Distinguishing exoplanet companions from field stars in direct imaging using Gaia astrometry

Philipp Herz, Matthias Samland, and Coryn A. L. Bailer-Jones

Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany
e-mail: calj@mpia.de

Received 5 November 2023 / Accepted 7 December 2023

ABSTRACT

Direct imaging searches for exoplanets around stars detect many spurious candidates that are in fact background field stars. To help distinguish these from genuine companions, multi-epoch astrometry can be used to identify a common proper motion with the host star. Although this is frequently done, many approaches lack an appropriate model for the motions of the background population, or do not use a statistical framework to properly quantify the results. For this study we used Gaia astrometry combined with 2MASS photometry to model the parallax and proper motion distributions of field stars around exoplanet host stars as a function of candidate magnitude. We developed a likelihood-based method that compares the positions of a candidate at multiple epochs with the positions expected under both this field star model and a co-moving companion model. Our method propagates the covariances in the Gaia astrometry and the candidate positions. True companions are assumed to have long periods compared to the observational baseline, so we currently neglect orbital motion. We applied our method to a sample of 23 host stars with 263 candidates identified in the B-Star Exoplanet Abundance Study (BEAST) survey on VLT/SPHERE. We identified seven candidates in which the odds ratio favours the co-moving companion model by a factor of 100 or more. Most of these detections are based on only two or three epochs separated by less than three years, so further epochs should be obtained to reassess the companion probabilities. Our method is publicly available as an open-source python package from GitHub to use with any data.

Key words. methods: statistical – techniques: high angular resolution – planets and satellites: detection – stars: early-type

1. Introduction

Young exoplanets with a favourable brightness and separation to their host star can be directly imaged. However, such exoplanets can be confused with more distant background stars that happen to lie in the line of sight. A common way to distinguish these scenarios is to observe both the host star and candidate over time to look for a common proper motion and/or parallax. An example of doing this is shown in Fig. 1 (data from Janson et al. 2021b and published in Squicciarini et al. 2022). The positions of the host star and a number of exoplanet candidates were measured at two epochs, 2018 and 2021. The orange crosses show the measured change in position (relative to the host star) of candidates between these two epochs. The host star has some proper motion and parallax between 2018 and 2021: the black dashed line shows how objects with zero proper motion and parallax would move over this time period (as our view is centred on the host star). Those orange points clustered near the black point labelled 2021 are therefore consistent with being distant background stars. The candidate labelled ‘b’, on the other hand, has a motion more consistent with the host star and so is more likely to be a true companion.

To make this procedure quantitative, we must take into account the measurement uncertainties and any covariance between them. Background stars do not have zero parallax and proper motion, so we need a proper model for their motions too. There may also be more than two epochs, so we want to take into account all of the data within a single assessment. In the literature these confounding factors are usually not considered (Lagrange et al. 2010; Lafrenière et al. 2011; Caron et al. 2013; Kuzuhara et al. 2013; De Rosa et al. 2015; Konopacky et al. 2016; Chauvin et al. 2017; Kepler et al. 2018; Bohn et al. 2021; Janson et al. 2021a; Squicciarini et al. 2022; Franson et al. 2023; Mesa et al. 2023; Chomez et al. 2023). Such studies often compare the second epoch position with that expected for a stationary background star. If they deviate significantly, the null hypothesis of being a background star is rejected and some alternative model, that is, a companion, is implicitly accepted. This classical hypothesis-testing approach does not, however, assess whether the data might be even more unlikely under the companion model.

The goal of this paper is to put this astrometric confirmation approach on a solid statistical footing. We developed a model to evaluate whether the multi-epoch motion of the candidate is more likely to be a co-moving exoplanet or a coincidental field star. Our model is based on the proper motion and parallax distributions of field stars in the same area of the sky as the candidate, and with similar magnitudes to the candidate being tested. Using the odds ratio, we compare this background model with a model in which the candidate is co-moving with the host star. We applied our method to multi-epoch measurements of candidates observed in the B-Star Exoplanet Abundance Study (BEAST) survey (Janson et al. 2021b), which has an abundance of candidate objects due to observing close to the Galactic plane. A python package implementing our method is accessible via GitHub.

Full Table 1 is available at the CDS via anonymous ftp to cdsarc.cds.unistra.fr (130.79.128.5) or via https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/682/A92

1 https://github.com/herzphi/compass
∆allactic motion. In a Cartesian plane projection, we write this as a linear motion with a superimposed parallax as a function of time. Assuming that the candidate does not use other present or future surveys.

We consider the position of a candidate object relative to the host star, and the latter additionally uses a magnitude-dependent fit to the proper motion and parallax distribution of field stars close to the host star’s line of sight. Our method can use an arbitrary number of epochs of astrometric observations to derive the likelihoods. We also implemented a special case of the general method that neglects the parallax. This is less realistic but is easier to visualize. The field star astrometric model is currently built on data from the Gaia (Gaia Collaboration 2023) and 2MASS (Two Micron All Sky Survey; Skrutskie et al. 2006) surveys, but could use other present or future surveys.

2. Methods

We developed a probabilistic method that compares the likelihoods of position measurements of an exoplanet candidate under two models: the first assumes the candidate is a co-moving companion, and the second assumes it is a field star. The former uses the proper motion and parallax of the exoplanet candidate’s host star, and the latter additionally uses a magnitude-dependent fit to the parallax and proper motion distributions of field stars close to the host star’s line of sight. Our method can use an arbitrary number of epochs of astrometric observations to derive the likelihoods. We also implemented a special case of the general method that neglects the parallax. This is less realistic but is easier to visualize. The field star astrometric model is currently built on data from the Gaia (Gaia Collaboration 2023) and 2MASS (Two Micron All Sky Survey; Skrutskie et al. 2006) surveys, but could use other present or future surveys.

