Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **511,** 4702–4709 (2022) https://doi.org/10.1093/mnras/stac391
Advance Access publication 2022 February 22

Astrometric star-cluster membership probability: application to the case of M 37 with Gaia EDR3

M. Griggio[1,2]★ and L. R. Bedin[2]★

[1]*Dipartimento di Fisica, Università di Ferrara, Via Giuseppe Saragat 1, Ferrara I-44122, Italy*
[2]*Istituto Nazionale di Astrofisica, Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, Padova I-35122, Italy*

Accepted 2022 February 8. Received 2022 February 7; in original form 2021 December 28

ABSTRACT

In this work, starting from the well-accepted relations in literature, we introduce a new formalism to compute the astrometric membership probabilities for sources in star clusters, and we provide an application to the case of the open cluster M 37. The novelty of our approach is a refined – and magnitude-dependent – modelling of the parallax distribution of the field stars. We employ the here-derived list of members to estimate the cluster's mean systemic astrometric parameters, which are based on the most recent *Gaia*'s catalogue (EDR3).

Key words: astrometry – *(Galaxy:)* open clusters and associations: individual: NGC 2099 (M 37) – catalogues.

1 INTRODUCTION

Star clusters represent one of our most important sources of knowledge of stellar formation and evolution: The measurements of their distance, age, and chemical composition provide strong constraints on astrophysical models of stellar evolution. Star clusters consist of gravitationally bound stars which share the same distance and centre-of-mass motion, and they appear as a stellar overdensity in a region of the sky. In the studies of these objects, one of the most crucial steps is the determination of the membership probability of the observed stars, to distinguish actual members of the cluster from field stars that lie in the same region but are not bound to the cluster.

Traditionally, the problem of estimating membership probabilities using the astrometric parameters of the sources has been treated with techniques that were developed in the pioneering work by Vasilevskis, Klemola & Preston (1958) and Sanders (1971). In their works, the distribution of sources in the vector-point diagram is modelled as a mixture of two Gaussian distributions, one for the cluster members and another one for the field sources. This method was further refined by the contribution of several authors (see Balaguer-Núnez, Tian & Zhao 1998; Tian, Zhao & Stetson 1998, and references therein).

An additional improvement of this technique introduced by Kozhurina-Platais et al. (1995) foresees the partition of the data in brightness (a *sliding window* in magnitude) and spatial bins when deriving the parameters of the distributions. One of the advantages of using a 'local sample' approach is that membership probabilities are not biased by possible differences in the shape of the field and cluster luminosity functions, or in proper motion accuracy for bright and faint stars.

In this work, we discuss an improvement of the astrometric method exploiting the *Gaia* astrometry to increase the separation between cluster and field stars.

★ E-mail: luigi.bedin@inaf.it (LB); massimo.griggio@inaf.it (MG)

Including parallaxes provides additional information to estimate membership probabilities. While multiple publications since 1998 have taken into account *Hipparcos* (e.g. Robichon et al. 1999; Baumgardt, Dettbarn & Wielen 2000), and later *Gaia* (e.g. Cantat-Gaudin et al. 2018; Castro-Ginard et al. 2018; Gagné et al. 2018; Monteiro et al. 2020) parallaxes, none of these works introduced a proper formalism, with the only exception of Monteiro et al. (2020), which however made an oversimplification that will be discussed later.

This paper is organized as follows: In Section 2, we review the classical formalism used to compute the membership probability, in Section 3, we introduce the new term to account for the parallax distribution, while in Section 4, we compare the membership calculated with this new term and without it, taking the open cluster M 37 as a test case. In Sections 5 and 6, we use the membership probability to select a list of cluster's members and we use them to derive a new estimate of the cluster's mean proper motion and parallax. We also publicly release a catalogue of all the sources with the membership probabilities. Finally, in Section 7, we provide a summary of this work.

2 MEMBERSHIP PROBABILITY: THE CLASSICAL APPROACH

In this section, we will review the formalism 'traditionally' employed to determine the membership probability of the i-th star using four out of its five astrometric parameters, namely its position (x_i, y_i) and its proper motion (μ_{x_i}, μ_{y_i}). We will follow the formulation from Tian et al. (1998) and Balaguer-Núnez et al. (1998).

In these works the cluster membership probability of the i-th star is calculated as

$$P_c(i) = \frac{\Phi_c(i)}{\Phi(i)}, \tag{1}$$

where Φ_c is the cluster distribution function and Φ is the total distribution given by

$$\Phi = \Phi_c + \Phi_f, \tag{2}$$

with $\Phi_{\rm f}$ being the distribution function of field stars. The distribution function of cluster (and field) stars is given by the contribution of two terms, i.e.

$$\Phi_{\rm c/f} = n_{\rm c/f} \cdot \left(\Phi^{v}_{\rm c/f} + \Phi^{r}_{\rm c/f}\right), \tag{3}$$

in which Φ^{v} is the distribution function in the velocity space, Φ^{r} is the distribution function in the position space, and n is the normalized number of stars ($n_{\rm c} + n_{\rm f} = 1$).

