Draft version August 22, 2018
Typeset using LaTeX **twocolumn** style in AASTeX61

DETECTION AND PARTIAL CHARACTERISATION OF STELLAR MULTIPLICITY WITH GAIA

Andrew R. Casey,[1,2] Daniel Foreman-Mackey,[3] Carles Badenes,[4] Rosemary Mardling,[5,6,7,3] John Lattanzio,[5,6,7,3] David W. Hogg,[5,6,7,3] and Adrian M. Price-Whelan[8]

[1] *School of Physics & Astronomy, Monash University, Wellington Rd, Clayton 3800, Victoria, Australia*
[2] *Faculty of Information Technology, Monash University, Wellington Rd, Clayton 3800, Victoria, Australia*
[3] *Flatiron Institute, 162 Fifth Ave, New York, NY 10010, USA*
[4] *Department of Physics and Astronomy, University of Pittsburgh, 3941 O'Hara Street, Pittsburgh, PA 15260, USA*
[5] *Center for Cosmology and Particle Physics, Department of Physics, New York University, 726 Broadway, New York, NY 10003, USA*
[6] *Center for Data Science, New York University, 60 Fifth Ave, New York, NY 10011, USA*
[7] *Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg*
[8] *Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA*

(Received 2018 XX XX; Revised 2018 XX XX; Accepted 2018 XX XX)

Submitted to AAS Journals

ABSTRACT

The frequency and properties of stellar multiples (e.g., binaries, triples) underpin much of astrophysics, and are difficult to estimate. Here we use *Gaia* data to detect and partially characterise XXX stellar multiple systems based on excess jitter in astrometry and radial velocity. We reliably detect binary systems with orbital periods up to $\approx$3.5 yr. We use eclipsing binary systems to calibrate excess astrometric jitter with tangential motion, allowing us to estimate inclination angles and directly constrain companion masses for millions of systems. We find that the distribution of inclination angles is isotropic. A comment on stellar multiplicity with stellar metallicity. A comment on stellar multiplicity in the field relative to clusters. A comment on systems with stellar remnants (e.g., NS/BHs).

Keywords: (stars:) binaries: general, (stars:) binaries: spectroscopic, (stars:) binaries (including multiple): close, astrometry, techniques: radial velocities, methods: statistical

Corresponding author: Andrew R. Casey
andrew.casey@monash.edu

1. INTRODUCTION

A higher order stellar system describes any star that has at least one stellar companion (e.g., binaries, trinaries). The presence of stellar companions complicates inferences on stellar and galactic properties, but these factors are often ignored because it is extremely challenging to detect stellar multiplicity for an unresolved system, and even harder to separate the contributions from individual sources.

The *Gaia* space telescope provides exquisite astrometry, photometry, and radial velocity measurements over many years for million of point sources in our galaxy (**?**). When a point source has an unresolved stellar companion within some range of orbital and stellar parameters, it is expected that there will be an excess in motion jitter relative to a single star of a similar type. This is nothing new: astronomers have used radial velocity (Butler & Marcy 1996) and astrometric variations (Muterspaugh et al. 2010) to infer the presence of stellar and exoplanet companions for decades. However, the *Gaia* data presents an opportunity to infer the presence of stellar companions for millions of point sources across both hemispheres. No ground- or space-based instrument has ever provided such data.

Literature review on what we do and don't know about the multiplicity fraction

In this work we make use of astrometry (including radial motion) and photometry? from the second *Gaia* data release to infer the presence of stellar companions for millions of point sources. We provide partial characterisations of these systems, based on the data available for each source. In Section 3 we describe our methods, and in Section 4 we outline our results in context of other stellar multiplicity surveys. We include comparisons of binary properties between clusters and the field, as a function of stellar properties (e.g., stellar metallicity), and highlight some particularly noteworthy systems that would benefit from additional observations.

2. DATA

There are many ways that binaries can be reliably identified using *Gaia* data, including radial velocity measurements, astrometry, and photometry. These data are sensitive to characterising stellar multiplicity in different ways. Here we describe the data that we use from the second *Gaia* data release (Lindegren et al. 2018) before defining our model.