2.1. Positional model

We consider the position of a candidate object relative to the host star as a function of time. Assuming that the candidate does not move relative to the star, then in a geocentric coordinate system this can be described as a linear motion with a superimposed parallactic motion. In a Cartesian plane projection, we write this as

\[ \Delta \mathbf{x}_i = \left( \Delta \mathbf{x}_i^a - \Delta \mathbf{x}_0^a \right) + \left( \mu \mathbf{a}_a - \mu \mathbf{s}_a \right) t_i + \left( \varpi \mathbf{a}_s - \varpi \mathbf{s}_s \right) s_i(t_i) \]

\[ \Delta \mathbf{y}_i = \left( \Delta \mathbf{y}_i^a - \Delta \mathbf{y}_0^a \right) + \left( \mu \mathbf{a}_a - \mu \mathbf{s}_a \right) t_i + \left( \varpi \mathbf{a}_s - \varpi \mathbf{s}_s \right) s_i(t_i), \]

where \( \mu \) is the proper motion and \( \varpi \) is the parallax. Subscript \( a \) refers to the candidate object, subscript \( \star \) to the host star, subscript \( 0 \) to the true position of the candidate at a reference epoch, and subscript \( i \) to the \( i \)th epoch relative to the reference epoch. The functions \( s_i \) and \( s_i \) are periodic phase factors for the parallax motion given by the orbital motion of Earth around the Sun at epoch \( i \). The primed variables denote the true position of the candidate, while the unprimed variables denote the measured position of the candidate. In this paper ‘true positions’ denote those we would obtain in the absence of noise in our positional measurements at the epochs, but they are still based on some externally provided values of parallax and proper motion. Those external values are considered to be noisy, and their covariances are taken into account.

Equation (1) can be generalized to \( N \) measurements of the candidate’s position over time \( \Delta \mathbf{x}_i, \Delta \mathbf{y}_i \) at times \( t = \{ t_i \} \) with \( i \in [1, N] \), where the bold font indicates a vector of measurements. The first position measurement defines the true position of the candidate in the first epoch \( \Delta \mathbf{x}_1, \Delta \mathbf{y}_1 \) without loss of generality.

2.2. Overall probabilistic model

Given the measured positions of the exoplanet candidate over time, we computed the likelihood of the data under two models. The first model, denoted \( M_\star \), assumes the exoplanet candidate is a co-moving companion, sharing the same proper motion and parallax as the host star. We made the simplifying assumption that the orbital period is long compared to the observational baseline and so neglected the candidate’s orbital motion. This may not be justified in all cases, which we elaborate on in Sect. 4. The second model, denoted \( M_b \), assumes the candidate to be a background object with a proper motion and parallax distribution constructed from a set of background stars in a narrow field of view around the host star (described in Sect. 2.3). We then compared these two models via the odds (likelihood) ratio

\[ r_{c:b} = \frac{P(\Delta \mathbf{x}, \Delta \mathbf{y} \mid M_\star)}{P(\Delta \mathbf{x}, \Delta \mathbf{y} \mid M_b)}, \] (2)

which indicates which model favours the data more. This does not give the posterior probability of the model given the data, however, for which we would first need to establish model prior probabilities.

Let us first consider the denominator in Eq. (2), the likelihood under the background model. This can be computed as a combination of two probability density functions (PDFs). The first of these is the probability of the noisy position measurements given the true positions, \( P(\Delta \mathbf{x}, \Delta \mathbf{y} \mid \Delta \mathbf{x}^\prime, \Delta \mathbf{y}^\prime) \), which reflects the noise in the determination of the centroid of a point spread function on the detector. We assumed this to be a Gaussian with mean \( \Delta \mathbf{x}^\prime, \Delta \mathbf{y}^\prime \) and a covariance matrix \( \Lambda \) that reflects the accuracy of, and correlations in, the measurements. The second term is \( P(\Delta \mathbf{x}^\prime, \Delta \mathbf{y}^\prime \mid \Delta \mathbf{x}_i, \Delta \mathbf{y}_i) \). This represents the spread in possible true positions of a background star at specific times arising from the spread in the proper motion and parallax of the background star population. We assumed this term to be a Gaussian distribution with a mean given by the right side of Eq. (1). The terms in that equation, as well as the covariance matrix of the Gaussian (which we denote \( \Lambda^\prime \)), come from our fit to the background star population (Sect. 2.3).

\[ \text{We use the term ‘background objects’ and ‘field stars’ interchangeably, as most field stars are more distant than the host star.} \]
two PDFs we then combined via a marginalization to give the required likelihood

\[
P(\Delta x, \Delta y \mid t, M_b) = \int_{\Delta x', \Delta y'} P(\Delta x, \Delta y \mid \Delta x', \Delta y')
\times P(\Delta x', \Delta y' \mid t, M_b) \, d\Delta x' \, d\Delta y'.
\]  

This shows that the background model likelihood is a convolution of the two PDFs. Given that these are both Gaussian, the result of the convolution is also a Gaussian, with mean \((\Delta x', \Delta y')\) and covariance matrix \((\Lambda + \Lambda')\). More details are provided in Appendix A.

We took the same approach to computing the likelihood of the data under the co-moving companion model \(M_c\), the numerator in Eq. (2). Here, however, the relative position of the exoplanet candidate was assumed to be constant, meaning that the relative position measurements have zero variance and so \(P(\Delta x', \Delta y' \mid t, M_c)\) is simply a delta function. The convolution in Eq. (3) is then trivial, resulting in \(P(\Delta x, \Delta y \mid T, M_c)\) being a Gaussian with mean \((\Delta x', \Delta y')\) and covariance matrix \(\Lambda\).