For the cluster velocity distribution they adopt an asymmetric 2D Gaussian in the form

$$\Phi^{v}_{\rm c}(i) = \frac{1}{2\pi(\sigma^2_{\mu_{\rm xc}} + \epsilon^2_{\mu_{x_i}})^{1/2}(\sigma^2_{\mu_{\rm yc}} + \epsilon^2_{\mu_{y_i}})^{1/2}} \times \exp\left\{-\frac{1}{2}\left[\frac{\left(\mu_{x_i} - \mu_{x_{\rm c}}\right)^2}{\sigma^2_{\mu_{\rm xc}} + \epsilon^2_{\mu_{x_i}}} + \frac{\left(\mu_{y_i} - \mu_{y_{\rm c}}\right)^2}{\sigma^2_{\mu_{\rm yc}} + \epsilon^2_{\mu_{y_i}}}\right]\right\}, \tag{4}$$

where $\left(\mu_{x_i}, \mu_{y_i}\right)$ are the proper motions of the i-th star, $\left(\mu_{x_{\rm c}}, \mu_{y_{\rm c}}\right)$ is the cluster proper motion centre, $(\sigma_{\mu_{\rm xc}}, \sigma_{\mu_{\rm yc}})$ is the intrinsic proper motion dispersion of member stars, and $(\epsilon_{\mu_{x_i}}, \epsilon_{\mu_{y_i}})$ are the observed errors of the proper motions of the i-th star. Similarly, for the field stars velocity distribution we have

$$\Phi^{v}_{\rm f}(i) = \frac{1}{2\pi\left(1-\gamma^2\right)^{1/2}(\sigma^2_{\mu_{\rm xf}} + \epsilon^2_{\mu_{x_i}})^{1/2}(\sigma^2_{\mu_{\rm yf}} + \epsilon^2_{\mu_{y_i}})^{1/2}} \times \exp\left\{-\frac{1}{2\left(1-\gamma^2\right)}\left[\frac{\left(\mu_{x_i} - \mu_{x_{\rm f}}\right)^2}{\sigma^2_{\mu_{\rm xf}} + \epsilon^2_{\mu_{x_i}}} - \frac{2\gamma\left(\mu_{x_i} - \mu_{x_{\rm f}}\right)\left(\mu_{y_i} - \mu_{y_{\rm f}}\right)}{(\sigma^2_{\mu_{\rm xf}} + \epsilon^2_{\mu_{x_i}})^{1/2}(\sigma^2_{\mu_{\rm yf}} + \epsilon^2_{\mu_{y_i}})^{1/2}} + \frac{\left(\mu_{y_i} - \mu_{y_{\rm f}}\right)^2}{\sigma^2_{\mu_{\rm yf}} + \epsilon^2_{\mu_{y_i}}}\right]\right\}, \tag{5}$$

where $\left(\mu_{x_i}, \mu_{y_i}\right)$ are the proper motions of the i-th star, γ is the correlation coefficient between μ_{x_i} and μ_{y_i}, $\left(\mu_{x_{\rm f}}, \mu_{y_{\rm f}}\right)$ the field proper motion centre, $(\epsilon_{\mu_{x_i}}, \epsilon_{\mu_{y_i}})$ the observed errors of the proper motions of the i-th star, and $(\sigma_{\mu_{x_{\rm f}}}, \sigma_{\mu_{y_{\rm f}}})$ the field intrinsic proper motion dispersion.

For the spatial distribution of cluster members a simple (and sufficient for the purpose) approximation is to use a Gaussian profile:

$$\Phi^{r}_{\rm c}(i) = \frac{1}{2\pi r^2_{\rm c}} \exp\left\{-\frac{1}{2}\left[\left(\frac{x_i - x_{\rm c}}{r_{\rm c}}\right)^2 + \left(\frac{y_i - y_{\rm c}}{r_{\rm c}}\right)^2\right]\right\}, \tag{6}$$

in which (x_i, y_i) is the position of the i-th star, $(x_{\rm c}, y_{\rm c})$ the centre of the cluster, and $r_{\rm c}$ the characteristic radius. The field star spatial distribution is assumed to be flat:

$$\Phi^{r}_{\rm f}(i) = \frac{1}{\pi r^2_{\rm max}}, \tag{7}$$

where $r_{\rm max}$ is the radius of the portion of the sky under exam (assuming it has a circular shape).

This method to compute the membership probabilities was applied in a number of papers in the recent literature (see e.g. Yadav et al. 2008; Bellini et al. 2009; Nardiello et al. 2018; Scalco et al. 2021).

3 INCLUDING THE PARALLAX

The Gaia EDR3 (Gaia Collaboration 2016, 2021) catalogue is an unprecedented astronomical data set in terms of its size and astrometric precision and accuracy. In particular, it provides the full 5-parameter astrometric solution (positions, proper motions, and parallaxes) and magnitudes in its three photometric bands (G, $G_{\rm BP}$, and $G_{\rm RP}$) for more than 1.4 billion sources, with a limiting magnitude of about $G \approx 21$ and a bright limit of about $G \approx 3$. Thanks to the Gaia EDR3 exquisite astrometry we can extend the formalism presented in the previous section including a new term to take into account the parallax distribution. Particularly, the parallax uncertainties in the EDR3 are 0.02–0.03 mas for $G < 15$, 0.07 mas at $G = 17$, 0.5 mas at $G = 20$, and 1.3 mas at $G = 21$ (Lindegren et al. 2021a). This unmatched level of precision allows us to include the parallax in the computation of the membership probability, thus achieving a more robust estimate for this fundamental quantity.

To account for the parallax distribution we rewrite equation (3) as

$$\Phi_{\rm c/f} = n_{\rm c/f} \cdot \left(\Phi^{v}_{\rm c/f} + \Phi^{r}_{\rm c/f} + \Phi^{\varpi}_{\rm c/f}\right), \tag{8}$$

where $\Phi^{\varpi}_{\rm c/f}$ is the distribution function of the parallaxes for the cluster members and for the field stars.

