Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **517,** 1946–1957 (2022)
Advance Access publication 2022 October 4

https://doi.org/10.1093/mnras/stac2829

Searching for young runaways across the sky

Marina Kounkel [1]★ Aidan Mcbride [2]★ Keivan G. Stassun [1] and Nathan Leigh [3,4]

[1]*Department of Physics and Astronomy, Vanderbilt University, VU Station 1807, Nashville, TN 37235, USA*
[2]*Department of Physics and Astronomy, University of Utah, JFB, 115 1400 E Salt Lake City, UT 84112, USA*
[3]*Departamento de Astronomía, Facultad de Ciencias Físicas y Matemáticas, Universidad de Concepción, Av. Esteban Iturra s/n Barrio Universitario, Casilla 160-C, Concepción, Chile*
[4]*Department of Astrophysics, American Museum of Natural History, New York, NY 10024, USA*

Accepted 2022 September 28. Received 2022 September 28; in original form 2022 August 12

ABSTRACT

We present a catalogue of 3354 candidate young stars within 500 pc that appear to have been ejected from their parent associations with relative speeds of >5 km s^{-1}. These candidates have been homogeneously selected through performing a 2D spherical traceback of previously identified pre-main-sequence candidates to various star-forming regions, ensuring that the traceback age as well as the estimated age of a star is consistent with the age of the population, and excluding contaminants from the nearby moving groups that follow the dominant velocity currents in the field. Among the identified candidates we identify a number of pairs that appear to have interacted in the process of the ejection; these pairs have similar traceback time, and their trajectory appears to be diametrically opposite from each other, or they have formed a wide binary in the process. As the selection of these candidates is performed solely in 2D, spectral follow-up is necessary for their eventual confirmation. Unfortunately, recently released *Gaia* DR3 radial velocities appear to be unsuitable for characterizing the kinematics of low-mass stars with ages <100 Myr, as the accretion, activity, and a variety of other spectral features that make them distinct from the more evolved stars do not appear to have been accurately accounted for in the data, resulting in significant artificially inflated scatter in their RV distribution.

Key words: proper motions – stars: kinematics and dynamics – stars: pre-main-sequence.

1 INTRODUCTION

Typically, young stars do not form in isolation – they tend to form in massive molecular clouds alongside hundreds, if not thousands, of other young stars. Some of them originate in relatively diffuse environments, some are born in dense clusters. Regardless of the overall stellar density surrounding them, many of the youngest stellar objects initially are found in denser 'hubs', consisting of more than two stars resembling proto-multiple systems (Chen et al. 2013). Some of them will go on to form stable binaries and other higher order systems, and some will rapidly dissolve (Tobin et al. 2016, 2022).

When three or more stars are found in close proximity to one another and they do not form a stable orbital configuration (such as a hierarchical triple), orbital energy exchange will occur (Leigh & Geller 2013; Stone & Leigh 2019; Manwadkar, Trani & Leigh 2020; Manwadkar et al. 2021). Such an exchange will typically lead to an ejection of one of the stars from the system if they have finite sizes, and always if they are assumed to be point-particles. The typical ejection speeds are only a few km s^{-1} (Reipurth et al. 2010), which would artificially broaden the overall velocity dispersion of the population in which these stars have been formed (Kounkel et al. 2022), but otherwise not be immediately apparent. The most extreme ejection events can accelerate stars to speeds of several tens of km s^{-1}; such stars are considered walkaway (if their speed is <30 km s^{-1}) and runaway (if their speed is >30 km s^{-1}, e.g. Schoettler et al. 2020), though such boundaries are arbitrary, largely driven by the precision in velocity in the early studies that were needed for a robust confirmation that such stars have been ejected, particularly if a point of origin is uncertain.

One of the most famous pair of runaways are two O stars, AE Aur and μ Col, moving in opposite directions from one another, having been ejected $\sim$2 Myr ago from the Orion Nebula Cluster (ONC) with speeds of $\sim$100 km s^{-1} (Blaauw 1961). A number of other OB runaways have also been identified (Hoogerwerf, de Bruijne & de Zeeuw 2001): as they are bright and fast-moving, proper motions for a number of OB runaways has been readily available for several decades. And, in many cases, they can be traced back to specific clusters (Bhat, Irrgang & Heber 2022) from which they are believed to originate.

However, identifying low-mass ejected stars has presented a greater challenge. An unequivocal confirmation of youth of a star located outside of notable star-forming regions has been difficult to come by. This is necessary to separate likely runaway candidates from ordinary field stars that happen to have similar kinematics – as the latter stars are far more numerous, even a small degree of contamination can overwhelm the sample of runaways. Furthermore, it is only recently, with the release of astrometry from *Gaia* (Gaia Collaboration 2018, 2021), that the sufficiently precise proper motions have become available for a large number of stars.

Since then, there have been several studies that searched for young runaway and walkaway stars, but such studies tended to primarily focus on those that originate from the ONC. In particular, McBride & Kounkel (2019) have found 26 stars with tangential velocities >10 km s^{-1} among the known members of the cluster. Schoettler

★ E-mail: marina.kounkel@vanderbilt.edu (MK); aidan.mcbride@utah.edu (AM)

et al. (2020) have identified 85 candidates younger than 4 Myr that traceback to the ONC, currently located within 100 pc of it. Similarly, Farias, Tan & Eyer (2020) have identified 17 000 candidates within 45° of the cluster that are expected to be young – while most of these sources suffer from contamination, 25 candidates have a particularly high likelihood of having been ejected. Other studies that have identified runaway/walkaway stars in the ONC include Kounkel et al. (2017), Platais et al. (2020), and Maíz Apellániz, Pantaleoni González & Barbá (2021).

Outside of the ONC, no systematic search of young ejected stars has been performed, and all the candidates so far are serendipitous, such as, e.g. Luhman (2018) identifying a dissolution of three stars in Taurus, or Tajiri et al. (2020) finding a fast-moving dipper star.

In this paper, we take advantage of newly available catalogues of likely pre-main-sequence (PMS) stars to search for runaway and walkaway candidates, tracing them back to a variety of star-forming regions within 500 pc. In Section 2, we describe the data that form the basis of this analysis. In Section 3, we present the methodology used to identify them. In Section 4, we discuss these results, and provide conclusions in Section 5.

2 DATA

2.1 Base catalogue

The release of data from *Gaia* has made it increasingly more possible to search for PMS stars with ages of a few to several tens of Myr. A large number of catalogues have taken advantage of the fact that young stars form in large populations that are dynamically cold to identify a large number of young moving groups (e.g. Kounkel & Covey 2019; Kounkel, Covey & Stassun 2020; Kerr et al. 2021; Prisinzano et al. 2022). However, as ejected stars have been dynamically processed, they may no longer share the kinematics with their parent association; as such it is necessary to confirm that a specific star is PMS without relying on its kinematics.

Other studies, such as Zari et al. (2018) and McBride et al. (2021), have performed a photometric selection of young stars along their HR diagram. In particular, McBride et al. (2021) have developed a neural net Sagitta that was trained on the 2MASS and *Gaia* photometry, as well as parallaxes, of the stars that are members of young moving groups from Kounkel et al. (2020).

Sagitta consists of two parts. The first component is a classifier that assigns a probability of a particular star being PMS, separating out the young stars from the stars on the main and binary sequences, as well as from reddened high mass and red giant branch stars. Sources with the PMS probability >0.95 represent the cleanest sample that have only negligible contamination from the evolved stars. At lower thresholds, contamination increases to as much as ∼50 per cent at PMS probability of ∼0.85; however, it becomes more sensitive towards stars with an age of a few tens of Myr, as they are closer to the binary sequence.

The second component of Sagitta is estimating the ages of the stars, using a neural net to interpolate across the empirical isochrones of the full census of members of young moving groups with known ages.

In total, McBride et al. (2021) have identified ∼450 000 PMS candidates in *Gaia* EDR3 data down to PMS probability of 0.7 at distances of up to ∼3 kpc. In this work, we limit the sample to only the sources with PMS > 0.85 (to ensure that the majority of sources would be bona fide PMS stars, despite the contamination), and with parallax $\pi > 2$ mas (to ensure high degree of precision in the distance and tangential velocity of stars, as well as sensitivity to low-mass stars, both of which decrease substantially beyond that limit), resulting in 70 528 stars. Such cuts ensure high degree of completeness across regions included in this volume, a relatively limited degree of contamination, as well as a high precision in the *Gaia* astrometry across all of the candidates to ensure a reliable traceback.

2.2 Initial clustering

In order to identify runaways, it is necessary to evaluate their position and velocity relative to a specific population. At the same time, stars that are members of one star-forming region may systematically traceback to a different star-forming region due to the global dynamics – while there may be some physical process that may relate these populations, such stars are highly unlikely to be runaways in the true sense of the word, i.e. all of their individual stars would not have been ejected through the dynamical interactions. As such, it is necessary to identify all the stars that are the co-moving members of star-forming region, to exclude them from the possible pool of candidate runaways, as well as to define typical positions and velocities to which runaways can track. This is possible to do with hierarchical clustering, such as with HDBSCAN (Campello, Moulavi & Sander 2013; McInnes, Healy & Astels 2017).

However, star-forming regions tend to be extended, often spanning >100 pc, and they may have a significant velocity gradients, which is important to take into consideration. A single position/velocity combination for an entire region (e.g. for Orion, or for Sco Cen) does not offer a sufficient degree of precision in performing such a traceback. But, similarly, if we consider positions and velocities of individual stars within a given star-forming region, then this introduces too much noise in the traceback. Thus, it is necessary to identify smaller but meaningful subgroups within each region.