Having now expressions for both likelihoods in terms of the measured relative positions and their covariances, and in terms of the parallax and proper motion distribution of the background model, we could compute the odds ratio in Eq. (2) and thus decide which model better explains the data.

We refer to the above general method as the proper motion and parallax covariance method. If the parallax is negligible, or if the observations are separated by almost exactly a year, then we can set the parallax in Eq. (1) to zero. If we also have only two epochs, then instead of using the two position measurements directly, we could convert them into a single proper motion. We call this the ‘proper motion-only’ method. In this special case, the ‘true’ proper motion \(\Delta \mu_t = (\Delta \mu_x', \Delta \mu_y')\) relative to the host star can be written

\[
\Delta \mu_t = \frac{\Delta x' - \Delta x}{t_2 - t_1}
\]

which we compared to the measured relative proper motions from the data (same equation without the accents). The likelihoods for the two models are then

\[
P(\Delta \mu \mid M_c) = N\left(0, \begin{bmatrix} \text{Var}(\Delta \mu_x) & \text{Cov}(\Delta \mu_x, \Delta \mu_y) \\ \text{Cov}(\Delta \mu_y, \Delta \mu_x) & \text{Var}(\Delta \mu_y) \end{bmatrix} \right)
\]

and

\[
P(\Delta \mu \mid M_b) = \int P(\Delta \mu \mid \Delta \mu') P(\Delta \mu' \mid M_b) \, d\Delta \mu'.
\]

Both equations are Gaussian distributions. In the co-moving companion model the likelihood is centred at zero with a covariance matrix given by the uncertainties in position measurements. In the background model the likelihood is centred at the proper motion \(\Delta \mu_t\) of the background star distribution relative to the host star, with a covariance matrix given by the convolution of the proper motions of this background star population with the measurement uncertainties.

A more complete derivation of both the full model and this simpler (but less representative) proper motion-only model is given in Appendix A\(^3\).

---

\( ^3 \) More details can be found at [http://mpia.de/homes/calj/astrometric_companion_model.pdf](http://mpia.de/homes/calj/astrometric_companion_model.pdf)
Fig. 3. Proper motion distribution of stellar objects with a $K_S$-magnitude between 18 and 19 in a 0.3° sky area around HIP 82545 ($\mu^2$ Sco). The elliptical contours are the boundaries that encompass 50, 90, and 99% of the stellar objects.

Fig. 4. Variation of the mean of the proper motion distribution in field stars as a function of stellar magnitude. Each bin includes 200 stellar objects in a 0.3° sky area around HIP 82545 ($\mu^2$ Sco). Each point is the mean value of the 2D Gaussian fit in each magnitude bin. Only those points lying within the 10th–90th percentile range of magnitudes were used for the linear fit.

Fig. 5. Variation of the standard deviation and correlation of the proper motion distributions for the field stars around HIP 82545 ($\mu^2$ Sco) as a function of magnitude. Each point comes from a 2D Gaussian fit over a narrow magnitude bin (shown in Fig. 3). Each bin includes 200 stars in a 0.3° sky area around the target star. Crosses denote fits using stars with 2MASS magnitudes, and circles denote those with magnitudes computed from the Gaia colour transformation.

with $a > 0$. This fit has a minimum value of $\sigma_{\text{min}}$, a fixed constant. The magnitude $m_0$ is the mean of the observed magnitudes. An example of this fit is shown in Fig. 5. A linear fit was used instead if the exponential fit residuals were larger than those of a linear fit. In making this linear fit we prevented $\sigma_{\text{fit}}(m)$ from becoming smaller than 1 mas yr$^{-1}$.

We did not identify any definitive and consistent trend in the correlation between parallax and proper motion. We therefore opted to fit a constant value for the correlation instead.

2.4. Verification of the method using simulations

A verification of our model would ideally be based on a large set of real data in which the true nature of each candidate is known, but this was not available. We therefore used simulations to assess the reliability of the odds ratio. We simulated 1000 trajectories of a co-moving companion and 1000 trajectories of background model objects.
Starting with the real star $\mu^2$ Sco, we constructed a background model for proper motion and parallax. The initial position of the candidate relative to $\mu^2$ Sco was randomly generated. The position at subsequent epochs (one year apart) is given by

$$x_{t+1} = x_t + \left( \mu_x + N(0, \sigma = 3 \text{ mas yr}^{-1}) \right) \Delta t$$

and similarly for $y$. This equation represents either a co-moving companion with a zero (relative) proper motion ($\mu_x = \mu_y = 0$), or a relative proper motion of the background model to $\mu^2$ Sco ($\mu_x = \mu_{\mu,\delta} - \mu_{\mu,\delta}^s$ Sco, where $\mu_{\mu,\delta}^s$ Sco is a fixed value defined by the brightness of the candidate), both with a normal distribution $N$ to simulate the propagation of the proper motion uncertainty. We chose a standard deviation of 3 mas yr$^{-1}$. We simulated 1000 paths the candidate could take under the companion model $M_c$, using Eq. (8) with $\mu = 0$, and another 1000 paths under the background object model $M_b$ using a single fixed proper motion for a background object with a similar brightness to the candidate. The spread seen in Fig. 6 thus arises from the noise we add at each epoch. Our method distinguishes between those paths via a logarithmic odds ratios greater than zero (companion model favoured, shown in blue) and smaller than zero (background model favoured, shown in black). We found that all 2000 trajectories were correctly identified as coming from the model from which they were generated.

