected separations s_i, the likelihood function is:

$$\mathcal{L} = p(\{s_i\}|\mathbf{m}) = \prod_i p(s_i|\mathbf{m})$$

where p(s_i|**m**) is the detection probability for the i-th binary given model parameters **m**, calculated as:

$$p(s_i|\mathbf{m}) = \frac{\phi(s_i|\mathbf{m}) f_{\Delta G}(s_i|d_i)}{\int_{s_{\min}}^{s_{\max}} \phi(s|\mathbf{m}) f_{\Delta G}(s|d_i) ds}$$

The probability p_i is proportional to the likelihood of finding a binary with distance d_i, magnitude difference ΔG, and projected separation s_i in the catalog. The denominator reflects the fraction of detectable binaries at distance d_i and magnitude difference ΔG, qualitatively indicating that at larger distances and magnitude differences, wide-separation binaries are more easily detected than close ones. ϕ(s|**m**) is normalized such that ϕ(s|**m**)ds = 1, and the selection function f_ΔG( ) = f_ΔG(s|d_i) is given by equation (4). We set s_{min} = 10^{-2} AU (small projected separations make the probability approach zero, rendering s_{min} irrelevant) and s_{max} = $10^{5\cdot 20}$ AU (~0.8 pc) due to high contamination beyond this limit.

5 Fitting Results

We uniformly divided the high-purity sample into 10 distance shell subsamples within 0–1 kpc and performed posterior sampling of the three power-law models for physical projected separation using the emcee library, adopting broad, flat priors for all parameters.

This section presents constraints on model parameters for the intrinsic projected separation distribution of high-purity subsamples in different distance shells, visualized using corner plots to show relationships between parameters and their marginal probability distributions. In plots showing power-law index γ (or γ_1) versus distance, "before calibration" and "after calibration" refer to using $_0$ and β values from El-Badry (2018) versus distance-corrected values, respectively.

We present model parameter variations before and after selection function calibration for each power-law model (SPL, DPL, SBPL), focusing on power-law index γ (or γ_1). For clearer visualization, we plot absolute values $|\gamma|$ in some cases.

5.1 SPL

We fitted both the full projected separation range s $10^{5 \cdot 20}$ AU and a restricted range s 10^4 AU for the 10 distance shell subsamples using the SPL model (ϕ(s) = ϕ_0s^γ), obtaining corresponding $|\gamma|$ values ([Figure 9: see original paper]).

[Figure 9: see original paper] shows that $|\gamma|$ increases with distance shell, rising from $1.70^{+0 \cdot 01}\,_{0 \cdot 01}$ (s $< 10^4$ AU) in the $0 < d < 100$ pc shell to $1.99^{+0 \cdot 02}\,_{0 \cdot 03}$ (s $< 10^4$ AU) in the $900 < d < 1000$ pc shell. After applying the calibrated selection function, $|\gamma|$ stabilizes near 1.5 across all shells.

5.2 DPL

Unlike SPL' s single parameter, DPL has three free parameters: $\mathbf{m} = (\gamma_1, \gamma_2$, log(s_{br}/AU)). Before selection function calibration, $|\gamma_1|$ shows a clear increasing trend with distance shell, rising from $1.64^{+0 \cdot 01}\,_{0 \cdot 01}$ in the $0 < d < 100$ pc shell to $1.87^{+0 \cdot 04}\,_{0 \cdot 05}$ in the $900 < d < 1000$ pc shell ([Figure 10: see original paper]). After recalibration, $|\gamma_1|$ fluctuates around 1.5.

Given the presence of break point s_{br}, we extended the projected separation range to s_{max} = $10^{5 \cdot 20}$ AU while acknowledging high contamination at large separations. We fitted the intrinsic distribution using DPL with pre- and post-calibration selection functions ([Figure 11: see original paper]), which shows marginal probability distributions for DPL parameters across distance shells.

5.3 SBPL

SBPL incorporates four free parameters: $\mathbf{m} = (\gamma_1, \gamma_2$, log(s_{br}/AU), Λ), adding smoothing factor Λ to quantify the steepness of the transition between power laws. Using the same separation limits as DPL (s $10^{5 \cdot 20}$ AU), we fitted the intrinsic distribution with pre- and post-calibration selection functions.