It is possible to 'tune' HDBSCAN to recover structures of different scales. But, unfortunately, if it is tuned to recover smaller subgroups, it would struggle to pick up stars in the more distributed parts of the star-forming regions that could be easily picked up if a full population is considered. As previously mentioned, clustering is important for rejecting false positives from global kinematics, thus their exclusion is not ideal.

Because of this, we perform a two-step approach to clustering – first to identify large-scale structure, and second to identify subgroups in the identified populations. This is similar to the approach considered by Zari, Brown & de Zeeuw (2019) and Kerr et al. (2021).

The clustering was performed in seven dimensions:

(i) $\arccos(\cos(l))/2$ deg
(ii) $\arcsin(\sin(l))/2$ deg
(iii) $b/2$ deg
(iv) $\log_{10}(1000/\pi)$
(v) $4.74\mu_{l*}/\pi$
(vi) $4.74\mu_{b*}/\pi$
(vii) Age (Myr)/10,

where π is parallax in mas, and μ_{l*} and μ_{b*} are proper motions in l and b, in mas yr^{-1}, converted to the local standard of rest (Schönrich, Binney & Dehnen 2010), with galactic rotation subtracted (see Kounkel et al. 2022). The implemented transformation converts them from proper motion to the tangential velocity space. In order to prevent splitting of groups at $l = 0° = 360°$ boundary, a circular encoding for l is implemented. We note that as the sample has $\sigma_\pi/\pi < 0.1$, the inversion of the parallax to get the distances is an acceptably precise approximation for the purposes of the analysis here. Additionally, while some of the scalings in the above parameters

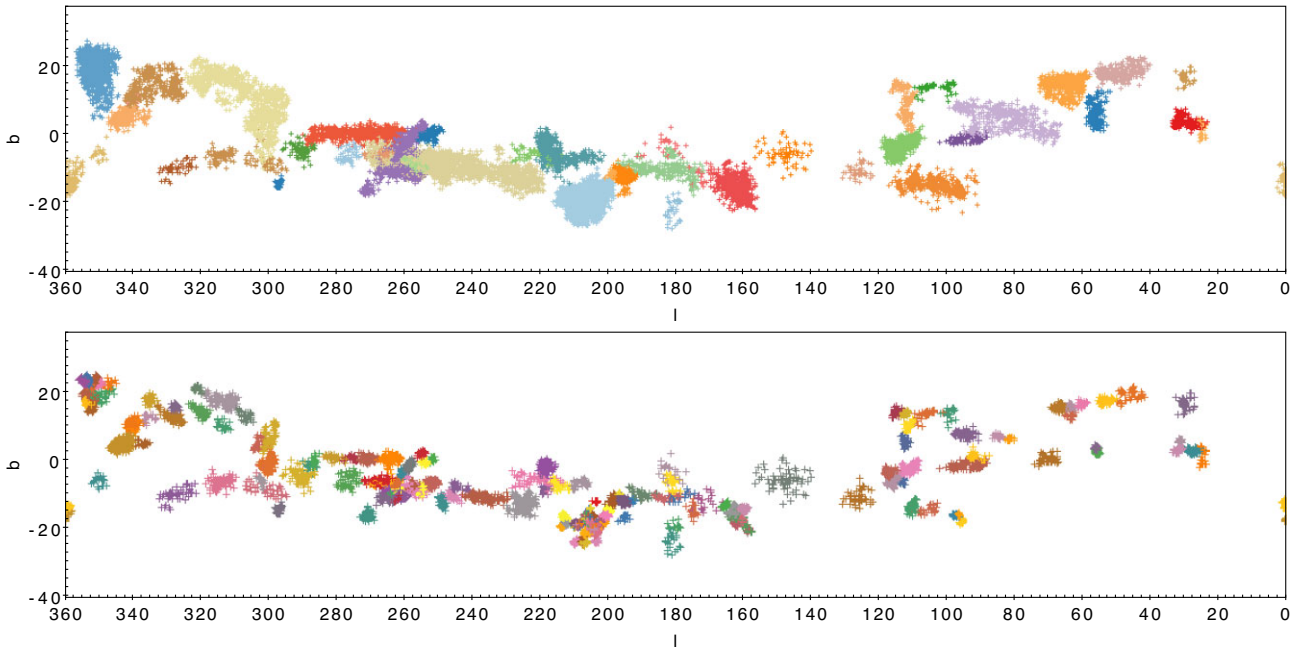

Figure 1. Spatial distribution of the identified structures, arbitrarily coloured by their assigned group. Top: large-scale clustering. Bottom: finer subgroups.

Table 1. Catalogue of clustered PMS candidates.

Column	Unit	Description
Source		*Gaia* DR3 source id
l_*	deg	Galactic longitude
b_*	deg	Galactic latitude
π_*	mas	*Gaia* DR3 parallax
$\sigma_{\pi,*}$	mas	Uncertainty in parallax
PMS		PMS probability from Sagitta
$\sigma_{\rm PMS}$		Uncertainty in PMS probability
t_*	[yr]	(log) Age from Sagitta
$\sigma_{t,*}$	[yr]	Uncertainty in age
μ_{l*}	mas yr^{-1}	Corrected proper motions in l
μ_{b*}	mas yr^{-1}	Corrected proper motions in b
Population		Name of the large-scale population
Subgroup		Name of finer subgroup in a larger population

of the mismatched data units are arbitrary, they were decided on iteratively through the visual examination of the outputs to ensure a homogeneous selection of known populations.

To identify large-scale structure, we used minimum cluster size of 40 stars, with the minimum sampling of 40 stars, and 'eom' as the cluster selection method. This has selected 27 511 (out of 70 528) stars into 42 different populations. Afterwards, we applied HDBSCAN again on the already clustered stars, changing the minimum cluster size to 20 stars, minimum samples of five stars, and 'leaf' as the cluster selection method. This selected 10 451 stars in 167 different subgroups (Fig. 1).

We note that these groups have a considerable degree of similarity with the structures identified in other all-sky clustering approaches (e.g. Kounkel & Covey 2019; Kerr et al. 2021; Prisinzano et al. 2022), but there are minor differences that sometimes prevent a precise one-to-one match. Fundamentally, all clustering techniques trace out similar overdensities in the distribution of stars, but they differ in the ability of recovering particular populations that tend to be more diffuse, as well as in the ability of separating out the neighbouring structures. The primary purpose of repeating the clustering compared to these previous works is to ensure the membership in these identified structures is as complete as possible relative to the base *Gaia* EDR3 catalogue from McBride et al. (2021) in particular, such that stars that may have been excluded in the initial selection in other works would be considered when identifying 'core' members of all the groups. Independent clustering has also enabled us to better control the relative scales of the recovered structures.

To the best of our ability, we have attempted to cross-match all of the identified subgroups to the structures that have been previously recognized in the literature, but it is not always possible. The membership of all of the groups is included in Table 1, and their average properties are shown in Table 2.

3 ANALYSIS

3.1 Selection of candidate ejected stars

In order to identify all of the candidate runaway and walkway stars, as well as their origin, we evaluate the position and velocity of all of the stars that were not identified as members of any of the moving groups relative to the average position and velocity of all substructures. Stars that have been ejected from a particular group would appear to be moving on a radial trajectory away from it in its rest frame.

We perform a traceback solely in the plane of the sky, similarly to Farias et al. (2020) using an approximation of a spherical surface. In such a geometry, a path of a star is described by a great circle.

A bearing θ represents an angle that an object needs to travel along a great circle that would connect a starting location with a particular point on a spherical surface. For each star, we calculate the bearing θ corresponding to its proper motions via[1]

$$\theta = \mathrm{atan2}(\sin(l_\mu - l_*)\cos b_\mu, \cos b_* \sin b_\mu - \sin b_* \cos b_\mu \cos(l_\mu - l_*)). \quad (1)$$

[1] http://www.movable-type.co.uk/scripts/latlong.html

Table 2. Identified groups and their properties.

Column	Unit	Description
Population		Name of the large-scale population
Subgroup		Name of finer subgroup in a larger population
l_g	deg	Average galactic longitude
$\sigma_{l,\,g}$	deg	Scatter in l
b_g	deg	Average galactic latitude
$\sigma_{b,\,g}$	deg	Scatter in b
π_g	mas	Average parallax
$\sigma_{\pi,\,g}$	mas	Scatter in parallax
$\mu_{l*,\,g}$	mas yr^{-1}	Average corrected proper motions in l
$\sigma_{\mu l,\,g}$	mas yr^{-1}	Scatter in μ_{l*}
$\mu_{b*,\,g}$	mas yr^{-1}	Average corrected proper motions in b
$\sigma_{\mu b,\,g}$	mas yr^{-1}	Scatter in μ_{b*}
t_g	[yr]	Average (log) age
$\sigma_{t,\,g}$	[yr]	Scatter in age
N		Number of stars

The galactic coordinates (l_*, b_*) correspond to the current position of the star, and (l_μ, b_μ) are the position of a star that are, respectively, offset by μ_{l*} and μ_{b*} in the reference frame of a particular group defined by that group's typical proper motions $\mu_{l*,\,g}$ and $\mu_{b*,\,g}$, i.e.

$$\begin{aligned} l_\mu &= l_* - (\mu_{l*} - \mu_{l*,g})\Delta t / \cos(b_*) \\ b_\mu &= b_* - (\mu_{b*} - \mu_{b*,g})\Delta t \end{aligned} \tag{2}$$

with Δt set arbitrarily small, to 1 yr, and ensuring appropriate unit conversion.