We can assume that the parallaxes of cluster members are normally distributed, such that

$$\Phi^{\varpi}_{\rm c}(i) = \frac{1}{\left(2\pi(\sigma^2_{\varpi_{\rm c}} + \epsilon^2_{\varpi_i})\right)^{1/2}} \exp\left\{-\frac{1}{2}\left(\frac{\varpi_i - \varpi_{\rm c}}{\sigma^2_{\varpi_{\rm c}} + \epsilon^2_{\varpi_i}}\right)^2\right\}, \tag{9}$$

where $(\varpi_i, \varpi_{\rm c})$ are the parallax of the i-th star and of the cluster, respectively, ϵ_{ϖ_i} are the observed errors of the parallax of the i-th star, and $\sigma_{\varpi_{\rm c}}$ the cluster intrinsic parallax dispersion (in the case where the size of the cluster is not negligible compared to its distance).

However, modelling the distribution function of the parallaxes of field stars, $\Phi^{\varpi}_{\rm f}$, is more complicated: We are not observing an ensemble of stars all at the same distance, or at an average distance with a normal distribution around the mean. In the case of the parallaxes of the field we are rather observing the closest stars to the Sun and stars potentially well into the Galactic Halo. The exact distribution function of stars in the Galactic field in different directions and at the various magnitudes is hard to model, and to derive an accurate distribution is well beyond the purpose of this paper. For our purposes, it will be sufficient to adopt a simple approximation, analogous to what described in equation (5) for the proper motion distribution of field objects $\Phi^{v}_{\rm f}$ (which is a widely accepted approximation in literature). To reproduce the field stars' parallax distribution we adopted a sum of two Gaussian functions, which are assumed to model as well the measurement errors in the parallaxes. This choice let us reproduce very well the parallax at each magnitude bin without complicating too much our formalism. Therefore, we have

$$\Phi^{\varpi}_{\rm f} = \frac{A_1}{(2\pi\sigma^2_{1_{\rm f}})^{1/2}} \exp\left\{-\frac{1}{2}\left(\frac{\varpi_i - \varpi_{1_{\rm f}}}{\sigma^2_{1_{\rm f}}}\right)^2\right\} + \frac{A_2}{(2\pi\sigma^2_{2_{\rm f}})^{1/2}} \exp\left\{-\frac{1}{2}\left(\frac{\varpi_i - \varpi_{2_{\rm f}}}{\sigma^2_{2_{\rm f}}}\right)^2\right\}, \tag{10}$$

where A_1, A_2, $\varpi_{1_{\rm f}}$, $\varpi_{2_{\rm f}}$, $\sigma_{1_{\rm f}}$, and $\sigma_{2_{\rm f}}$ are the parameters of the two Gaussian and ϖ_i the parallax of the i-th star. We verified that this simple model for the distribution of the parallaxes for field objects is a general valid approximation. To this aim, we downloaded portions of the Gaia EDR3 catalogue in various directions of the sky, to probe different parts of the Galactic field. The obtained Gaia EDR3 distributions of the parallaxes for field objects at various magnitudes were always represented – within the statistical sampling errors – by our simple model.

We note here that Monteiro et al. (2020) followed a method that is qualitatively similar to ours, but they do not use the sliding window approach and they adopted a single Gaussian model; while this assumption works well in the case of parallaxes dominated by

errors (faint stars), it does not represents well the intrinsic parallax distribution of field stars when uncertainties are small. We show a detailed comparison of the two models in the next section.

4 EXAMPLE: THE CASE OF M 37

We considered the open cluster M 37 (NGC 2099) as a test case for this new formalism for the computation of the membership probabilities. We downloaded a portion of the Gaia EDR3 catalogue centred on M 37 ($\alpha_c = 88.074$ deg, $\delta_c = +32.545$ deg; Cantat-Gaudin & Anders 2020, hereafter CG20) with a radius of 1.5 deg, and we computed the membership probability for each source both including and neglecting the contribution from the *Gaia* parallaxes. We adopted a sliding window in magnitude of 1.5 mag, which we found as a good compromise between having a good statistics at all magnitudes and considering sources with magnitude similar to the target star. As initial guess, we employed for the cluster's systemic proper motion[1] $(\mu_{x_c}, \mu_{y_c}) = (\mu_{\alpha_c}, \mu_{\delta_c}) = (1.924, -5.648,)$ mas yr^{-1} and for the systemic parallax $\varpi_c = 0.666$ mas, which are the values given by CG20.

We started by estimating the intrinsic dispersion of the proper motions of the cluster, i.e. $\sigma_{\mu_{x_c}}$ and $\sigma_{\mu_{y_c}}$ of equation (4). We selected the members of M 37 for which CG20 give their clustering score equal to one (the highest score), choosing only the sources with $G < 17$ (where the *Gaia* errors are of the order, 10^{-2} mas yr^{-1}). We σ-clipped the values at 3σ around the median, and then we calculated the 68.27$^{\rm th}$ percentile of the residuals from the median of μ_α and μ_δ, which we assumed as the observed dispersion. Subtracting in quadrature from the observed dispersion the median observational relative errors provided by Gaia EDR3 gives a reasonable estimate of the cluster intrinsic dispersion. We obtain $\sigma_{\mu_{x_c}} = \sigma_{\mu_\alpha} = 137$ μas and $\sigma_{\mu_{y_c}} = \sigma_{\mu_\delta} = 138$ μas. Note that at a corresponding distance of 1.5 kpc (for $\varpi = 0.666$ mas), these translate into transverse velocities of less than 1 km s^{-1}, which is a reasonable value for such an open cluster (see e.g. CG20, and references therein). The values for proper motions and estimated errors of individual sources in the equation (4) are taken straight from the Gaia EDR3 catalogue, i.e.: $\mu_{x_i} = \mu_{\alpha_i}$, $\mu_{y_i} = \mu_{\delta_i}$, $\epsilon_{x_i} = \epsilon_{\mu_{\alpha_i}}$, and $\epsilon_{y_i} = \epsilon_{\mu_{\delta_i}}$.