2.1. *Radial velocities*

Radial velocity measurements are not available for individual epochs in the second *Gaia* data release (Lindegren et al. 2018; Cropper et al. 2018), but a point estimate and an error in radial velocity is available for 7,224,631 sources (Cropper et al. 2018). The reported radial velocity error ($\sigma_{\mathrm{V_R}}^{\mathrm{MTA}}$; column name `radial_velocity_error`) is a function of the number of transits N (i.e., the number of observations; given by column `rv_nb_transits`) and the standard deviation among those measurements. This allows us to calculate the standard deviation in radial velocity for each source, independent of the number of measurements, which we will refer to as the *radial velocity jitter* $\sigma(\mathrm{V_R^t})$,

$$\sigma(\mathrm{V_R^t}) = \sqrt{\frac{2N}{\pi}}\sigma_{\mathrm{V_R}}^{\mathrm{MTA}} \quad . \tag{1}$$

For a single star, or a star without a significant mass companion, the source radial velocity jitter represents the minimum uncertainty with which *Gaia* can measure radial velocity for a star of that colour, apparent magnitude, and absolute magnitude. Stars with bluer colours have fewer absorption lines with deeper wings, and fainter stars have on average lower signal-to-noise (S/N) ratios, both of which result in noisier radial velocity measurements. The absolute magnitude is similarly important, as giant stars have narrow absorption lines than main-sequence stars of the same temperature and colour. For stellar multiples, the source radial velocity jitter will be the quadrature sum of the expected jitter for the primary (if it were a single star) and a contribution based on the orbital parameters and the observation epochs.

2.2. *Astrometry*

Many higher-order star systems are unresolved point sources in *Gaia*, but the stellar companion(s) may have sufficient mass to measurably affect the tangential velocity of the primary and produce a detectable jitter in the astrometric position. There are many other sources of astrometric jitter, but here we show that the principle effect of astrometric excess noise is due to stellar companions. Note that due to the astrometry 'degree of freedom bug' in the processing of the second *Gaia* data release (e.g., see Appendix A1 of Lindegren et al. 2018), that resulted in 80% of sources having zero astrometric excess noise, we instead make use of the astrometric unit weight error u

$$u = \sqrt{\frac{\chi^2_{al}}{N_{obs,al} - 5}} \tag{2}$$

where χ^2_{al} is the astrometric χ^2 value in the along-scan direction (column `astrometric_chi2_al`) and $N_{obs,al}$ is the number of astrometric observations in the along-scan direction (column `astrometric_n_obs_al`). The values of this quantity can be interpreted similarly to how the astrometric excess noise would be interpreted (e.g., large values are less consistent with a single star model), although like the radial velocity and astrometric jitter, it must be considered in context with stars of similar colour, apparent magnitude, and preferably, absolute magnitude.

3. METHOD

We assume that for sources with similar *bp* - *rp* colours, apparent *rp* magnitudes, and absolute *rp* magnitudes, the jitter distribution (in radial velocity or astrometry) is a mixture of two populations:

1. *single-star systems*, where the jitter represents the minimum noise that *Gaia* can measure for a star of that *bp* - *rp* colour, apparent *rp* magnitude, and absolute *rp* magnitude;
2. *higher-order star systems*, where the jitter is significantly above that for single-star systems of a similar colour, apparent magnitude, and absolute magnitude.

We also assume that the mean single-star jitter will change depending on the *bp* - *rp* colour, the apparent magnitude, and the absolute magnitude. For example, the mean radial velocity jitter of single stars with $T_{\rm eff} \approx 6000\,{\rm K}$ will be higher than the jitter of single stars with $T_{\rm eff} \approx 5000\,{\rm K}$, with all else being (nearly) equal. We similarly assume that the mean jitter of stellar *multiples* will depend on the source colour and its position in the Hertzsprung-Russell diagram. For these reasons, in order to evaluate whether a source is more likely to be a single star or a stellar multiple, we must consider the jitter in context with other sources that have similar: *bp* - *rp* colour, apparent *rp* magnitude, and absolute *rp* magnitude. In principle if we knew the distribution stellar multiple properties (or assumed one) then we could model the observability of binary (and higher order) systems across the Hertzsprung-Russell diagram, but this is beyond the scope of this work.

additional assumptions

Given these assumptions, we chose to adopt a non-parametric model for stellar multiplicity, primarily because we do not know the distribution of properties of stellar multiples. Our approach is illustrated in Figure 1 and can be qualitatively described by the following steps.

For each *Gaia* source:

1. Select N sources that have a similar *bp* - *rp* colour, apparent *rp* magnitude, and absolute *rp* magnitude.
2. Fit the distribution of radial velocity jitter for sources selected in Step 1 using a mixture of a normal and a log-normal distribution.
3. Evaluate whether the source of interest is more consistent with being a single star component (drawn from the normal distribution) or a stellar multiple (drawn from the log-normal distribution).
4. If the source is more likely to be a stellar multiple, use the estimated properties of single stars with similar properties to estimate the excess radial velocity jitter due to the stellar companion(s).