We also calculate an angular distance d from the star to the median position of each given group (l_g, b_g) as

$$\phi = \mathrm{acos}(\sin b_* \sin b_g + \cos b_* \cos b_g \cos(l_g - l_*)). \tag{3}$$

Finally, we determine the initial traceback position (l_o, b_o) from which a star would have originated having moved along the bearing θ over the angular distance ϕ.

$$\begin{aligned} b_o &= \mathrm{asin}(\sin b_* \cos\phi + \cos b_* \sin\phi \cos\theta) \\ l_o &= l_* + \mathrm{atan2}(\sin\theta \sin\phi \cos b_*, \cos\phi - \sin b_* \sin b_o) \end{aligned} \tag{4}$$

We then establish a set of criteria to evaluate whether a star could have originated from a given group.

(i) $|l_o - l_g| < 3\sigma_{l,\,g}$, $|b_o - b_g| < 3\sigma_{b,\,g}$, where $\sigma_{lb,\,g}$ is the standard deviation in l, b of all the identified members of a group;

(ii) $\phi < 50$ deg (to prevent the approximation of the great circle from breaking down at large angles);

(iii) Relative velocity between the star and a moving group $v_{\rm rel} > 5$ km s^{-1};

(iv) Traceback time $t_{\rm rel} \equiv \phi / v_{\rm rel}$ (ensuring appropriate unit conversion) is less than the age of a star (t_*);

(v) Traceback time is less than the typical age of a group (t_g);

(vi) $|t_* - t_g| < 3$ Myr, OR $|t_* - t_g| < (\sigma_{t,\,g})$, where $\sigma_{t,\,g}$ is the standard deviation of ages of all of the identified members of a group (3 Myr was chosen to ensure that very young stars would be recovered if they have been ejected from a cluster like the ONC which has sustained its star formation for a few Myr);

(vii) $|1000/\pi_* - 1000/\pi_g| < 100$ pc, where π_* and π_g are the parallaxes of a star and of a group (since RVs are typically not available, this is an arbitrarily small distance that should be possible to be traversed by an ejected star in a few Myr, within errors);

(viii) A star is not already a member of any of the moving groups.

Table 3. Catalogue of candiate runaway and walkaway stars.

Column	Unit	Description
Source		*Gaia* DR3 source id
l_*	deg	Galactic longitude
b_*	deg	Galactic latitude
π_*	mas	*Gaia* DR3 parallax
$\sigma_{\pi,*}$	mas	Uncertainty in parallax
PMS		PMS probability from Sagitta
σ_{PMS}		Uncertainty in PMS probability
t_*	[yr]	(log) Age from Sagitta
$\sigma_{t,*}$	[yr]	Uncertainty in age
μ_{l*}	mas yr^{-1}	Corrected proper motions in l
μ_{b*}	mas yr^{-1}	Corrected proper motions in b
Subgroup		Name of finer subgroup in a larger population
θ	deg	Bearing of a stellar motion relative to the subgroup
ϕ	deg	Current separation between the star and the subgroup
v_{rel}	km s^{-1}	Relative velocity between the star and the subgroup
t_{rel}	Myr	Estimated traceback time to the subgroup
l_o	deg	Estimated galactic longitude of the point of origin
b_o	deg	Estimated galactic latitude of the point of origin
ξ_1		Angle-based metric for moving group contaminants
ξ_2		Distance-based metric for moving group contaminants
Wide		Index of wide binary candidates
Opposing		Index of opposing pairs candidates

The catalogue of all of the identified candidate runaway and walkaway stars is presented in Table 3

3.2 Remaining structure identification

The initial clustering analysis of the catalogue has allowed to exclude a number of false positives runaway candidates through limiting a sample of stars that traceback to other moving groups due to dominant global kinematics in the region. However, a significant fraction of runaway candidates is still dominated by such sources. The primary source of contamination are very sparse moving groups (e.g. Taurus), comprehensive membership of which is difficult to recover with clustering in a presence of structures that are much more massive. Additionally, older groups that are in the process of dissolving have members that may be missed due to these stars having kinematics just outside of the typical velocity dispersion, or have positions just outside of where the bulk of the stars in the moving group still reside. Such stars still follow the typical dominant velocity currents of a region in which they reside, but they may not necessarily be identified among the 'core' members of a moving group.

The ability to trace young stars in bulk to other regions may help to reveal the processes that have shaped the velocity currents in a region. However, when several stars are found in a similar direction and at a similar distance away from a particular group, such sources are unlikely to have been ejected from said group. To evaluate the probability of a given source being a viable ejected star candidate, we derive two metrics: ξ_1 and ξ_2.

$$\xi_1 = n_{\Delta\theta<10}/n_{tot}, \tag{5}$$

where $n_{\Delta\theta\,<\,10}$ is the number stars with the bearing θ within 10° of θ of a given star, and $n_{\rm tot}$ is the total number of stars that traceback to a given subgroup. The more concentrated the distribution of stars is originating from a particular direction, the larger ξ_1 becomes, signifying a possibly unaccounted moving group. If $n_{\rm tot} < 12$, then we set $n_{\rm tot} = 12$ – this is to prevent extreme cases of e.g. if only a

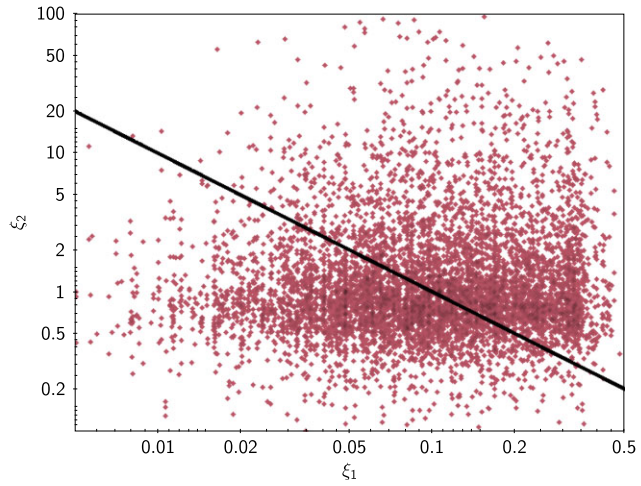

Figure 2. Distribution of ξ_1 and ξ_2 parameters for all of the 6803 candidates (see the text for definition). The black line corresponds to $\xi_1\xi_2 < 0.1$.

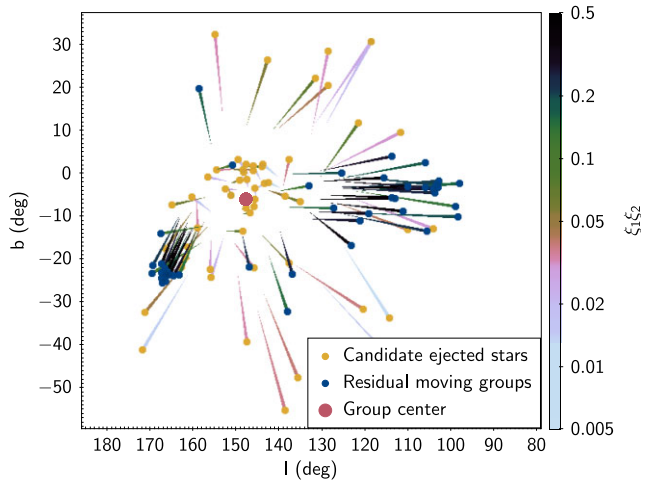

Figure 3. Candidate ejected stars that traceback to one group (α Per), colour-coded by $\xi_1\xi_2$. The sources with large $\xi_1\xi_2$ reveal the presence of two moving groups in this region that have not been recovered in the initial clustering.

single star traces back to a particular subgroup, ξ_1 becomes 1.

$$\xi_2 = \sigma_{\phi_\theta}/|\overline{\phi_\theta} - \phi| \quad (6)$$

if $n_{\Delta\theta<10} > 5$, ϕ_θ are the angular distances of all stars within 10° of θ of a given star, otherwise, if $n_{\rm tot} > 5$, angular distances of all stars tracing back to a given group are used. Average $\overline{\phi_\theta}$ and the standard deviation σ_{ϕ_θ} are computed using only 10–90 percentile range of the distribution of ϕ_θ. This tests for the commonality of angular distances that different stars traceback over: the greater the similarity, the larger ξ_2 becomes. If $n_{\rm tot} \leq 5$, ξ_2 is set to 1.

Conservatively, we set the threshold of $\xi_1\xi_2 = 0.1$ (Fig. 2), separating members of possibly unaccounted of moving groups and possible runaway/walkaway stars, chosen through examining all stars tracing back to individual groups (e.g. Fig. 3), which roughly filters out all of the overdensities in the distribution of sources that are apparent to the eye.

To ensure that neighbouring groups do not 'shield' each other, preventing sources to trace towards them from a particular side, initially, a star could be matched to multiple subgroups. Following calculation of $\xi_1\xi_2$, only the group that is located the closest to a given star is chosen. In total, this selection has identified 6803 candidate runaway and walkaway stars, of which 1266 have PMS probability >0.95. Of these, 3449 have been excluded by $\xi_1\xi_2$ test, with only 3354 stars remaining at PMS probability >0.85, and only 546 at PMS probability >0.95 (Fig. 4).

4 DISCUSSION

4.1 Radial velocity

Currently, the traceback of motion of the ejected candidates is done solely in the plane of the sky. A full 3D traceback would be able to more definitively confirm the point of origin of these fast-moving PMS stars, as RVs are able to affect the overall trajectory of a star away from following a great circle.