The parameters $\mu_{x(y)_f}$ and $\sigma_{\mu_{x(y)_f}}$ of equation (5) have been estimated from the sources in the magnitude window of the star under exam. We adopted as $\mu_{x(y)_f}$ the median values of the proper motion of field objects, and the 68.27$^{\rm th}$ percentile of the residuals around these median as the observed dispersion: $\sigma^{\rm obs}_{\mu_{x(y)_f}}$. We then calculated the average observational error of the sources in the magnitude window, $\epsilon_{\mu_{x(y)_f}}$, by clipping the errors at 3σ and computing the median. This value is then used to calculate the intrinsic dispersion for field proper motion as: $\sigma^2_{\mu_{x(y)_f}} = (\sigma^{\rm obs}_{\mu_{x(y)_f}})^2 - \epsilon^2_{\mu_{x(y)_f}}$, which are the ones to be used in equation (5). Again, for individual sources the values for proper motions and proper motion errors in the equation (5) are taken from the Gaia EDR3 catalogue.

To deal with the spatial distribution, we projected the *Gaia* coordinates (α_i, δ_i) on the tangent plane (ξ, η), adopting the centre of the cluster as tangent point (α_c, δ_c), employing standard relations (see e.g. equation 3 in Bedin & Fontanive 2018). Therefore, the coordinates on the tangent plane became $x_i = \xi_i$ and $y_i = \eta_i$ in equation (6). The estimate of r_c of equation (6) have been performed from the stars in the magnitude bin of the target; we calculate r_x (r_y) as the 68.27$^{\rm th}$ percentile of the residuals from x_c (y_c), and we adopt as cluster radius in the magnitude bin $r_c^2 = r_x^2 + r_y^2$. This procedure allows us to account for the different distributions of the stars in each magnitude bin, which is a proxy for different mass-bins (at least along the main sequence). The parameter $r_{\rm max}$ of equation (7) is the radius of the Gaia EDR3 slice that we considered, i.e. $r_{\rm max} = 1.5$ deg.

The intrinsic dispersion of the parallaxes (σ_{ϖ_c} in equation 9) is negligible for M 37 and it can be set equal to zero. Again, for this initial computation we adopt for the cluster average parallax the value from CG20, while the values of the parallax (ϖ_i) and parallax error (ϵ_{ϖ_i}) for individual sources are those given by the Gaia EDR3 catalogue. In the next sections, we will then derive our own estimates for the cluster mean proper motions and parallax, employing Gaia EDR3 instead of the DR2, and finally recompute the updated membership probabilities.

For the distribution of parallaxes in the Galactic field we used equation (10), so we fitted the distribution in each magnitude window with a sum of two Gaussian functions. From the fit we obtain the $\tilde{\sigma}_{1(2)_f}$ parameters, which contain also the contributions of errors in parallaxes at the considered magnitude. To account for the observational errors, we compute the quantity $\sigma'^2_{1(2)_f} = \tilde{\sigma}^2_{1(2)_f} - \bar{\epsilon}^2_{1(2)_f}$, where $\bar{\epsilon}_{1(2)_f}$ is the median error of the stars in the magnitude window that we are considering, calculated after performing a 3σ clip. These are the values employed in equation (10), which are the sum in quadrature of intrinsic distributions and errors, $\sigma^2_{1(2)_f} = \sigma'^2_{1(2)_f} + \epsilon^2_{\varpi_i}$. We then used the Gaia EDR3 values for the parallax ϖ_i and parallax error ϵ_{ϖ_i}. In Fig. 1, we show the distribution of the parallaxes in the M 37 field of view, for different magnitude bins, where for comparison we show the fitted distribution of the parallaxes of cluster+field stars obtained in the case of a two-Gaussian model (solid line) and in the case of one-Gaussian model (dotted line).

We finally computed the membership probability, hereafter P_ϖ, following equation (8), which includes the parallax, and we compared it with the membership probability calculated without including the parallax, i.e. P, computed according to equation (3). The results show that including the parallax term allow us to better separate the cluster members from the field stars, in particular at fainter magnitudes ($16 < G < 18$).

In the top panel of Fig. 2, we show the membership probability calculated with the standard approach, without the parallax term, while on the bottom panel we show the results obtained including the parallax. In the left-hand panels (0 per cent $< P <$ 95 per cent), we can see that there are considerably less sources with magnitudes in the range $10 < G < 19$ within 25 per cent $< P <$ 75 per cent if we account for the parallax distribution (372 sources in the top panel, 282 in the bottom), confirming that the discrimination between member and field stars is better.

In Fig. 3 (top panel), we plotted the maximum membership probability per magnitude bin versus the *Gaia G* magnitude. The blue and red lines represent the probability calculated with and without taking into account the parallax, respectively. This plot shows us that very high membership probabilities extend deeper when considering the parallax contribution in the calculation (about 1 mag at $P_{\rm max} =$ 99 per cent).