### 3. Results

#### 3.1. General results

BEAST (Janson et al. 2021b) employs the VLT (Very Large Telescope)/SPHERE (Spectro-Polarimetric High-contrast Exoplanet REsearch) Beuzit et al. (2019) extreme adaptive optics instrument to conduct a direct imaging survey of 85 B-type stars in the young (5–20 Myr) and relatively nearby (120–150 pc) Scorpius-Centaurus (Sco-Cen) region (Fig. 7). The targets have similar distance and $G$-band magnitude, but vary in colour. By targeting young B-type stars, their exoplanets should be relatively bright in the near-infrared. Before this survey, B-stars had not been systematically surveyed for exoplanets. Radial velocity surveys, which are more sensitive to close-in planets, reveal comparatively few planets around massive stars (Reffert et al. 2015). Thus, BEAST addresses the question of whether massive stars can form massive planets at larger separations. Certain formation scenarios, such as disk instability, might primarily take place in the outermost areas of extensive disks surrounding massive stars (Helled et al. 2014). Sco-Cen is located close to the Galactic plane, so observations often include many spurious candidates in the field-of-view. This makes the survey a prime target for testing our method of distinguishing bound companions from faint field stars.

Of the 85 stars in BEAST, 23 have candidates with at least two epochs of observation (as of 2022-04-01). These 23 stars have in total 263 candidates. The projected separations between the candidates and their host stars range from 49 to 1457 au (Fig. 8).

We applied both our proper motion-only model and our proper motion and parallax covariance model to all 263 candidates. The results are shown in Table 1, sorted by descending logarithmic odds ratio $\log_{10} \ell_c(\mu, \pi)$, that is, the most likely real companions appear first. In what follows we focus on the results using the parallax and proper motion model in the final column. Visualizations of all 20 candidates are shown later. The complete table of candidate astrometry and photometry from the BEAST survey will be published in Delorme et al. (in prep.).

Most interesting are the seven candidates that have an odds ratio greater than ten ($\log_{10} \ell_c(\mu, \pi) > 1$). Two of these, b Cen b and $\mu^2$ Sco b, are well-established as exoplanets (Janson et al. 2021b; Squicciarini et al. 2022). We further infer...
Table 1. Candidates with logarithmic odds ratio greater than zero under the parallax and proper motion model, sorted by decreasing $\log_{10} r_{cb}(\mu)$ (Eq. (2)).

<table>
<thead>
<tr>
<th>Host Star</th>
<th>Host Star</th>
<th>Candidate ID</th>
<th>$K_s$-band (mag)</th>
<th>$N_{obs}$</th>
<th>$\Delta t_{obs}$ (yr)</th>
<th>$\log_{10} r_{cb}(\mu)$</th>
<th>$\log_{10} r_{cb}(\mu, \sigma)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIP 71865</td>
<td>b Cen</td>
<td>b (1oh0dfaj)</td>
<td>17.17</td>
<td>3</td>
<td>20.9</td>
<td>15.78</td>
<td>$\geq 300$</td>
</tr>
<tr>
<td>HIP 61257</td>
<td>HD 109195</td>
<td>'B' (58pwu3dv)</td>
<td>12.81</td>
<td>3</td>
<td>2.8</td>
<td>1.81</td>
<td>$\geq 300$</td>
</tr>
<tr>
<td>HIP 82545</td>
<td>HD 149274</td>
<td>B</td>
<td>13.63</td>
<td>2</td>
<td>2.7</td>
<td>$\sim 2.71$</td>
<td>91.66</td>
</tr>
<tr>
<td>HIP 81208</td>
<td>HD 93563</td>
<td>x9ld1uh0</td>
<td>12.03</td>
<td>2</td>
<td>1.1</td>
<td>1.57</td>
<td>12.56</td>
</tr>
<tr>
<td>HIP 76048</td>
<td>HD 138221</td>
<td>vslvj1zp</td>
<td>14.50</td>
<td>2</td>
<td>0.2</td>
<td>1.42</td>
<td>2.00</td>
</tr>
<tr>
<td>HIP 60009</td>
<td>(\zeta) Cru</td>
<td>rtolbhl8</td>
<td>20.49</td>
<td>2</td>
<td>1.0</td>
<td>0.07</td>
<td>0.32</td>
</tr>
<tr>
<td>HIP 62327</td>
<td>HD 110956</td>
<td>n1lk1oif</td>
<td>19.64</td>
<td>2</td>
<td>1.0</td>
<td>0.08</td>
<td>0.30</td>
</tr>
<tr>
<td>HIP 69011</td>
<td>HD 123247</td>
<td>962kt2wv</td>
<td>18.16</td>
<td>2</td>
<td>0.2</td>
<td>0.81</td>
<td>0.26</td>
</tr>
<tr>
<td>HIP 60009</td>
<td>(\zeta) Cru</td>
<td>ebmx222k</td>
<td>19.97</td>
<td>2</td>
<td>1.0</td>
<td>0.05</td>
<td>0.18</td>
</tr>
<tr>
<td>HIP 74100</td>
<td>HD 133937</td>
<td>ldg61h81</td>
<td>19.97</td>
<td>2</td>
<td>1.0</td>
<td>0.08</td>
<td>0.18</td>
</tr>
<tr>
<td>HIP 52742</td>
<td>HD 93563</td>
<td>mmn9uw3ha</td>
<td>19.49</td>
<td>2</td>
<td>1.1</td>
<td>0.01</td>
<td>0.16</td>
</tr>
<tr>
<td>HIP 60009</td>
<td>(\zeta) Cru</td>
<td>qhsttxkw</td>
<td>20.14</td>
<td>2</td>
<td>1.0</td>
<td>0.03</td>
<td>0.15</td>
</tr>
<tr>
<td>HIP 60009</td>
<td>(\zeta) Cru</td>
<td>n4iz1pt</td>
<td>19.51</td>
<td>2</td>
<td>1.0</td>
<td>0.04</td>
<td>0.15</td>
</tr>
<tr>
<td>HIP 60009</td>
<td>(\zeta) Cru</td>
<td>4h805ve</td>
<td>18.85</td>
<td>2</td>
<td>1.0</td>
<td>0.06</td>
<td>0.14</td>
</tr>
<tr>
<td>HIP 60009</td>
<td>(\zeta) Cru</td>
<td>a7214wz3</td>
<td>19.39</td>
<td>2</td>
<td>1.0</td>
<td>0.09</td>
<td>0.12</td>
</tr>
<tr>
<td>HIP 60009</td>
<td>(\zeta) Cru</td>
<td>5s2kr75</td>
<td>20.11</td>
<td>2</td>
<td>1.0</td>
<td>0.05</td>
<td>0.10</td>
</tr>
<tr>
<td>HIP 60009</td>
<td>(\zeta) Cru</td>
<td>d9oc07ww</td>
<td>19.01</td>
<td>2</td>
<td>1.0</td>
<td>0.17</td>
<td>0.07</td>
</tr>
<tr>
<td>HIP 60009</td>
<td>(\zeta) Cru</td>
<td>g7art0xk</td>
<td>19.23</td>
<td>2</td>
<td>1.0</td>
<td>0.12</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Notes. Those candidates with logarithmic odds ratio above 1.0 we identify as likely companions (in bold face). The column ‘Candidate ID’ contains unique identifier strings assigned to each candidate by Janson et al. (2021b) in most cases, and/or the commonly accepted companion name (‘B’ referring to a very probable stellar-mass companion). Here we show the first 20 candidates. The full table including results for 263 candidates (including six candidates published in Viswanath et al. 2023 that did not have follow-up when our list was compiled) is available at the CDS.