The RV data for PMS stars remain very sparse. The recent release of *Gaia* DR3 has provided RVs for over 30 million stars (Katz et al. 2022), representing a major achievement in providing comprehensive data of the radial velocity structure of the Galaxy. But, the sources for which RVs are available tend to be very luminous, disfavouring low-mass PMS dwarfs. Only ∼10 per cent of the base catalogue of PMS stars has reported *Gaia* DR3 RVs, and their uncertainties tend to be 5–10 km s^{-1}. Such data are insufficient to even establish an RV reference frame for most of the populations, let alone to perform a detailed traceback of all of the individual stars fully in 3D.

Furthermore, evaluating the available *Gaia* RVs for PMS stars paints a somewhat grim picture regarding their quality. We compare these RVs versus those that have been measured by APOGEE in various star-forming regions, such as Orion, Taurus, Perseus, and NGC 2264 (Kounkel et al. 2019). We limit the sample to only those sources for which at least three high-resolution spectroscopic observations exist to ensure RV stability, excluding any of the SB1s, SB2s (or higher order multiples) that have been identified in that work.

While the APOGEE RVs of these young stars tend to have low velocity dispersion of less than a few km s^{-1}, same sources result in the velocity dispersion of several 10s of km s^{-1} with *Gaia* RVs. This scatter is not accounted for in the uncertainties: even though the the typical reported errors in *Gaia* RVs are 5–10 km s^{-1}, the difference between *Gaia* and APOGEE RVs on the order of 10σ is common (Fig. 5).

Class II young stellar objects (YSOs), i.e. sources that still have infrared excess due to the presence of a protoplanetary disc, sources that typically have strong accretion signature in optical spectrum have RVs that are qualitatively worse than in the sources that have already depleted their disc. Class III objects have poor RVs as well, however. These sources still have low surface gravity, and often may have strong activity features that may have been unaccounted in the spectral fitting. As such, *Gaia* RVs of nearly all young stars are unstable.

This issue persists for >100 Myr. For example, in Pleiades, where almost all of the stars are already on the main sequence, but many still show high activity. While the solar-type stars ($T_{\rm eff} \sim 6000$ K) have reasonable RV precision, with the scatter in RVs reproduced by the uncertainties (typically <2 km s^{-1}), the cooler dwarfs in Pleiades (∼1/4 of the sample) also have RVs with unrealistically large scatter. More evolved field dwarfs in the APOGEE data that are not associated with any young cluster or a moving group do not appear to be affected to a similar degree, even in comparison to the Pleiades. As such, this issue appears to affect younger and active stars specifically.

As our sample consists entirely of young cool dwarfs, we chose not to use *Gaia* RVs even in the cases where these sparse data are available. Furthermore, we advise extreme caution to other

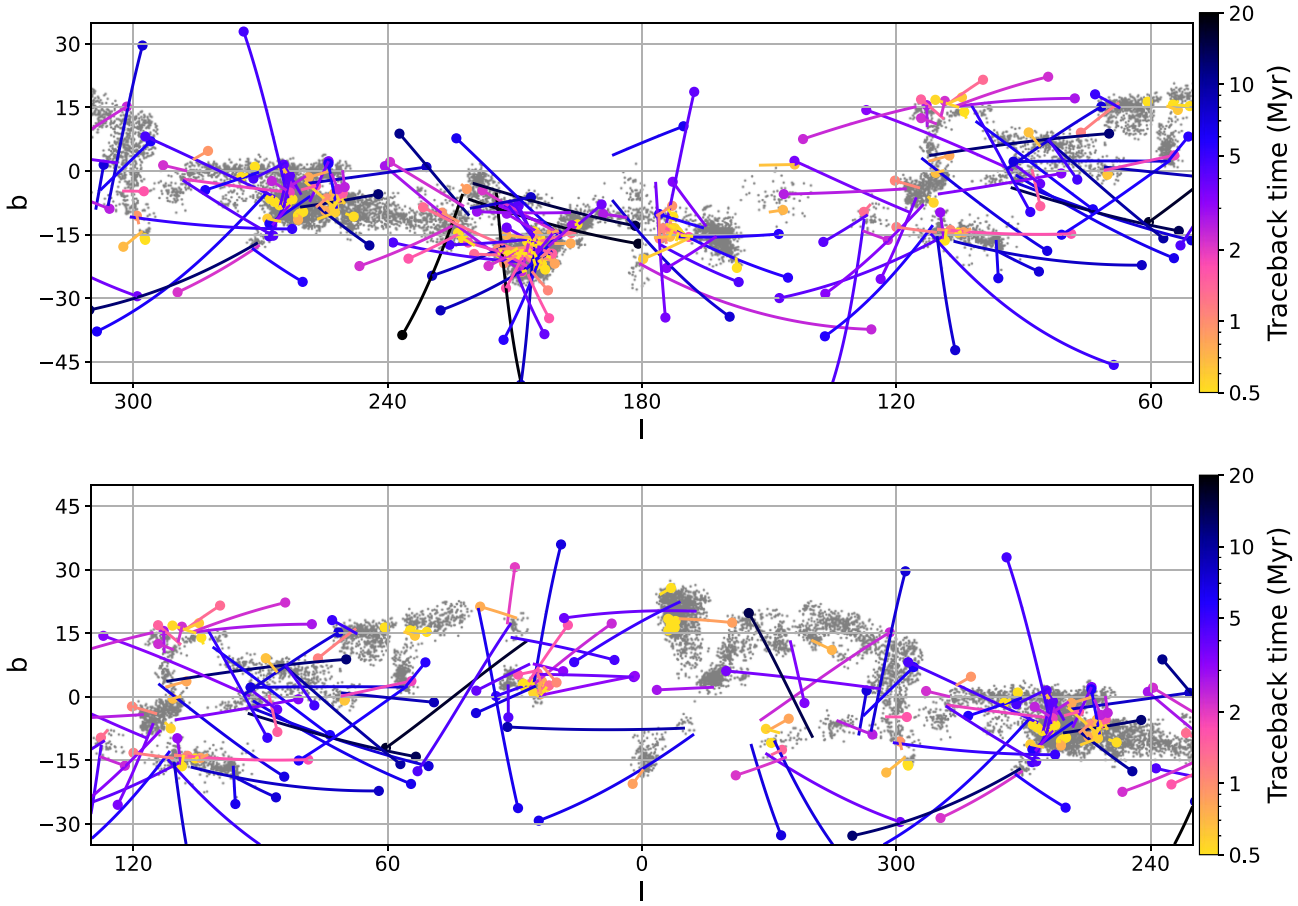

Figure 4. Candidate ejected stars with PMS > 0.95 and relative velocity >10 km s^{-1}. The vectors show their apparent path from the estimated l_o and b_o to their current position (signified by larger dot); they are colour-coded by the traceback traveltime. The greyscale background shows the distribution of the clustered structures.

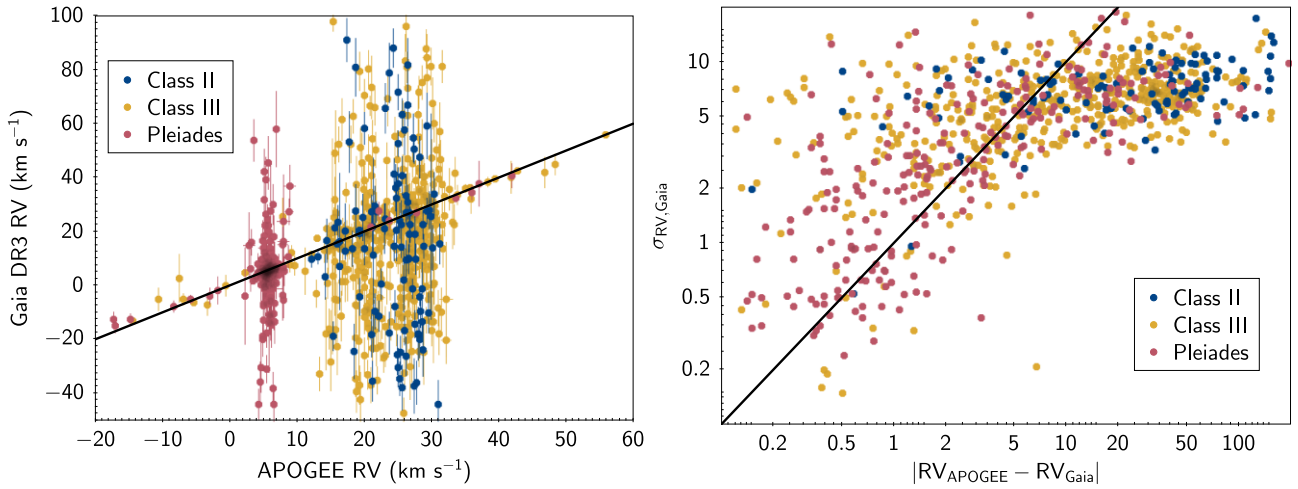

Figure 5. A comparison of the *Gaia* DR3 and APOGEE radial velocities of young stars that appear to have RVs from Kounkel et al. (2019), from Orion, Taurus, Perseus, and NGC 2264, separated into the disc-bearing and non-disc-bearing stars. The data for the Pleiades are also shown to highlight the differences that the age of a star may create in the sample. The left-hand panel shows the direct comparison of the two sets of RVs. The right-hand panel shows the magnitude of the difference between the RVs as a function of the reported uncertainty in RV from *Gaia* DR3.

studies that attempt to overinterpret the observed features of the RV distribution in the young stars (such as Zucker, Peek & Loebman 2022) – the improper data processing of *Gaia* spectra provides yet another source of RV scatter that needs to be considered carefully in determining such parameters as e.g. velocity dispersion of young moving groups.