We then divided the sources in magnitude bins of 1 mag, and calculated the number of sources in each bin with $P > 97.5$ per cent. In Fig. 3 (bottom panel), we plot the results: In blue we show the points obtained without accounting for the parallax, in red those obtained including the parallax in the membership calculation. It is clear that for $G \gtrsim 14$ we find more member stars if we use the parallax term. In total we found 1266 sources with $P \geq 97.5$ per cent, and 1824 sources with $P_\varpi \geq 97.5$ per cent in the region $10 < G < 20$, where P_ϖ is the membership probability calculated with the formalism introduced in this work.

[1]Where for conciseness in the notation we indicate $\mu_{\alpha\cos\delta_c}$ with μ_{α_c}.

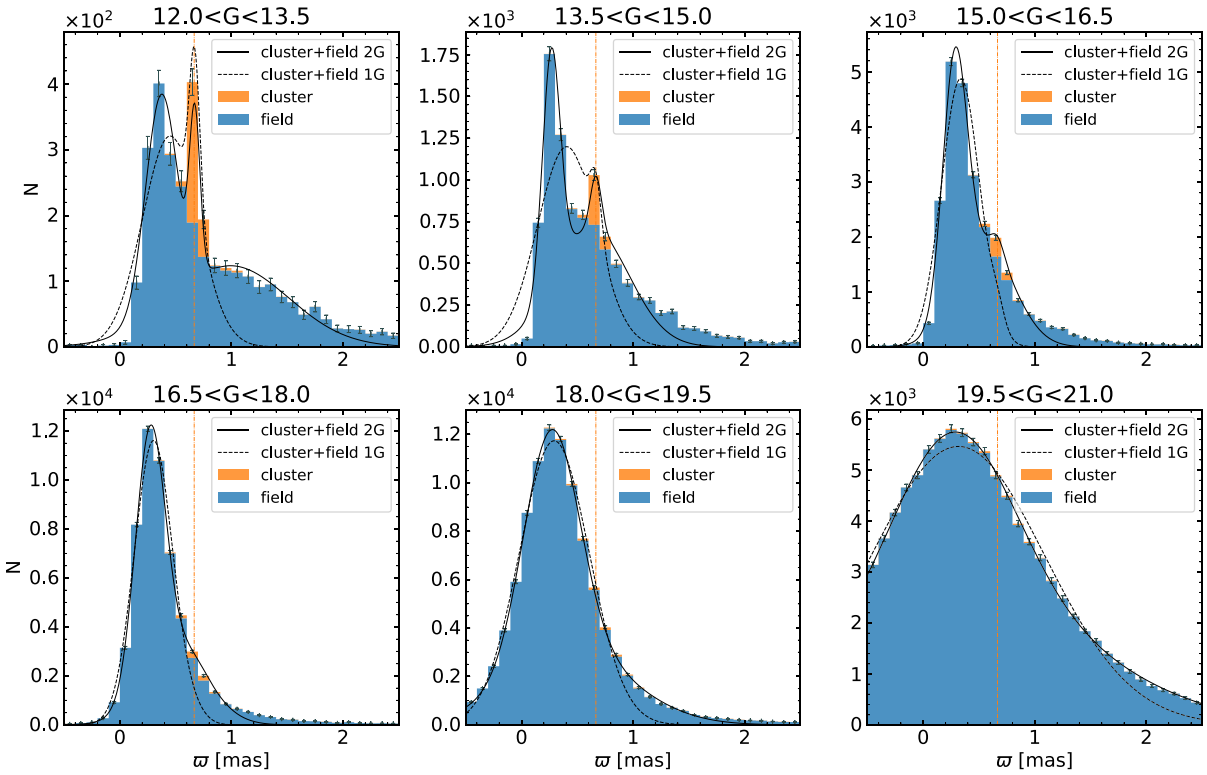

Figure 1. Parallax distribution in different magnitude bins in the M 37 field of view. We plotted in orange the stars with parallax within 0.2 mas from the cluster's mean parallax and proper motions within a circle of 0.5 mas yr^{-1} in radius centred on the cluster's mean proper motions. Mean values were obtained from Cantat-Gaudin & Anders (2020). The solid black line represents the sum of the Gaussian of the cluster and the field modelled with two Gaussian functions, while the dashed black line is the same but using a single Gaussian to model the field's parallax distribution.

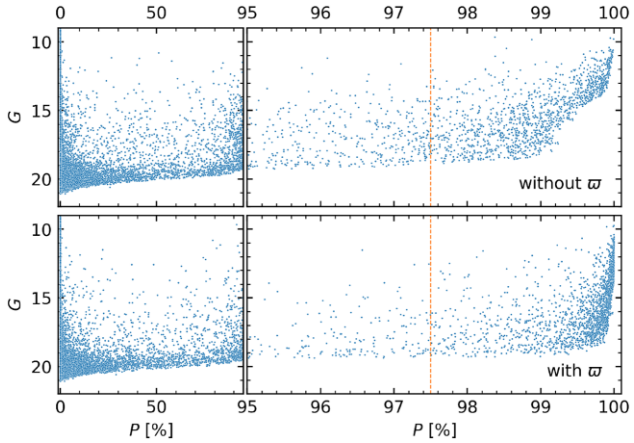

Figure 2. Comparison between the membership probability calculated without the parallax term (top panel) and accounting for the parallax of the sources (bottom panel); $P > 95$ per cent are zoomed on the right-hand panels. Vertical orange line indicates $P = 97.5$ per cent.

Fig. 4 shows a comparison between CG20 and this work. In the left-hand column we show all the sources in the catalogue; in this figure, we limit the sample only to sources within a radius of 0.3 deg (slightly more than the cluster radius given by Dias et al. 2002) from the centre of the cluster, as the number of field objects beyond this limit would overwhelm the plot, making the comparisons less clear.