STEP 1: SELECTING SIMILAR SOURCES

We construct a k-dimensional tree (k-d tree; **?**) for all sources in the dimensions of *bp* - *rp* colour, apparent *rp* magnitude, and absolute *rp* magnitude. When doing so we scale the dimensions such that 0.1 mags in *bp* - *rp* colour is approximately equal to 1 magnitude in apparent *rp* magnitude and 1 magnitude in absolute *rp* magnitude. We use this k-d tree for selecting a 'ball' of similar stars for a given source.

We enforce a number of constraints when selecting similar stars. For each source we require the multi-dimensional 'ball' to include at least 128 sources. We further require that the radius of the ball extends to at least 0.05 mags in *bp* - *rp* colour, 0.5 mags in apparent *rp* magnitude, and 0.5 mags in absolute *rp* magnitude. If the initial query of 128 points does not extend wide enough to meet our constraints on minimum radius, we extend the radius (linearly in each scaled) dimension until the minimum constraints on radii are met. Our minimum radii constraints imply that there will be many (e.g., $>>$ 128) points returned in dense regions of the Hertzsprung-Russell diagram. In these situations we then subsample the selected points to return a random 1024 sources in any ball, while enforcing our constraints on minimum radii. As a result, the number of 'similar' sources selected will range from a minimum of 128 to a maximum of 1024, and the minimum radii constraints on various dimensions are always maintained.

STEP 2: FITTING THE TWO-COMPONENT MIXTURE

We fit a two-component mixture model to the distribution of radial velocity jitter for all similar sources selected through Step 1. This mixture model includes a normal distribution to represent the intrinsic jitter for single stars, and a log-normal distribution to represent stellar multiples. Specifically the model parameters include a mixing weight parameter θ that is constrained

Figure 1. Illustration of the method we adopt for identifying and characterising stellar multiplicity across the Hertsprung-Russell diagram. For each *Gaia* source we select between 128 and 1024 sources that have similar *bp* - *rp* colours, apparent *rp* magnitudes, and absolute *rp* magnitudes (Step 1). With these similar sources we fit a two-component model to distinguish single star sources from stellar multiples (Step 2). With the model parameters optimised from Step 2, we evaluate probability of stellar multiplicity and partially characterise the higher order systems by assuming binarity and estimating the system semi-amplitude (Step 3).

between $(0, 1)$, the mean μ_s and standard deviation σ_s of the normal distribution, and the mean μ_m and standard deviation σ_m of the log-normal distribution.

We assume a beta prior on θ such that $\theta \sim \mathcal{B}\left(5, 5\right)$ and require that σ_s and σ_m are larger than zero. Depending on the distribution of the radial velocity jitter, we often found this mixture challenging to optimize because the log-normal distribution could easily account for most of the observed distribution, which tended the mixing weight to zero and implied that all selected sources were stellar multiples. To avoid this situation we placed a constraint on the mean of the log-normal distribution to ensure that the *mode* of the log-normal distribution is larger than $\mu_s + 1\sigma_s$ such that

$$\mu_s + 1\sigma_s < \exp\left(\mu_m - \sigma_m^2\right) \quad . \tag{3}$$

Wherever possible we initialised the model parameters from nearby points (within the ball selected by the k-d tree) where an optimised result already exists. This was primarily to minimise computational cost; the initialisation did not have an impact on the result. We required the log probability and the gradient of the log probability to reach machine precision before considering the optimisation converged. After convergence, we queried whether any points within the ball of similar stars had not been optimized. If no result existed, we would next move to that point. We refer to this process as *swarm optimisation*, which we conducted in parallel, and illustrate in Figure **??**.

To minimise computational cost we assigned the closest 8 stars with the result from others

STEP 3: EVALUATING PROBABILITY AND SYSTEM PROPERTIES

Bayes factors for each dimension?

source properties: semi-amplitude (easy)

Can we recover period and eccentricity using The Joker?