that the candidates of HIP 81208 (two candidates), HIP 61257, HIP 52742, and HIP 76048 are likely true companions (but not necessarily in the exoplanet mass regime). The remaining candidates in Table 1 are still formally favoured by the co-moving companion model, but not by much. These will require more epochs or other observations in order to determine their nature. We now examine the results on some of the individual, high odds ratio candidates, which also serve to illustrate how our method works.

3.2. Results on individual candidates

\(\mu^2\) Sco b. The astrometric field star model for the target \(\mu^2\) Sco is shown in Fig. 9. The black line is our fit to the background stars (described in Sect. 2.3), which can be compared to the various BEAST candidates in this field in orange. Nearly all of the candidates agree with the field star model, showing the large degree of field star contamination that can be present in exoplanet searches. Just one does not agree; this is the exoplanet \(\mu^2\) Sco b. The PDF over the proper motion (the likelihood) for this candidate under the background model is shown by the black curves and contours in Fig. 10. The likelihood for the companion model is shown by the blue curves and contours, and the measured proper motion is shown as the orange point and line. In this example we see that the measurement is far more consistent with the companion model than with the background model. This is a rather clear-cut case even by visual inspection; many other cases in Appendix B are more ambiguous.

b Cen b. This candidate, which was identified as an exoplanet by Janson et al. (2021a), has three epochs in our analysis.
Demonstration of the proper motion-only method for the known exoplanet $\mu^2$ Sco b. The likelihood under the background model is shown in black and the likelihood under the co-moving companion model is shown in blue. The bottom left panel shows these as two-dimensional (Gaussian) distributions. The other two panels show the one-dimensional marginal distributions. The measured relative proper motion is shown in orange: the uncertainties in this is not shown because it is included in the two likelihoods (see Eqs. (5) and (6)).

and the measured positions of the candidate (orange). The first observation of this candidate was at the epoch 2000.4 (circle) based on archival images by Shatsky & Tokovinin (2002). In the relative reference frame to the host star, a co-moving companion would not move from its 2000.4 position in our model, and so is not shown for other epochs. The next observation of the candidate took place in 2019.2 with BEAST. Based on the background proper motion and parallax distribution, as well as the host star’s proper motion, a background object in the same area of the sky as the host star would have moved about 850 mas since 2000.4. However, the figure shows that by 2019.2 the candidate has moved much less (orange point). Another epoch at 2021.3 confirms that the motion of the candidate is much closer to the host star’s motion than it is to that of the background population. The background star hypothesis is clearly ruled out in this case. But because the measured positions lie near the 99% contour of our co-moving companion model, some residual motion compared to the co-moving companion model likely remains. This can be explained by orbital motion – which is not taken into account – over the 21-yr observational baseline (Janson et al. 2021a).

HIP 61257 B'. This highly probable candidate of HIP 61257 can be further examined by including non-BEAST data. Kouwenhoven et al. (2005) discuss a potential companion around HIP 61257, but they ultimately classify it as a background object. This potential companion has a $K_s$-band magnitude of 12.43 mag and a separation of 5540 mas, which coincides with the magnitude and separation of one of our candidates with logarithmic odds ratio greater than zero in both modelling frameworks. The astrometric motion plot (Fig. 12) shows that this candidate is co-moving with the host star. As the astrometric data from Kouwenhoven et al. (2005) are reported without separation and position angle uncertainties, we adopted an uncertainty of 10 mas in the relative right ascension and declination. This companion was also discussed by Gratton et al. (2023), who identified it to be a low-mass star (0.083 $\pm$ 0.01$M_\odot$) based on its $Gaia$ and $K$-band magnitudes. The logarithmic odds ratios without the archival data are $\log r_c(\mu, \sigma) = 12.91$ and $\log r_c(\mu) = 2.45$. We therefore confirm this candidate as a genuine binary companion based on its astrometry.