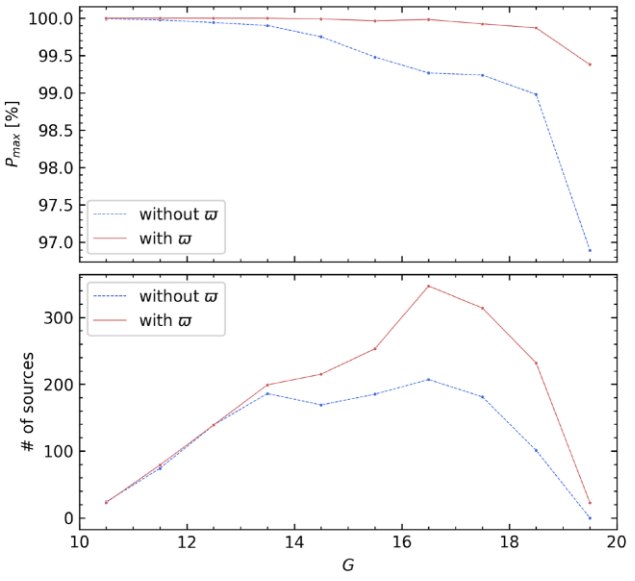

Figure 3. Top panel: Maximum membership probability per magnitude bin. The dashed blue (solid red) line is the membership probability computed without (with) taking into account the parallax distribution. Bottom panel: Number of sources per magnitude bin. The dashed blue (solid red) line is the number of sources with $P > 97.5$ per cent ($P_{\varpi} > 97.5$ per cent).

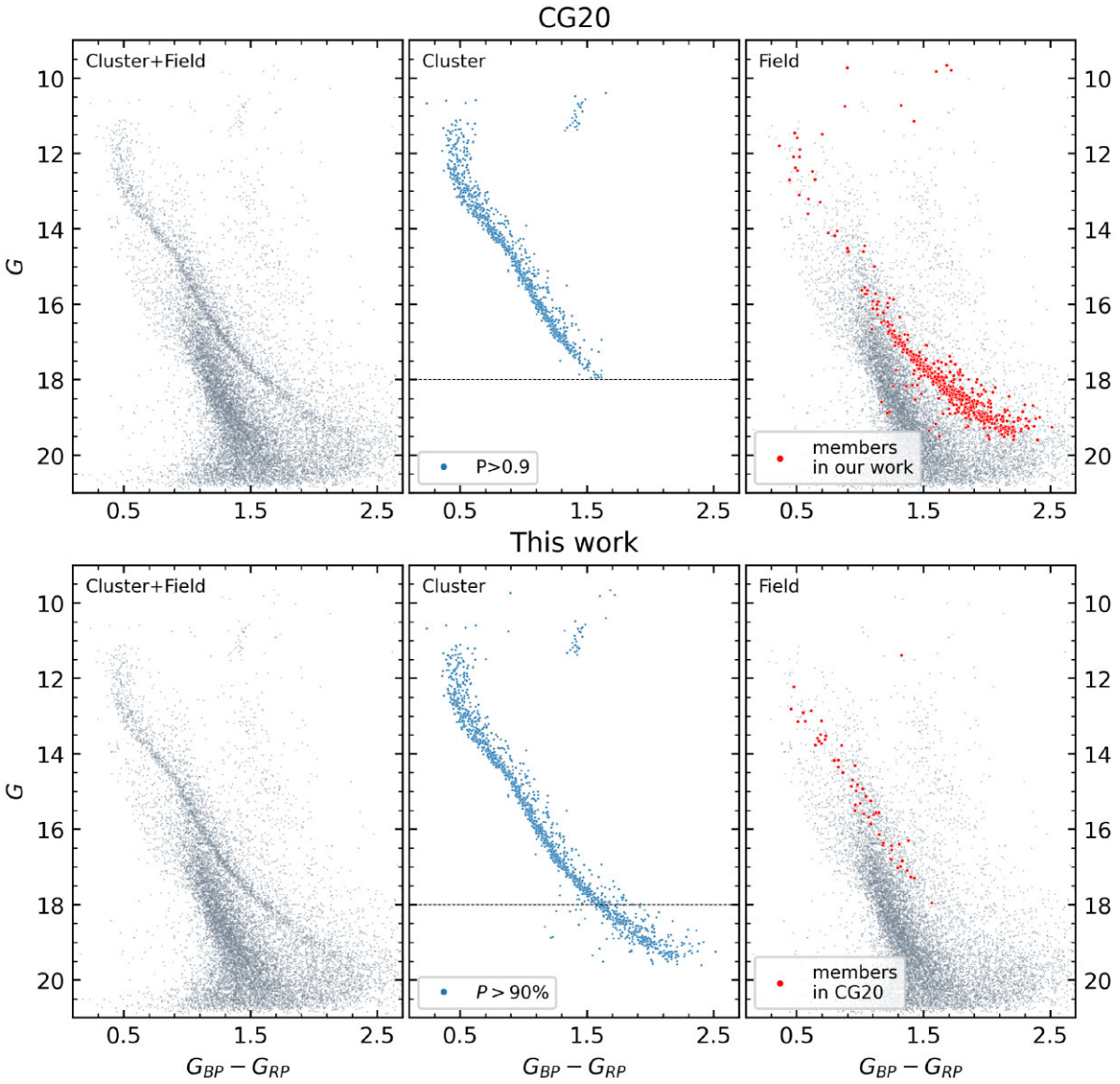

Figure 4. Colour–magnitude diagram (CMD) of the sources in M 37 field of view: Top panels show members identified by CG20, while bottom panels show the members selected in this work. Left-hand panel: All the sources. Central panel: Sources with $P \geq 0.9$ per cent (top panel) and $P_{\varpi} \geq 90$ per cent (bottom panel). Right-hand panel: Sources with $P < 0.9$ per cent (top panel) and $P_{\varpi} < 90$ per cent (bottom panel); in the top (bottom) panel we highlighted in red the sources that passed the membership cut in the bottom (top) row. See text for more details.