Unfortunately, at the moment, no other RV survey provides a comprehensive coverage of data across all of the candidates. This will

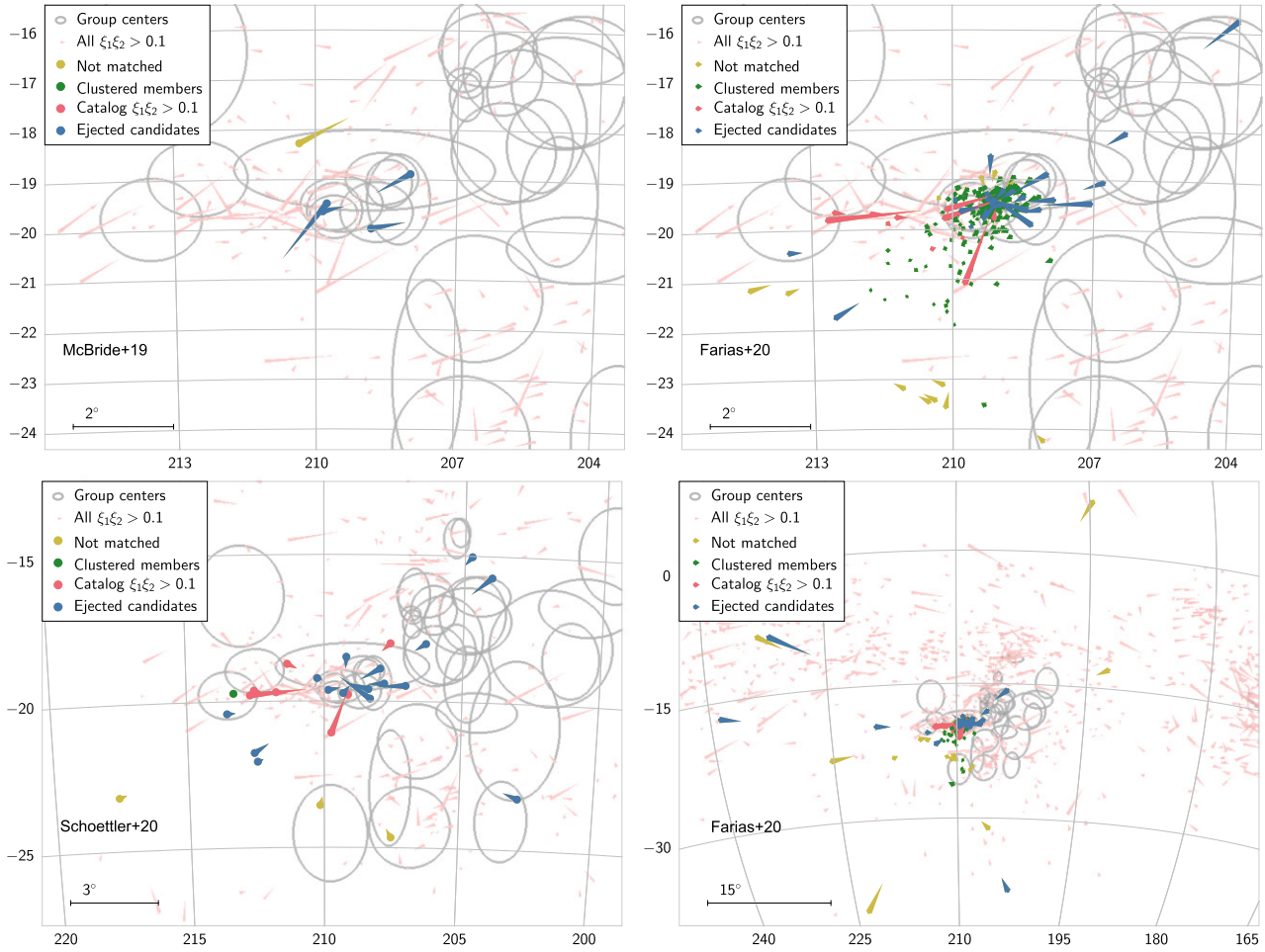

Figure 6. A comparison of the trajectories of the ejected candidates in the works of McBride & Kounkel (2019), Schoettler et al. (2020), and Farias et al. (2020) that can be matched to the base catalogue of PMS stars. Sources are colour-coded based on whether the stars have met the thresholds to be selected as runaway/walkaway stars that we use in this work. The pale pink vectors show the trajectories of all sources that have $\xi_1\xi_2 > 0.1$ presented here, to show the dominant velocity currents in the field when evaluating the identified candidates in each catalogue. The grey circles show the area on the sky corresponding to $3\sigma_{lb,\,g}$ for each of the subgroups in Orion. The plots are shown in the galactic coordinates.

eventually change: over the next 5 yr, SDSS-V is set to obtain high-resolution spectra ∼100 000 young stars (Kounkel, in preparation), however, these data are not yet available. It is also possible that in the future a more careful independent processing of the spectroscopic data released by *Gaia* will enable a more optimal RV extraction of the young stars – once RVS spectra from *Gaia* are released in full, or in the subsequent data releases.

Ultimately, a comprehensive confirmation of the candidate ejected stars, fully in 3D, may not be possible now, but such an analysis will be possible at a later date.

4.2 Comparison to previous works

Previously, McBride & Kounkel (2019), Schoettler et al. (2020), and Farias et al. (2020) have searched for stars ejected from the Orion Nebula. We compare the identified candidates from these works to the catalogue we derive here (Fig. 6).

McBride & Kounkel (2019) have identified 26 stars among known members of the ONC that have high proper motions. Of these, due to limits in colour imposed by Sagitta (typically, the recovered YSO candidates have $G_{BP} - G_{RP} > 2$ mag), and due to the quality of photometry that has been required in the creation of the base catalogue, only five of these 26 sources are included in the analysis. Of these five sources, two appeared to traceback to the centre of the ONC, two appeared to traceback away from the ONC as visitors to other clusters in the Orion Complex, and one had its origin unknown. In total, only one of them (a visitor, *Gaia* EDR3 3017044689550345856) does not meet the initial set of criteria due to a slight mismatch in age of a star (∼0.5 Myr) and the traceback time (∼0.6 Myr) – we note that the uncertainty in age is not considered for the purpose of this exercise. The remaining four stars do meet the initial selection criteria and indeed can be traced back to various parts of Orion.

Schoettler et al. (2020) have identified 85 candidates that can be traced to the ONC from up to 100 pc away from it. Of these, only 27 can be matched to the base catalogue from McBride et al. (2021), again, usually due to the colour limits. Of these 27 stars, one has been selected as a core member of the Orion A molecular cloud. Three stars do not meet the traceback criteria, due to the traceback time being longer than the age of a star and/or group. Seven stars have $\xi_1\xi_2 > 0.1$, and, by the strictest set of criteria, 16 sources can be recovered as runaway/walkaway candidates.

Finally, Farias et al. (2020) have identified ∼17 000 candidates. The sources have been initially selected using a variety of different tracers of youth, including infrared excess, optical variability, position on the HR diagram, and others. Some of these criteria have a