3.1. *SB2 Systems: Double-lined spectroscopic binaries*

Many sources in the second *Gaia* data release do not have a reported radial velocity, despite being bright enough ($G \lesssim 13$) and in a suitable temperature range (e.g., between $\approx 4000\,\mathrm{K}$ and $\approx 6500\,\mathrm{K}$) for radial velocities to be measured (Cropper et al. 2018). The principle reason why radial velocity measurements are *not* reported for these stars is because the *Gaia*/DPAC team have identified the source to be a double-lined spectroscopic binary (a so-called SB2-type system), either through a composition of two stellar sources present in the spectra, or from multiple (significant) modes in the cross-correlation function. In these situations it is not sensible to report a point estimate of the radial velocity of the point source. For fainter or bluer sources, however, the radial velocity may not be reported simply because it is too faint, blue, or red.

We calculated the completeness of radial velocity measurements (e.g., the fraction of sources with reported radial velocities) as a function of all available source properties that might affect whether the radial velocity may be reported or not. This included position (α, δ, l, b), parallax, proper motions, apparent magnitudes, *bp* - *rp* colour, properties of the radial velocity templates, stellar parameters (T_{eff}, $R_\circ$, $L_\circ$), and other properties. The full list of properties investigated is described in Appendix **??**, with corresponding figures.

In Figure **??** we show the completeness as a function of some pertinent properties, which demonstrate that the radial velocity completeness is relatively flat until a source becomes too faint (low *rp* flux), or is either too blue or too red. We adopt conservative limits for when the radial velocity completeness starts to drop with these properties, and we assume that any source within the following range of source parameters

$$\begin{aligned} \text{phot rp mean flux} &\in [10^4, 10^8] \\ \text{bp} - \text{rp} &\in [X, Y] \end{aligned} \tag{4}$$

is likely to be a double-lined spectroscopic binary (SB2) if no radial velocity is reported. That is to say that we are explicitly assuming that if a point source meets the criteria in Equation 4 and does not have a reported radial velocity measurement, then the point source is an unresolved double-lined spectroscopic binary. In principle we could make more realistic attempts to model the radial velocity completeness as a function of stellar properties rather than simply stating "sources within this parameter range should have radial velocities unless they are SB2 systems", but the radial velocity completeness within our specified range is reasonably flat.

Some confirmation of this? The properties of SB2 candidate systems relative to others. Figure 3.

We stress that our sample of SB2 type systems is calculated (or rather, deduced) separately to our calculations of multiplicity probability given the *measured* quantities from *Gaia* like radial velocity, astrometry, and photometry. In Section 5 we detail some of the immediate limitations or interpretations of this assumption, and we emphasise that the main conclusions of this work do not depend on SB2 systems. Nevertheless, we *strongly* caution that our deductive inference of SB2 systems will be far more uncertain than the binary probabilities we derive from other information. Our indicator whether a star is likely a SB2 system likely suffers from more contamination and lower completeness compared to the probabilities we derive from other information (e.g., the radial velocity jitter, astrometric excess noise, and photometric colours), and the contamination and completeness likely vary over some complex (unknown) function of parameter space.

4. RESULTS

5. DISCUSSION

- Binary fraction in the spaces that we were fitting: rp flux, colour, etc, just showing the transition of probabilities
- Binary fraction as a function of fitting properties (e.g., colour, absolute RP mag, apparent RP mag)
- Binarity across the H-R diagram
- What are the distributions of orbital parameters of binary systems that we would be able to detect?

5.1. *Comparison with El-Badry et al. (2018)*

5.2. *Comparison with Badenes et al. (2018)*

5.3. *Comparison with Raghavan et al. (2010)*

5.4. *Comparison with Troup et al. (2016)*

5.5. *Comparison with Price-Whelan et al. (2018)*

Figure 2. Fraction of *Gaia* sources with reported radial velocities as a function of source properties. Within the adopted source parameter ranges (gray; see Section 3.1) the radial velocity completeness is approximately flat. We flag sources within this range that do not have reported radial velocities to be candidate SB2 systems.

- Binarity with metallicity
- binarity in clusters vs the field?
- binarity among extremely metal-poor stars?

It is a pleasure to thank Berry Holl (Observatoire de Genéve), Jose Hernandez (ESAC), Daniel Michalik (ESA/ESTEC), Kevin C. Schlaufman (Johns Hopkins University), Lorenzo Spina (Monash University), and Sergey Koposov (Carnegie Mellon University). This work was supported in part by the Australian Research Council through Discovery Grant DP160100637. This work has made use of data from the European Space Agency (ESA) mission *Gaia* (`https://www.cosmos.esa.int/gaia`), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, `https://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* Multilateral Agreement. This research was developed in part at the NYC Gaia DR2 Workshop at the Center for Computational Astrophysics of the Flatiron Institute in 2018 April.