In the central column, we show the stars with $P \geq 0.9$ (top panel) and with $P_{\varpi} \geq 90$ per cent (bottom panel). The clustering score given by CG20 is provided only for stars brighter than $G = 18$ (black dashed line) and according to the authors is a proxy for cluster membership probability. However, the main sequence of M 37 clearly extends (and it is well populated) also to fainter magnitudes than that limit. In the common region analysed by both works ($G \leq 18$) we found about 200 extra sources with membership probability greater than 90 per cent, with respect to those with clustering score greater than 0.9 by CG20. Nevertheless, the most interesting plots are shown in the right-hand column, where we plotted the sources that did not pass the membership selection; in the top panel we highlighted in red the sources that did not pass the $P > 0.9$ selection in the middle-top panel, but that passed the $P > 90$ per cent membership probability of the present work. Conversely, the stars in red in the bottom-right panel are those members according to $P > 0.9$ in CG20, but not to the here-derived $P > 90$ per cent. Apart from the obvious improvement of the present work in finding members beyond $G > 18$, we note a significant improvement in identifying members also in the magnitude range $16 < G < 18$.

5 ASTROMETRIC PARAMETERS OF M 37

To derive the mean astrometric parameters of M 37 from the EDR3 catalogue, we first need to select the most probable cluster members. The selection procedure is illustrated in Fig. 5. In the top-left panel, we show the membership probability plotted against the G-band magnitude. We started by rejecting all the sources with membership probability lower than 50 per cent. Among these sources we rejected those falling outside the area delimited by the two red dashed line on the top-right panel. To define these red lines we proceeded as follows:

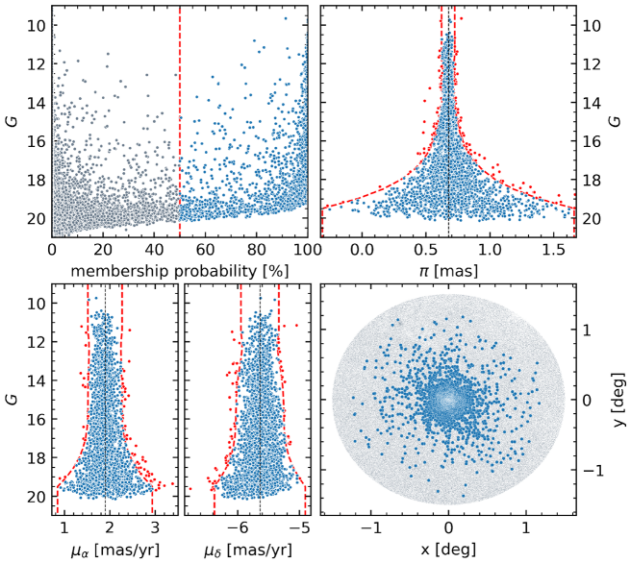

Figure 5. Members' selection. Top-left panel: Membership probability for all the sources. We reject the stars with $P < 50$ per cent. Top-right panel: G magnitude versus parallax. Here, we reject the stars that fall outside the region delimited by the two red lines. The black dashed line represents the median parallax. Bottom-left panel: Proper motion of the sources versus their G magnitude. We kept the sources between the red dashed lines. The vertical black line is the median proper motion. Bottom-right panel: Spatial distribution of the sources. Blue markers are the selected members of M 37.

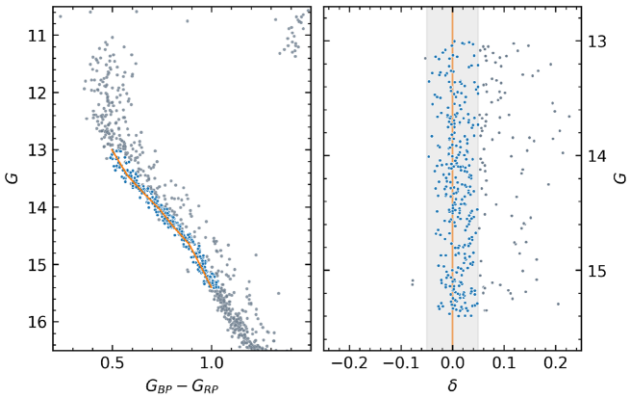

Figure 6. Left-hand panel: CMD showing in blue the sources that we used to estimate the mean parameters of M 37. Right-hand panel: Colour residuals from the fiducial for the stars in the selected sample; we used the sources in the shaded grey area which corresponds to 1σ.

First, we divided the stars into G-magnitude bins of 0.5, for each bin we calculated the 3σ-clipped median of the errors on the parallax given by the EDR3, and we took this median – multiplied by a factor of 2.5 – as the maximum error for members at the given G magnitude. We then define the red lines as a spline through these maxima. In the bottom-left panels, we applied a similar cut, but we did not use the measurements errors on the proper motion from the Gaia catalogue as the errors are much smaller than the intrinsic proper motion dispersion of cluster members (especially at the brighter magnitudes). Therefore, to define the widths of each bin we used instead the 68.27$^{\rm th}$ of the observed residuals from the median (defined after a 3σ-clipping), and again multiplied by a factor of 2.5. On the bottom-right panel, we show the spatial distribution of the stars that passed all these four selections and which we then we consider as most probable members of M 37.