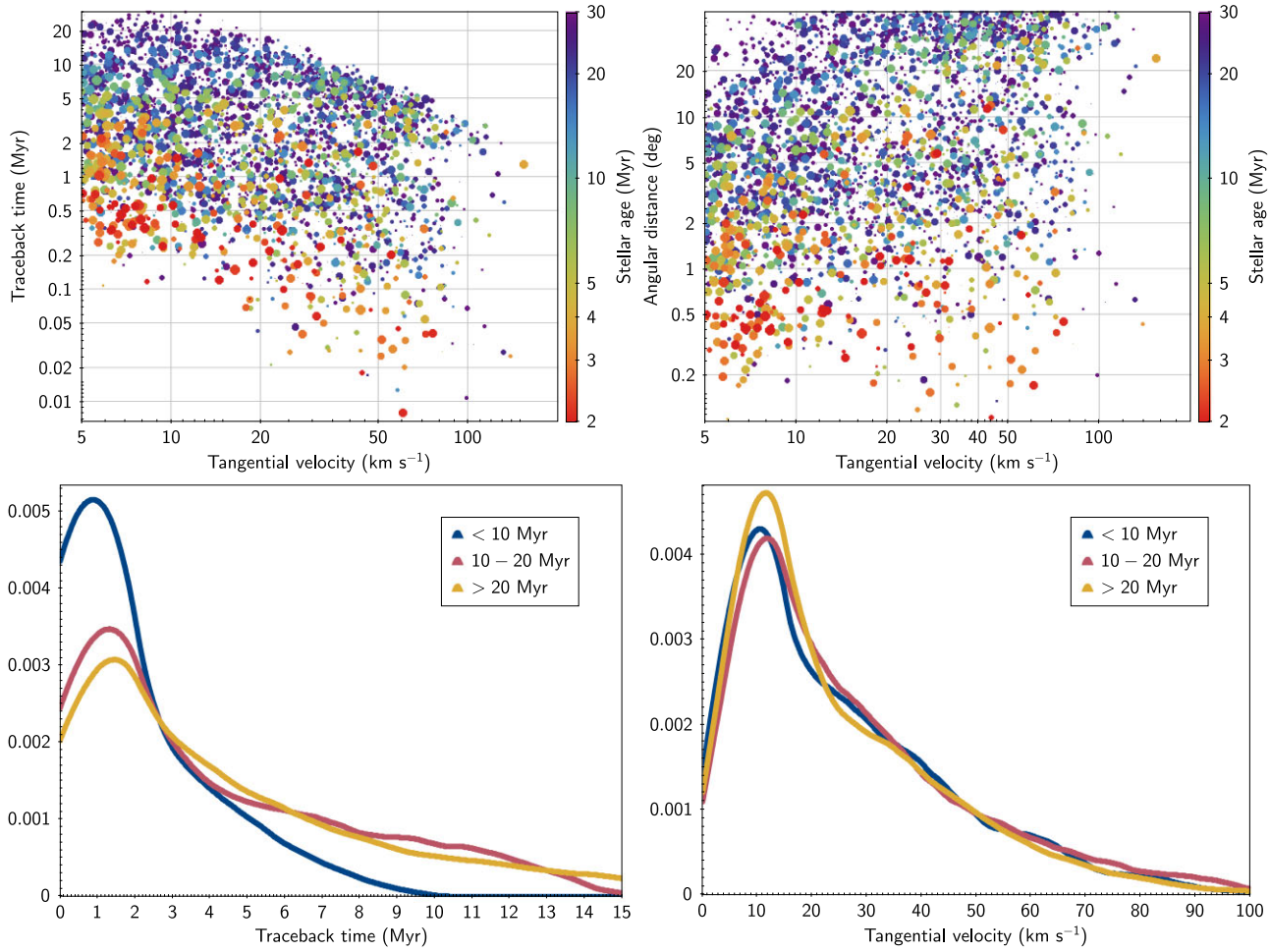

Figure 7. Top: correlation between the projected tangential velocity, traceback time, and the angular distance for the candidate ejected stars with $\xi_1\xi_2 < 0.1$, colour-coded by their age. The size of the symbol corresponds to the probability of a star being PMS, with the larger circles corresponding to the higher confidence sources, and smaller circles having a greater degree of contamination from the field stars. Bottom: kernel density estimate showing a distribution of tangential velocity and traceback time for the sample, separated into three different age bins.

significant degree of contamination from the more evolved stars, as such, only 331 can be cross-matched against our base catalogue. Of these, 267 can be recovered as core members of Orion. 22 stars do not meet traceback criteria, commonly because the conversion to the local standard of rest reference frame has significantly affected their trajectory in comparison to the heliocentric reference frame due to the large separation in distance between a star and the cluster, some due to the traceback time exceeding the age of a star and/or cluster. Of the remaining 42 stars, 11 have $\xi_1\xi_2 > 0.1$, and 31 meet the strictest set of criteria as walkaway/runaway candidates.

In total, we identify 249 ejected star candidates with $\xi_1\xi_2 < 0.1$ that can be traced back to the Orion Complex, of which 117 appear to originate from the ONC. Of these, 36 and 11 stars respectively have tangential velocity >30 km s^{-1}, and they can be considered as runaways.

4.3 Statistics

We examine the properties of the candidate ejected stars in Fig. 7. Across all ages of the stars, the candidates that the selection is most sensitive to are the sources that appear to have been ejected recently, within the last 2 Myr. Older stars have a longer tail in the distribution of their traceback times, as the sources that have been ejected early in their formation would have more time to travel further outwards, but they may be more difficult to detect with the methodology used here.

The overall projected tangential velocity distribution of the identified candidates does not appear to vary significantly between the older and the younger stars. As such, the angular distance between the star and its projected point of origin correlates strongly with their tangential velocity, with the bulk of the sources being within 10 deg of their parent group, and the older sources having a slightly more pronounced excess of sources at larger distances than the younger sources.

Restricting the selection to the runaways with the relative tangential speeds >30 km s^{-1} favours the sample with more recent ejections, as they would quickly disperse away from their parent population, and they may be more difficult to recover without a traceback that is fully in 3D, since the approximation of the spherical geometry may not be sufficiently precise from identifying them at larger distances.

4.4 Ejected pairs

4.4.1 Selection

In the process of disrupting a multiple system, all the stars will undergo acceleration, the precise magnitude of which depends on

both the initial configuration of their orbits as well as the masses of all stars. As such it is possible to eject two stars simultaneously, each travelling away from their parent population in opposite directions from each other, similarly to AE Aur and μ Col (Blaauw 1961; Hoogerwerf et al. 2001).

For most ejected stars, finding their precise pair may be a difficult task, as most such encounters result in low ejection velocities (Reipurth et al. 2010), not easily distinguishable from their parent association. In the process of mixing with their neighbours, the initial trajectory of the sources with low ejection velocity may be erased. This may be particularly common for systems with low mass ratio, in which the more massive companion is more likely to remain in a cluster, ejecting a lower mass neighbour.

As such, the probability of both stars being ejected with sufficiently high velocity for them to be recovered in the catalogue presented here is expected to be relatively low. None the less, there do appear to be certain pairs of candidates that may have a common origin.

To identify such pairs, we require

(i) Using the bearing θ_i that has been inverted, measured from (l_o, b_o) to (l, b) (to prevent the distortion from a spherical geometry that may occur using θ using the current position of a star as a point of origin), the difference in θ_i between the pair of stars should be in the range 175–185°, i.e. they are moving in near-opposite directions within an arbitrarily small precision that would ensure recovery of AE Aur and μ Col.

(ii) The traceback time for both stars should be within 10 per cent of each other, to ensure a probability of a simultaneous event, within uncertainties.

(iii) The distance between their initial traceback positions (l_o, b_o) is <2°, or <10 per cent of their distance at their current positions (l, b), whichever one is smaller.

(iv) Although currently RVs of the candidates are in large part not available, it is possible to evaluate the distances of the candidates to reject obvious false positives. If two stars are travelling in opposite directions, and they originate from the same star-forming region, the distance to their parent population has to be between the distances of both stars, within the tolerance for the uncertainties and the depth of the region.

(v) Both stars have $\xi_1\xi_2$ <0.1.

In total, 42 pairs satisfy these cuts; they are shown in Fig. 8. There may be chance coincidences in this sample, consisting of pairs that only appear to move opposite of one another, and that do indeed originate from the cluster, but would not necessarily traceback to the same system. However, repeating the above selection for angles other than 175–180° results in a typical number of 19 systems that are selected, i.e. there is a significant excess in the number of systems moving opposite of one another (Fig. 9). We further note that if we change some of the criteria – e.g. keep all of the requirements regarding the traceback to a given cluster the same, but force the minimum separation between stars at the initial position to be >2°, there is no preferred direction of motion between these random matches. As such, even though this sample of the opposing pairs may be contaminated by chance coincidences, we expect a subset of these systems to truly be simultaneous ejections.

Additionally, it may be possible for the two of the ejected stars to form a binary system in the process of the encounter. Compact binaries can be detected in the future through spectroscopic follow-up; however, in wide binaries, both stars may be fully resolved with *Gaia*, in which case, both stars are expected to have very similar kinematics, and have very similar trajectories. An example of such a system are Brun 259 and V1961 Ori (McBride & Kounkel 2019), two stars with separation of ∼7000 au that appear to have been concurrently ejected from the ONC 0.1 Myr ago. This system is notable, as binary systems with such wide separations are not expected to be stable in such a dense cluster. We use criteria similar to the above to identify new candidates:

(i) θ of both stars is within 5 deg of each other.

(ii) The traceback time for both stars should be within 10 per cent of each other.

(iii) Their projected separation is within 10 pc.

(iv) Their parallaxes are consistent with 10 pc separation, within errors.

(v) Both stars have $\xi_1\xi_2$ <0.1 and ξ_1 <0.1, this cut is stricter, to further minimize any contamination from the sources that may belong to various moving groups, however distantly.

This results in a catalogue of 19 candidates, shown in Fig. 8. Of these, six have projected separation within 1 pc. Recently, El-Badry, Rix & Heintz (2021) have produced a catalogue of wide binary candidates across the entire sky, containing >1 million pairs with 3D <1 pc. This catalogue is able to recover five of these six stars. We note that cross-matching against full sample sources that meet our initial set of criteria results produces 697 stars in common; however, there is a significant contamination in this sample from various moving groups and cases of clear mismatch in age of the stars in the pair, such as PMS star and a much more evolved counterpart, which are unlikely to be true binaries. As such the selection of 19 pairs is more conservative, however there may be other systems in the sample that are missed by our selection.

4.4.2 *Properties*

We examine the properties that these stellar pairs have. To estimate masses of the individual stars, we cross-match the catalogue of the identified systems against TESS Input Catalog (Stassun et al. 2019), which has an estimate of mass through photometry. We note that these young stars are moving almost completely horizontally on the HR diagram along the Hayashi tracks, as such minor discrepancies in the age of the star does not have a substantial effect on the mass.

Comparing the mass ratio q within each pair, we find that the opposing pairs tend to have q consistent with being uniform (with the detection bias against identifying low q systems, due to the typical masses of the stars recovered in the base catalogue). However, wide binaries show a preference towards unity, almost half of the systems have $q > 0.9$, with a significant deficit of systems with q in the range 0.7–0.9 in comparison to the opposing pairs (Fig. 10). There may be a slight trend of wide binaries to also be increasingly more common at lower q, but due to the small number of such systems, and due to the aforementioned detection bias it is difficult to definitively state it.