**HIP 81208 B and C.** These two candidates are identified as co-moving companions by Viswanath et al. (2023). They identify HIP 81208 B as a 67 MJup object, mostly likely a brown dwarf, and HIP 81208 C as a 0.135 $M_\odot$ low-mass star. Our results for these are shown in Figs. 13a,b (using astrometry and photometry from Table C.1 in Viswanath et al. 2023). Both candidates are favoured by the co-moving companion model with our more sophisticated parallax and proper motion method. The proper motion-only model rejects HIP 81208 B but accepts HIP 81208 C. This system was recently analysed further and
found to be a gravitationally bound hierarchical quadruple system comprised of low mass objects with a newly discovered companion to the C component (Cb; Chomez et al. 2023).

**HIP 52742 x9ld1uh0 and mm9uw3ha.** Two candidates with magnitudes of $K_S = 12.03$ mag and $K_S = 19.49$ mag are formally favoured by our companion model. The brighter one we identify as a companion with high significance (log odds ratio 12.6). Gratton et al. (2023) also identified this as a companion: Adopting an age of 82.5 Myr (obtained from assuming membership of a comoving group outside Sco-Cen; Janson et al. 2021b) they estimated a mass of 0.51 $M_\odot$ at a projected separation of 176 au. In a search for astrometric acceleration from a comparison of Hipparcos and *Gaia* DR3 proper motions, Brandt (2021) found marginally significant evidence for an acceleration of this star, which might be evidence of this companion. The fainter candidate companion to this star that we identify has a log odds ratio of just 0.16, which is not significant.

**HIP 76048 vslv1zp.** We identify one potential companion to this star with an odds ratio of 100 in our analysis, albeit with a very short baseline of just 0.2 yr. Brandt (2021) found no significant evidence of acceleration from their HIPPARCOS–*Gaia* DR3 proper motion study.

### 4. Discussion

#### 4.1. Interpretation

Our analysis favours the co-moving companion model for five candidates from the BEAST data set accessible for this study, as well as two candidates in the HIP 81208 system that have a second epoch from Viswanath et al. (2023). Of these seven candidates, two are confirmed exoplanets: b Cen AB b (Janson et al. 2021a) and $\mu$2 Sco b (Squicciarini et al. 2022). HIP 61257 ‘B’ is very likely a stellar-mass companion (Gratton et al. 2023). The two candidates of HIP 81208 are discussed as being a brown dwarf and a low-mass stellar companion by Viswanath et al. (2023). The remaining two companions of the host stars HIP 52742 and HIP 76048 – with a measurement baseline of only 0.2 yr – are unconfirmed at the time of writing.

Most of the targets in our analysis have only two observation epochs with small temporal baselines (of order one year), so many candidates show little motion relative to the host star between the epochs. It is precisely these cases where ad hoc approaches to assessing companionship are inconclusive, and our statistical framework is most useful. For many of these short baselines, the co-moving model is not favoured. Longer temporal baselines make it easier to distinguish between the models, especially as the targets of direct imaging campaigns tend to be nearby stars with large proper motion.

#### 4.2. Model assumptions

Our model does not currently take into account orbital motion, so the co-moving companion model may not be favoured even if the candidate is a true companion with significant orbital motion over the observational baseline. Orbital periods for directly imaged planets tend to be long, of the order of centuries or even millennia. In those cases, neglecting the orbital motion does not affect our analysis. Planets with relatively short orbits, in contrast, such as $\beta$ Pic b and c with periods of 23.6 and 3.3 yr respectively (Lacour et al. 2021), can show significant orbital motion.

If orbital motion can be directly observed then the candidate is very likely to be confirmed as a companion (for example Marois et al. 2008). This could be included as a relatively straight forward extension to our modelling approach by including a model with path curvature. Plausible paths (priors) could be generated from a Keplerian model (using the model of Blunt et al. 2020, for example) and then marginalizing over them to determine the model likelihood. An alternative approach is to show that a candidate’s motion is consistent with that of field stars while also showing that the range of orbits that can explain the data are beyond the escape velocity of the system (see for example Nielsen et al. 2017; Wagner et al. 2022).

We built our field star astrometric model using *Gaia* parallaxes and proper motions. On account of the limited depth of *Gaia*, we then had to extrapolate our model to the fainter magnitudes of our candidates. This can introduce biases, and thus incorrectly favour or disfavour the background model.

Another drawback of *Gaia* is that it observes in the optical, whereas direct imaging surveys for planets are currently done mostly in the near-infrared. To obtain the necessary infrared magnitudes of the *Gaia* sources we had to crossmatch to an infrared survey. We chose 2MASS because it is all-sky. However, it is not very deep, so for many field stars we instead had to predict the $K_S$-band magnitude using colour transformations from *Gaia*. We used the transformations of Riello et al. (2021) for objects in the colour range $-0.5 < G_{RP} - G < 2.5$. The accuracy of this for brighter stars around $\mu^2$ Sco is demonstrated in Fig. 14. The 2MASS objects span a $K_S$-band range of only 10 to 18 mag while the transformed photometry extends to 21 mag. Some exoplanet surveys are also conducted at shorter wavelengths, such as $J$ and $H$, for which colour transformations are more reliable as they are closer to the observed *Gaia* bands. Ideal for our purposes, of course, would be an infrared version of *Gaia*. 
Fig. 13. Astrometric motion of the two candidates of HIP 81208 with an observation baseline of three years. Viswanath et al. (2023) report both objects as co-moving companions based on their proper motion analysis. Our analysis supports this claim.

Fig. 14. Difference between the $K_S$ magnitude calculated from $Gaia$ via colour transformations and the and measured 2MASS $K_S$ magnitudes for 934,023 objects from $Gaia$ with 2MASS counterparts in the vicinity of $\mu^2$ Sco.