With the uncalibrated selection function, $|\gamma_1|$ again increases with distance shell, from $1.64^{+0 \cdot 01}\,_{0 \cdot 01}$ in the $0 < d < 100$ pc shell to $1.87^{+0 \cdot 04}\,_{0 \cdot 05}$ in the

900 < d < 1000 pc shell ([Figure 12: see original paper]). After recalibration, $|\gamma_1|$ fluctuates around 1.35. [Figure 13: see original paper] shows marginal probability distributions for SBPL parameters across distance shells.

5.4 Comparison of Results Before and After Calibration

We recorded best-fit power-law index values for each distance shell and model. Table 1 and numerically show power-law index variations before and after calibration. Post-calibration indices are significantly reduced, and under equivalent conditions satisfy $\gamma\{SBPL\}$ γ{DPL} γ_{SPL}.

Using the 200 < d < 300 pc shell as a test sample after calibration, we obtained marginal probability distributions for parameters from all three models and visualized the intrinsic distribution using best-fit values.

[Figure 14: see original paper] compares observed and model-fitted intrinsic distributions before and after selection effect calibration for the test sample. The black dashed line (SPL with $\gamma = -1.5$) represents the expected distribution. Compared to pre-calibration, post-calibration fits show larger break points s_{br} for DPL and SBPL, with all models approaching the expected distribution at s   s_{br}.

6 Summary and Outlook

Using Gaia EDR3 data, we constructed a high-purity wide binary catalog containing 144,458 pairs within 1 kpc of the Sun and with projected separations s < 1 pc. We investigated selection effects on the intrinsic projected separation distribution (particularly power-law index γ) using three models (SPL, DPL, SBPL). We found that uncalibrated selection functions cause γ to deviate from the expected value (γ -1.5), with deviations increasing with distance shell. To correct this, we proposed distance-dependent calibration of the selection function by uniformly dividing the 0–1 kpc range into 10 shells and statistically analyzing the relationship between angular separation   and completeness for different magnitude differences ΔG in each shell, then applying these relationships to model fitting. Results show that recalibrating the selection function with distance dependence stabilizes γ, providing more reliable theoretical foundations for wide binary formation and evolution.

After applying the distance-calibrated selection function, SPL and DPL yield power-law indices γ that remain stable across all distance shells, fluctuating around -1.5. SBPL produces γ -1.35, slightly lower than SPL and DPL due to the smoothing factor Λ's influence on the s   s_{br} regime. The four-parameter SBPL model more comprehensively describes the intrinsic distribution, leading us to speculate that previous fits to p(s)   s^γ may have overestimated γ, with the true value being closer to Öpik’ s law.

Due to Gaia’ s angular resolution limits and brightness difference effects, our catalog remains incomplete, causing significant biases in fitted intrinsic distributions (especially γ), manifested as increasing deficiency at small projected separations and γ deviations from -1.5. Notably, restricting SPL’ s separation range from $s < 10^{5} \cdot {}^{20}$ AU to $s < 10^{4}$ AU (s_{br}) shifts γ toward the normal value of -1.5, which is expected since SPL doesn’ t account for s_{br} while γ is defined for the s   s_{br} region. For any given sample, best-fit γ values typically satisfy γ*{SBPL}* γ{DPL} γ_{SPL}.

This study calibrates biases in power-law model parameters (particularly γ) caused by selection effects, but limitations remain: the sample doesn’ t cover smaller projected separations, and false binary contamination continues to affect catalog construction. Our catalog represents a trade-off between completeness and purity, without fundamentally resolving selection effects or incompleteness from observational characteristics and search methods.

Future Gaia data releases (e.g., Gaia DR3) will provide more information on selection effects. Combining Gaia with other surveys (e.g., LAMOST low-resolution spectroscopy) will enable better studies of projected separation distributions and selection effect variations across different binary populations.