We also examine the tangential velocity ratio between the lower mass and higher mass star in the system (Fig. 10). We find that in wide binaries, by construction of the catalogue, velocities of the pairs are very similar, however the lower mass star tends to have slightly higher speed than the higher mass one: typically $v_2/v_1 \sim 1.05$. This is unlikely to be caused by the orbital motion. For a system with the separation of 1 pc, the orbital period is ∼100 Myr, increasing to 3 Gyr for separations of 10 pc, i.e. much longer than the age of these stars. Rather, any orbit between these stars is likely to be unstable, and they are unlikely to necessarily be gravitationally bound. While they most likely have been ejected in the same event, conservation of angular momentum has imparted a slightly lower mass star greater

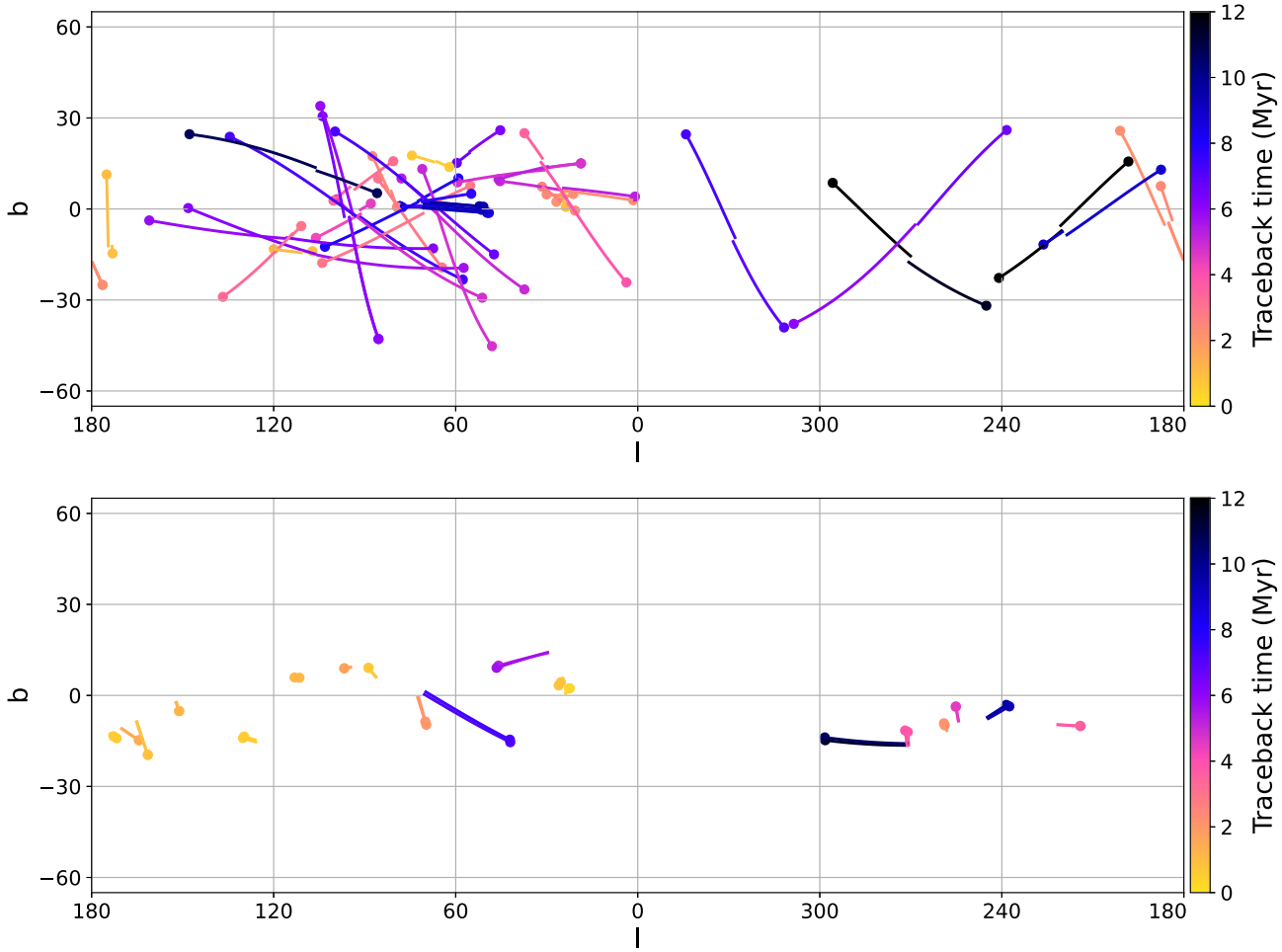

Figure 8. Top: spatial distribution of the candidate interacting pairs of stars that appear to have been ejected concurrently from the same location, travelling in opposite directions. Bottom: candidate wide binary systems that may have formed in the process of the ejection (note, due to their proximity, vectors for both stars may overlap). The current position of each star is highlighted by a large circle. The vectors show their apparent path from the estimated l_o and b_o to their current position (signified by larger dot), they are colour-coded by the traceback traveltime.

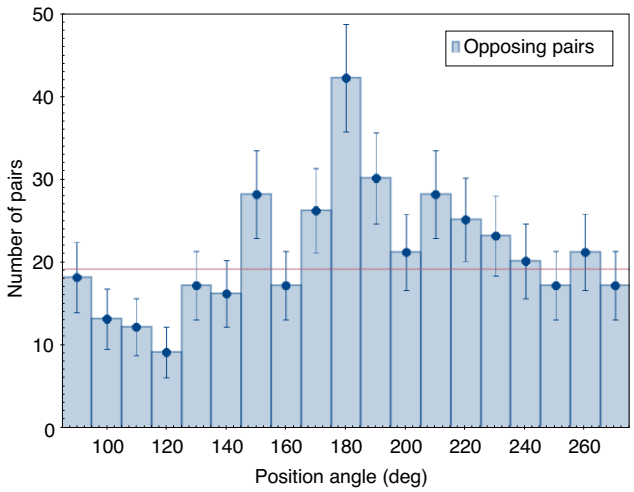

Figure 9. The number of system that are selected using the criteria for the opposing pairs, but using different requirement for the angle. The typical number of systems that appear to move relative to one another at random angles, but project back to a similar origin within a similar period of time is 19 pairs (shown as red line), there is a significant excess at the angle of 180 deg, corresponding to the stars moving in the opposite directions.

velocity, and they will continue to grow further apart. The only reason why they can be observed as a pair is due to their youth, since their velocities have not evolved significantly post ejection. Indeed, on average, these wide binaries tend to be on average younger than the full sample of candidate ejected stars, and they tend to be marginally slower, as they are more likely to lose coherence the further they are from their parent population (Fig. 11).

An opposite trend is observed among the opposing pairs: as the precise location of the ejection within a region is unknown, at larger distances minute deviations in the initial angle of ejection become less apparent. As such to get further out, these opposing pairs tend to be somewhat older and somewhat faster moving than the full sample (Fig. 11). And, surprisingly, a more massive star tends to have a slightly faster velocity than a lower mass one (Fig. 10). Pairs selected through similar set of criteria, but moving relative to each other at any angle other than 180°, tend to have a more symmetric distribution of velocities between v_2 and v_1 (Fig. 9). The excess in the opposing pairs is slight; however, and due to a small number of pairs, the statistical significance is low.

None the less, we consider possibilities of how such an excess of faster moving massive stars in the pair may arise. One of them is that initially, both stars may have belonged to an unstable quadruple system. Initially, the least massive star in the system may have been

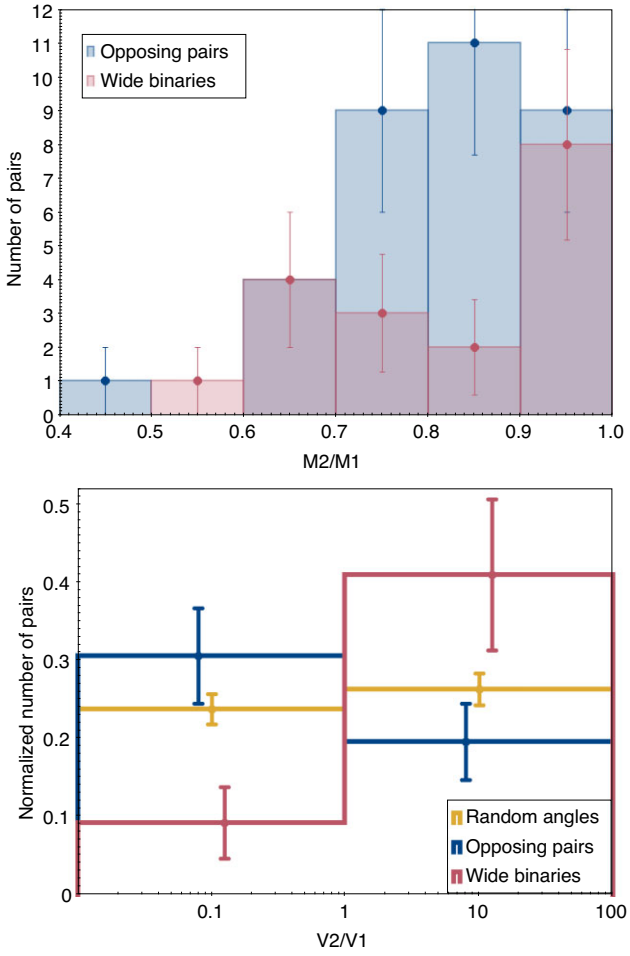

Figure 10. Top: the comparison of mass ratios between the lower and higher mass star in the system of the opposing pairs, and among the wide binaries. Bottom: the comparison of the tangential velocity ratios between lower mass and higher mass star between the opposing pairs, wide binaries, as well as pairs chosen as moving at random angles relative to one another. The uncertainties are derived using sqrt(n) approximation.