As explained in Sect. 2.3, in our background model the parallax and proper motion distributions depend only on direction and magnitude. A further improvement would be to add an additional dependence on the colour, if that is provided by the imaging survey. We could then also use the measured colour to infer something about the intrinsic properties of the object, such as its spectral type (Parviainen et al. 2019). Although that may help determine whether or not it is a low-mass object, it would not tell us whether it is gravitationally bound.

In computing the odds ratio we only considered kinematics. We do not take into account the number density of background stars or the angular separation between the candidate and target star. Yet for a given separation, the less dense the background, the more likely the candidate is a genuine companion (Tamura 2016; Squicciarini et al. 2022). This information could be incorporated as an additional multiplicative odds ratio, although it requires a model (or measurement) of the stellar density at the faint magnitudes at which we conduct our exoplanet survey.

Finally, we emphasize that we report our results as a ratio of likelihoods of two models, where each likelihood is the probability of the data given the model. To convert each likelihood into a posterior probability of the model given the data, we would need to adopt a prior probability for the model, and know that our models are exhaustive. The latter is not yet the case, as we have neglected orbital motion, for example. The prior could incorporate the direction-dependent variation of the background star number density.

5. Summary

This work has introduced a statistical method that uses multi-epoch astrometry of an imaged exoplanet candidate to compare a co-moving companion model with a chance-aligned field star model. It puts what is commonly referred to as the ‘common proper motion test’ on a probabilistic footing.

Our statistical model enables a quantitative analysis of an arbitrary number of epochs, a task that cannot be achieved effectively through visual inspection. We consider the proper motion and parallax of the host star and the candidate and evaluate the likelihoods under two different models for the candidate: one in which it is a co-moving companion with negligible orbital motion, the other in which it is a member of the field star population. For the latter we built a probabilistic model of the distribution of the proper motions and parallaxes of field stars as a function of magnitude, using a fit to $Gaia$ data in the field of each target star.

We applied our method to a sample of 263 candidates around 23 stars from the B-Star Exoplanet Abundance Study. We first developed a purely proper motion based method, which we then extended to take into account the parallax. This model accommodates the covariance in the astrometry both between the
measurements and across multiple epochs, for both Gaia astrometry and the direct measurements. We identify seven candidates as co-moving companions. Five of these have been identified as real companions in the literature, including the two exoplanets $\mu^2$ Sco b and b Cen(AB) b. The remaining two candidates are priority targets for further investigation.

Our modelling approach is publicly available as an open-source Python package on GitHub, allowing for easy evaluation and visualization of existing and new data. While this work presents an improvement over current practices, there is scope for further improvement. Most useful would be the inclusion of exoplanet orbital motion in the companion model, the incorporation of stellar number densities, and discriminating field stars from exoplanets based on their spectral information.

Acknowledgements. We are grateful for the helpful comments from the referee that improved the paper. This study is based on observations acquired at the European Southern Observatory (ESO) VLT telescope (program 1101.C-025). This publication makes use of data from the European Space Agency (ESA) mission Gaia (http://www.cosmos.esa.int/gaia), processed by the Gaia Data Processing and Analysis Consortium (DPAC, http://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multi-lateral Agreement. This publication also makes use of data from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation.

References


4 https://github.com/herzphi/compass
Appendix A: Mathematical description of the multi-epoch probabilistic model

We describe here in more detail the probabilistic model introduced in Sect. 2.2. In high-contrast imaging, the set of positions of candidates are generally determined relative to the host star, namely

$$\Delta x' = x'_c - x'_s$$
$$\Delta y' = y'_c - y'_s$$  \hspace{1cm} \text{(A.1)}$$

where the subscript $c$ refers to the candidate and $s$ to the host star. The accents on the coordinates $x'_c$ and $x'_s$ indicate that they are ‘true’, that is, without positional measurement uncertainties. The noisy, observed versions are $x_a$ and $x_s$ at the (noise-free) times $t$. The times we use in our models are measured relative to the Gaia epoch of observation of the host star, because we use Gaia data to build our proper motion and parallax distribution models for the field stars.

As the true positions are nonetheless predictions based on noisy parallaxes and proper motions (and propagated in time using Eq. 1), we write the probability density functions of these true positions as

$$P(\Delta x', \Delta y' \mid t, M(\varpi, \mu))$$

where $M$ can be $M_2$ or $M_b$ and the PDF can be written as Gaussians with the variances and covariances between the two positions $\Delta x'$ and $\Delta y'$. For the companion model, these (co)variances are derived from the Gaia data for the host star. For the background model, these (co)variances come from our fit to the distribution of the Gaia parallaxes and proper motions of a set of field stars, as described in Sect. 2.3. The likelihoods for both of these model (which appear in the odds ratio of Eq. 2) are obtained by marginalizing over the unknown true positions ($\Delta x'$, $\Delta y'$) predicted by the model, as follows

$$P(\Delta x, \Delta y \mid t, M(\varpi, \mu))$$

$$= \int P(\Delta x, \Delta y, \Delta x', \Delta y' \mid t, M(\varpi, \mu)) d\Delta x' d\Delta y'$$

$$= \int P(\Delta x, \Delta y \mid \Delta x', \Delta y') \times$$

$$P(\Delta x', \Delta y' \mid \Delta x, \Delta y') d\Delta x' d\Delta y'$$  \hspace{1cm} \text{(A.3)}$$

$P(\Delta x, \Delta y \mid \Delta x', \Delta y')$ is the likelihood of the measured positions given the true positions, and so reflects the noise in the determination of the image centroid. We take this to be a Gaussian with mean ($\Delta x'$, $\Delta y'$) and $2N$-dimensional Gaussian $\Lambda$, where $N$ is the number of measurement epochs. Equation A.3 is a convolution of two Gaussians, which results in another Gaussian.