This work has made use of CosmoHub. CosmoHub has been developed by the Port d'Informació Científica (PIC), maintained through a collaboration of the Institut de Física d'Altes Energies (IFAE) and the Centro

Figure 3. Distribution of (a) astrometric unit weight error (see Eq. 2), (b) photometric *bp* - *rp* excess factor, and (c) photometric *rp* variability (see Eq. **??**) of candidate SB2 type systems relative to a control sample of the same size.

de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), and was partially funded by the "Plan Estatal de Investigación Cientfica y Técnica y de Innovación" program of the Spanish government.

APPENDIX

A. REPRODUCIBILITY

This project was developed in a `git` repository that is publicly accessible at `https://github.com/andycasey/velociraptor`. The repository includes notebooks that detail our work, LaTeX to compile this manuscript, and scripts to reproduce our analysis in its entirety. Reproducing this work in full will require at least X Gb of disk space and Y compute hours. To reproduce this work (including data retrieval, analysis, production of figures, and manuscript compilation) use the following commands on a modern terminal:

B. RELEVANT ASTRONOMICAL DATA QUERY LANGUAGE (ADQL) QUERIES

Include relevant ADQL queries

C. RADIAL VELOCITY COMPLETENESS AS A FUNCTION OF *Gaia* SOURCE PROPERTIES

Figures

Software: `Astropy` (Astropy Collaboration et al. 2013), `IPython` (Pérez & Granger 2007), `matplotlib` (Hunter 2007), `numpy` (Walt et al. 2011), `scipy` (Jones et al. 2001–), `Stan` (Carpenter et al. 2017), `CosmoHub` (Carretero et al. 2017), `TensorFlow` (Abadi et al. 2015) `Jupyter Notebooks` (Kluyver et al. 2016)

REFERENCES

Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, , , software available from tensorflow.org

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33

Badenes, C., Mazzola, C., Thompson, T. A., et al. 2018, ApJ, 854, 147

Butler, R. P., & Marcy, G. W. 1996, ApJL, 464, L153

Carpenter, L., Gelman, A., M., H., et al. 2017, Journal of Statistical Software, 76, doi:10.18637/jss.v076.i01

Carretero, J., et al. 2017, PoS, EPS-HEP2017, 488

Cropper, M., Katz, D., Sartoretti, P., et al. 2018, ArXiv e-prints, arXiv:1804.09369

El-Badry, K., Ting, Y.-S., Rix, H.-W., et al. 2018, MNRAS, 476, 528

Hunter, J. D. 2007, Computing in Science and Engineering, 9, 90

Jones, E., Oliphant, T., Peterson, P., et al. 2001–, SciPy: Open source scientific tools for Python, ,

Kluyver, T., Ragan-Kelley, B., Pérez, F., et al. 2016, in Positioning and Power in Academic Publishing: Players, Agents and Agendas, ed. F. Loizides & B. Schmidt, IOS Press, 87 – 90

Figure 4. Fraction of sources (per sky bin) without reported radial velocities for all sources with $G \lesssim 13$ (top) and sources in the ranges specified by Eqs **??** (bottom), where we assert that a missing radial velocity measurement signifies a likely SB2 candidate . The color scale is arbitrarily set to highlight structure, where black indicates a higher fraction of sources do not have reported radial velocities. Crowding in the galactic plane likely results in some sources not having radial velocities reported. The large scale structure visible in both axes is a combined effect of the initial *Gaia* source list, the scanning law, and star forming regions (i.e., where emission in the Ca II triplet likely causes issues for radial velocity determination).

Lindegren, L., Hernandez, J., Bombrun, A., et al. 2018, ArXiv e-prints, arXiv:1804.09366

Muterspaugh, M. W., Lane, B. F., Kulkarni, S. R., et al. 2010, AJ, 140, 1657

Pérez, F., & Granger, B. E. 2007, Computing in Science and Engineering, 9, 21

Price-Whelan, A. M., Hogg, D. W., Rix, H.-W., et al. 2018, AJ, 156, 18

Raghavan, D., McAlister, H. A., Henry, T. J., et al. 2010, ApJS, 190, 1

Troup, N. W., Nidever, D. L., De Lee, N., et al. 2016, AJ, 151, 85

Walt, S. v. d., Colbert, S. C., & Varoquaux, G. 2011, Computing in Science and Engg., 13, 22