References

[1] Moe M, Di Stefano R. ApJS, 2017, 230(2): 15
[2] El-Badry K, Rix H W, Heintz T M. MNRAS, 2021, 506(2): 2269
[3] Halbwachs J L, Mayor M, Udry S, et al. A&A, 2003, 397(1): 159
[4] Lépine S, Bongiorno B. AJ, 2007, 133(3): 889
[5] Tobin J J, Kratter K M, Persson M V, et al. Nature, 2016, 538(7626): 483
[6] Tokovinin A, Moe M. MNRAS, 2020, 491(4): 5158
[7] Offner S S, Kratter K M, Matzner C D, et al. ApJ, 2010, 725(2): 1485
[8] Klein R L, Fisher R, Krumholz M, et al. Rev Mex Astron Astrofis, 2002, 15: 92
[9] Kouwenhoven M B N, Goodwin S P, Parker R J, et al. MNRAS, 2010, 404(4): 1835
[10] Perets H B, Kouwenhoven M B N. ApJ, 2012, 750(1): 83
[11] Reipurth B. Mem Soc Astron Ital, 2017, 88: 611
[12] Reipurth B, Mikkola S. AJ, 2015, 149(4): 145
[13] Tokovinin A. MNRAS, 2017, 468(3): 3461
[14] Chanamé J, Gould A. ApJ, 2004, 601(1): 289
[15] Yoo J, Chaname J, Gould A, et al. ApJ, 2004, 601(1): 311
[16] Quinn D P, Wilkinson M I, Irwin M J, et al. MNRAS, 2009, 396(1): L11
[17] Monroy-Rodríguez M A, Allen C. ApJ, 2014, 790(2): 159
[18] El-Badry K, Rix H W. MNRAS, 2018, 480(4): 4884
[19] Jiang Y F, Tremaine S. MNRAS, 2010, 401(2): 977
[20] TIAN H J, El-Badry K, Rix H W, et al. ApJS, 2019, 246(1): 4
[21] Andrews J J, Chanamé J, Agüeros M A. MNRAS, 2017, 472(1): 675
[22] Pittordis C, Sutherland W. MNRAS, 2018, 480(2): 1778
[23] Livernois A R, Vesperini E, Pavlík V. MNRAS, 2023, 521(3): 4395
[24] Peñarrubia J, Ludlow A D, Chanamé J, et al. MNRAS, 2016, 461(1): L72

[25] Fellhauer M, Lin D N C, Bolte M, et al. ApJ, 2003, 595(1): L53
[26] Brown A G, Vallenari A, Prusti T. A&A, 2021, 649: A1
[27] Oh S, Price-Whelan A M, Hogg D W, et al. AJ, 2017, 153(6): 257
[28] Babusiaux C, Van Leeuwen F, Barstow M A, et al. A&A, 2018, 616: A10
[29] Lindegren L, Hernández J, Bombrun A, et al. A&A, 2018, 616: A2
[30] Evans D W, Riello M, De Angeli F, et al. A&A, 2018, 616: A4
[31] Arenou F, Luri X, Babusiaux C, et al. A&A, 2018, 616: A17
[32] Leike R H, Enßlin T A. A&A, 2019, 631: A32
[33] Widmark A, Leistedt B, Hogg D W. ApJ, 2018, 857(2): 114
[34] Gaia Collaboration, et al. The Gaia Mission. https://arxiv.org/pdf/1609.04153, 2016
[35] Herschel M, Watson D. Philosophical Transactions of the Royal Society of London, 1782, 72: 112
[36] Herschel W. Philosophical Transactions of the Royal Society of London, 1803, 93: 339
[37] Sesar B, Ivezić Ž, Jurić M. ApJ, 2008, 689(2): 1244
[38] Duchêne G, Kraus A. ARA&A, 2013, 51: 269
[39] Dhital S, West A A, Stassun K G, et al. AJ, 2010, 139(6): 2566
[40] Oelkers R J, Stassun K G, Dhital S. AJ, 2017, 153(6): 259
[41] Ziegler C, Law N M, Baranec C, et al. AJ, 2018, 156(6): 259
[42] Foreman-Mackey D, Hogg D W, Lang D, et al. PASP, 2013, 125(925): 306
[43] Foreman-Mackey D. JOSS, 2016, 1(2): 24

Note: Figure translations are in progress. See original paper for figures.