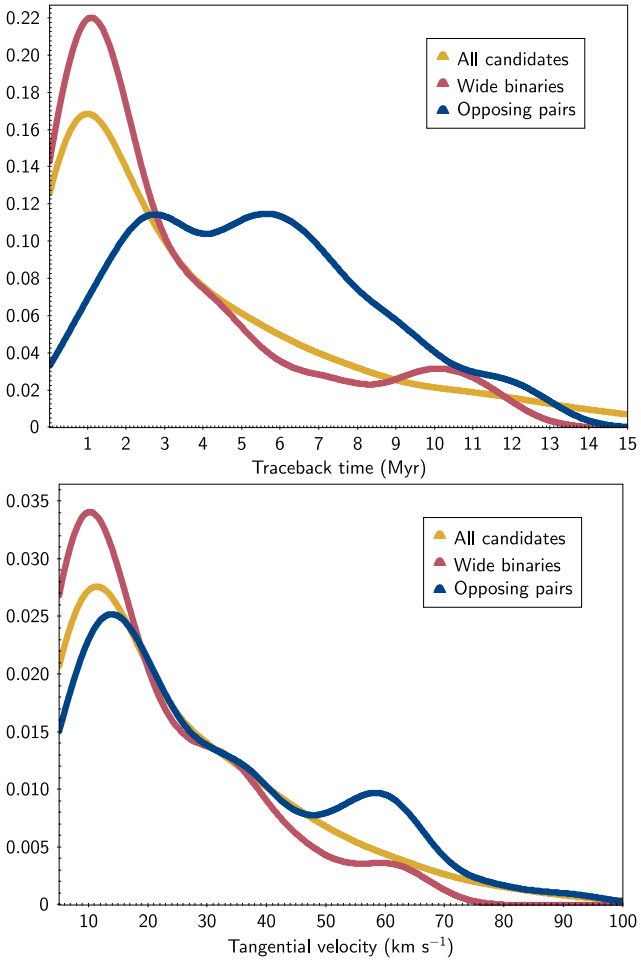

Figure 11. Kernel density estimate showing a distribution of tangential velocity and traceback time for the full sample of candidate ejected stars, as well as the sources that are a part of opposing pairs and wide binaries.

ejected, removing energy from the system in the process, making it more compact but still unstable. The second least massive star would then also be ejected shortly afterwards, and with a new configuration the resulting ejection would often (e.g. Leigh et al. 2016; Ryu, Leigh & Perna 2017a, b) produce slightly higher velocities, provided the virial ratio is close to zero (Leigh et al. 2016). However, a problem with such a scenario is that an ejection would not be simultaneous, but rather two separate ejections in a relatively rapid succession of each other, each launched in a random direction. Such systems may be present among the systems that meet the criteria for the opposing pairs but have different angles of motion; future follow-up may help to more definitively identify such systems, separating them from chance coincidences, enabling a way to traceback their formation history. However, while this scenario can contribute to a 'floor' in the number of pairs across all orientations (Fig. 9), this cannot explain the preference for the pairs to move 180° away from one another.

Another method would be possible through a disintegration of a triple system: through ejecting one of the stars, the newly formed binary system may have received enough recoil to also have been ejected – both this system and the ejected stars would be moving 180° relative to one another. And, the binary system would most likely be more massive than the ejected single, even though individual stars in that binary may be of lower mass, which could be responsible for the observed velocity signature.

In large part, a companion with separations <0.7 arcsec is less likely to be present as a part of the *Gaia* catalogue. Although *Gaia* DR3 has produced various catalogues of stellar multiplicity (Gaia Collaboration 2022), including astrometric, spectroscopic, and eclipsing binaries, they are highly incomplete (especially given the aforementioned issues with RVs in the sample of these young stars). As such, a future search for closer companions among these opposing pairs may better enable understanding their formation.

5 CONCLUSIONS

In this work, we present a first homogeneous search for candidate young runaway/walkaway stars across the entire sky within the solar neighbourhood, resulting in a catalogue of 3354 sources. These candidates were selected from a photometric catalogue of PMS stars that can be separated from the field stars with a relatively high degree of confidence. These sources can be traced back to the nearby star-forming regions within the lifetime of a star and their parent association, most commonly within a few Myr. They are also

distinct from the dominant velocity currents of dissolving young moving groups that are found in their vicinity.

As part of these analysis, we also identify candidate interacting pairs that appear to have been ejected concurrently in the same event, of which 42 appear to be travelling in opposite directions, and 19 appear to have formed a wide binary in the process of their ejection.

The traceback is performed solely in the plane of the sky, as currently only sparse data are available pertaining to the radial velocity information for these stars. Despite the wealth of recently released *Gaia* DR3 wealth of RVs across the solar neighbourhood, these data offer poor constraints on the radial velocity of low-mass stars younger than 100 Myr.

As such, these candidates require spectroscopic follow-up observations to enable the unequivocal confirmation of both their youth that would complement the initial photometric selection, as well as their status as walkaway/runaways through a full 3D traceback. None the less, once confirmed, these candidates can offer insight on the initial dynamical conditions within their parent associations that have led to these stars being ejected.

ACKNOWLEDGEMENTS

This work has used 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. NWCL gratefully acknowledges the generous support of a Fondecyt Iniciación grant 11180005, as well as support from Millenium Nucleus NCN19-058 (TITANs) and funding via the BASAL Centro de Excelencia en Astrofisica y Tecnologias Afines (CATA) grant PFB-06/2007. NWCL also thanks support from ANID BASAL project ACE210002 and ANID BASAL projects ACE210002 and FB210003.

DATA AVAILABILITY

This work uses the data previously made available in McBride et al. (2021). The full tables are included in the Supplemental Materials and will be made available on Vizier.

REFERENCES

Bhat A., Irrgang A., Heber U., 2022, A&A, 663, A39
Blaauw A., 1961, Bull. Astron. Inst. Neth., 15, 265
Campello R. J. G. B., Moulavi D., Sander J., 2013, in Pei J., Tseng V. S., Cao L., Motoda H., Xu G., eds, Advances in Knowledge Discovery and Data Mining. Springer, Berlin, Heidelberg, p. 160
Chen X. et al., 2013, ApJ, 768, 110
El-Badry K., Rix H.-W., Heintz T. M., 2021, MNRAS, 506, 2269
Farias J. P., Tan J. C., Eyer L., 2020, ApJ, 900, 14
Gaia Collaboration, 2018, A&A, 616, A1
Gaia Collaboration, 2021, A&A, 649, A1
Gaia Collaboration, 2022, A&A, preprint (arXiv:2206.05595)
Hoogerwerf R., de Bruijne J. H. J., de Zeeuw P. T., 2001, A&A, 365, 49
Katz D. et al., 2022, preprint (arXiv:2206.05902)
Kerr R. M. P., Rizzuto A. C., Kraus A. L., Offner S. S. R., 2021, ApJ, 917, 23
Kounkel M., Covey K., 2019, AJ, 158, 122
Kounkel M. et al., 2017, ApJ, 834, 142
Kounkel M. et al., 2019, AJ, 157, 196
Kounkel M., Covey K., Stassun K. G., 2020, AJ, 160, 279
Kounkel M., Stassun K. G., Covey K., Hartmann L., 2022, MNRAS, 517, 161
Leigh N. W. C., Geller A. M., 2013, MNRAS, 432, 2474
Leigh N. W. C., Stone N. C., Geller A. M., Shara M. M., Muddu H., Solano-Oropeza D., Thomas Y., 2016, MNRAS, 463, 3311
Luhman K. L., 2018, AJ, 156, 271
Maíz Apellániz J., Pantaleoni González M., Barbá R. H., 2021, Res. Notes Am. Astron. Soc., 5, 232
Manwadkar V., Trani A. A., Leigh N. W. C., 2020, MNRAS, 497, 3694
Manwadkar V., Kol B., Trani A. A., Leigh N. W. C., 2021, MNRAS, 506, 692
McBride A., Kounkel M., 2019, ApJ, 884, 6
McBride A., Lingg R., Kounkel M., Covey K., Hutchinson B., 2021, AJ, 162, 282
McInnes L., Healy J., Astels S., 2017, J. Open Source Softw., 2
Platais I. et al., 2020, AJ, 159, 272
Prisinzano L. et al., 2022, A&A, 664, A175
Reipurth B., Mikkola S., Connelley M., Valtonen M., 2010, ApJ, 725, L56
Ryu T., Leigh N. W. C., Perna R., 2017a, MNRAS, 467, 4447
Ryu T., Leigh N. W. C., Perna R., 2017b, MNRAS, 470, 2
Schoettler C., de Bruijne J., Vaher E., Parker R. J., 2020, MNRAS, 495, 3104
Schönrich R., Binney J., Dehnen W., 2010, MNRAS, 403, 1829
Stassun K. G. et al., 2019, AJ, 158, 138
Stone N. C., Leigh N. W. C., 2019, Nature, 576, 406
Tajiri T. et al., 2020, ApJS, 251, id.18
Tobin J. J. et al., 2016, ApJ, 818, 73
Tobin J. J. et al., 2022, ApJ, 925, 39
Zari E., Hashemi H., Brown A. G. A., Jardine K., de Zeeuw P. T., 2018, A&A, 620, A172
Zari E., Brown A. G. A., de Zeeuw P. T., 2019, A&A, 628, A123
Zucker C., Peek J. E. G., Loebman S. R., 2022, ApJ, 936, 160

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

table1.fits
table2.fits
table3.fits

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

This paper has been typeset from a TEX/LATEX file prepared by the author.