In the background model $M_b$, $P(\Delta x', \Delta y' \mid t_1, M_b)$ is a Gaussian with mean ($\Delta x'$, $\Delta y'$) and covariance matrix

$$\Lambda' = \begin{bmatrix}
\text{Var}(\Delta x') & \text{Cov}(\Delta x', \Delta y') & \text{Cov}(\Delta x', \Delta x_s') & \text{Cov}(\Delta x', \Delta y_s') \\
\text{Var}(\Delta y') & \text{Cov}(\Delta y', \Delta x') & \text{Cov}(\Delta y', \Delta x_s') & \text{Cov}(\Delta y', \Delta y_s') \\
\text{Var}(\Delta x_s') & \text{Cov}(\Delta x_s', \Delta x') & \text{Cov}(\Delta x_s', \Delta y') & \text{Cov}(\Delta x_s', \Delta x_s') \\
\text{Var}(\Delta y_s') & \text{Cov}(\Delta y_s', \Delta x') & \text{Cov}(\Delta y_s', \Delta y') & \text{Cov}(\Delta y_s', \Delta y_s') \\
\vdots & \vdots & \vdots & \vdots
\end{bmatrix}$$  \hspace{1cm} \text{(A.4)}$$

where terms are given in the order ($\Delta x'$, $\Delta y'$, $\Delta x_s'$, $\Delta y_s'$, ...). The symmetry of the covariance matrix fills in the lower triangle of this matrix.

Eq. A.4. The individual terms of $\Lambda'$ are computed from Eq. 1 to be

$$\text{Var}(\Delta x_i) = t_i^2 \left[ \text{Var}(\mu_{b,x}) + \text{Var}(\mu_{s,x}) \right]$$
$$+ 2t_i s_i(t_i) \left[ \text{Cov}(\sigma_{b,x}, \mu_{b,x}) + \text{Cov}(\sigma_{s,x}, \mu_{s,x}) \right]$$

$$\text{Var}(\Delta y_i) = t_i^2 \left[ \text{Var}(\varpi_{b,y}) + \text{Var}(\varpi_{s,y}) \right]$$
$$+ s_i(t_i)^2 \left[ \text{Var}(\sigma_{b,y}) + \text{Var}(\sigma_{s,y}) \right]$$
$$+ 2n_i s_i(t_i) \left[ \text{Cov}(\sigma_{b,y}, \mu_{b,y}) + \text{Cov}(\sigma_{s,y}, \mu_{s,y}) \right]$$

$$= t_i \left[ \text{Cov}(\mu_{b,x}, \mu_{b,y}) + \text{Cov}(\mu_{s,x}, \mu_{s,y}) \right] + t_i s_i(t_i) \left[ \text{Cov}(\mu_{b,x}, \varpi_{b,y}) + \text{Cov}(\mu_{s,x}, \varpi_{s,y}) \right]$$
$$+ t_i s_i(t_i) \left[ \text{Cov}(\sigma_{b,y}, \mu_{b,x}) + \text{Cov}(\sigma_{s,y}, \mu_{s,x}) \right]$$

$$+ t_i \left[ \text{Cov}(\mu_{b,x}, \varpi_{b,y}) + \text{Cov}(\mu_{s,x}, \varpi_{s,y}) \right] + t_i s_i(t_i) \left[ \text{Cov}(\sigma_{b,y}, \mu_{b,x}) + \text{Cov}(\sigma_{s,y}, \mu_{s,x}) \right]$$

$$+ t_i s_i(t_i) \left[ \text{Cov}(\sigma_{b,y}, \varpi_{b,y}) + \text{Cov}(\sigma_{s,y}, \varpi_{s,y}) \right]$$

$$+ t_i s_i(t_i) \left[ \text{Cov}(\varpi_{b,y}, \mu_{b,x}) + \text{Cov}(\varpi_{s,y}, \mu_{s,x}) \right]$$

$$+ t_i s_i(t_i) \left[ \text{Cov}(\varpi_{b,y}, \varpi_{b,y}) + \text{Cov}(\varpi_{s,y}, \varpi_{s,y}) \right]$$

From the convolution in Eq. A.3, we see that $P(\Delta x, \Delta y \mid t, M_b)$ is a $2N$-dimensional Gaussian with mean ($\Delta x'$, $\Delta y'$) and covariance matrix $\Lambda + \Lambda'$. Equations A.5 assume that there is no covariance between the field star proper motion and parallax and the host star proper motion and parallax. This is not strictly true, because both the background model and the host star data are drawn from Gaia, and Gaia shows correlations between sources separated by small angles on the sky. But in practice our background model describes the distribution of an ensemble of stars, and so is less affected by Gaia’s correlations between individual sources.

In the co-moving companion model $M_c$, the position of the companion relative to the star is constant, leading to a zero covariance for the second term under the integral in Eq. A.3 making $P(\Delta x', \Delta y' \mid t, M_c)$ a delta function. $P(\Delta x, \Delta y \mid t, M_b)$ is therefore a $2N$-dimensional Gaussian with mean ($\Delta x'$, $\Delta y'$) and covariance $\Lambda$.

Appendix B: Candidates with positive log odds ratios

Figures B.1, B.2, and B.3 show the modelling results for the 18 BEAST-only candidates that have logarithmic odds ratio $\log L_{bh} > 0$. The results for the two targets with additional data are shown in Fig. 13.
Fig. B.1: Candidates from BEAST with a logarithmic odds ratio \( \log r_{\text{ab}} > 0 \). The schematics of the plots are explained in Fig. 11.
Fig. B.2: Candidates from BEAST with a logarithmic odds ratio $\log r_{\text{EC}} > 0$. The schematics of the plots are explained in Fig. 11.
Fig. B.3: Candidates from BEAST with a logarithmic odds ratio $\log r_{\alpha, i} > 0$. The schematics of the plots are explained in Fig. 11.