We then further restrict this sample to the very best stars, requiring:

(i) $P_{\varpi} > 99.5$ per cent, i.e. high-confidence members;
(ii) magnitudes in all the three *Gaia* filters (no colour trends);
(iii) $13 \leq G \leq 15.4$, were the astrometric calibration of the EDR3 catalogue provide homogeneous errors (cf. Fabricius et al. 2021);
(iv) $\sigma_{\varpi}/\varpi < 0.1$, $\sigma_{\mu_\alpha}/\mu_\alpha < 0.1$, and $\sigma_{\mu_\delta}/\mu_\delta < 0.1$; and
(v) passing a number of quality cuts on the diagnostic parameters provided within the Gaia EDR3, as done by Soltis, Casertano & Riess (2021).

Specifically, these applied quality-parameters cuts are:

`astrometric_excess_ noise` < 1;
`astrometric_excess_noise_sig` $<= 10$;
`phot_bp_rp_excess_factor` < 1.6;
`phot_proc_mode` $= 0$; and
`astrometric_gof_al` < 4.

After these selections we considered the CMD of member stars (Fig. 6, left-hand panel) and applied a constraint in the G versus $(G_{\rm BP}-G_{\rm RP})$ plane to exclude the region of the CMD populated by high-mass ratio photometric binaries (real or blends) that may have lower precision astrometry. This is achieved as follows: We divided the sample into G-magnitude bins of 0.3 mag and we arbitrarily defined a specific colour for each bin as the 30$^{\rm th}$ percentile of the colour distribution of the stars in the bin. We then interpolated these points at any given G magnitude with a spline. The fiducial line defined in this way follows the bluer envelope of the main sequence, as shown in the left-hand panel of Fig. 6. We then calculated the colour residuals δ from the fiducial and discarded the sources with $|\delta| > 1\sigma$ (Fig. 6, right-hand panel).

With this tight selection of the very best measured and most likely members just defined for M 37, we now proceed with our own derivation of the cluster mean astrometric parameters. We first compute the 3σ-clipped median of ϖ, μ_α, and μ_δ for each G bin of 0.5 mag, with σ defined as the 68.27$^{\rm th}$ percentile of the residuals around the median. The error associated with each bin is defined as $\epsilon_k = \sigma/\sqrt{(N-1)}$, with N the number of sources in the bin. The values for the mean parallax and proper motions are calculated as a weighted mean through all the bins, with $1/\epsilon_k^2$ as weight.

As the astrometric parameters for M 37 are now better determined, thanks of the use of EDR3 (instead of being based on DR2 as in CG20) and the improved memberships, it makes more sense to use the newly determined cluster's mean parameters as starting values for our algorithm and to redetermine the membership probabilities. Therefore, we repeated the analysis just discussed to derive our final estimate of the mean proper motion and parallax of M 37. The values of mean parallax and proper motions for each magnitude bin are plotted in Fig. 7, with the weighted mean through all the bins shown on the top right of each panel. These final values are also reported in Table 1. We point out that neglecting the last selection on the CMD (displayed in Fig. 6) our estimates do not change significantly (less than 0.3σ).

Finally, Lindegren et al. (2021b) found that EDR3 parallaxes of sources identified as quasars are systematically offset from the expected distribution around zero by a few tens of μas. They give an attempt to account for this offset which depends non-trivially on the magnitude, colour, and ecliptic latitude of the source. We used their `Python` code to correct the parallaxes of M 37 members and then recomputed the mean value. However, as they point out in their

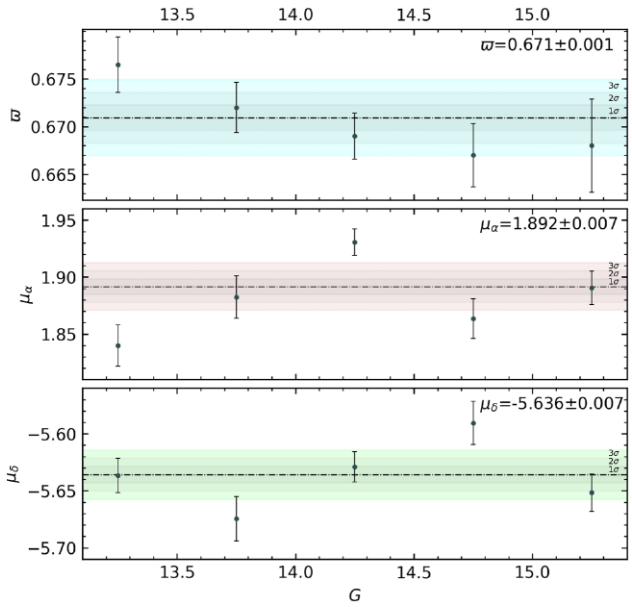

Figure 7. Mean values of the parallax (top panel) and proper motion components (middle and bottom panels) in each magnitude bin; the dash-dotted line is the overall weighted mean value, which is reported on the top right corner of each panel and in Table 1.

Table 1. M 37 mean parallax and proper motion. The value of $\varpi^{\rm L21}$ is the parallax corrected for the bias as in Lindegren et al. (2021b).

Parameter	Value	Unit
ϖ	0.671 ± 0.001	mas
$\varpi^{\rm L21}$	0.629 ± 0.001	mas
μ_α	1.892 ± 0.007	mas yr^{-1}
μ_δ	-5.636 ± 0.007	mas yr^{-1}

work, this correction is still under development and has problems, which seems to be supported by the disagreement with the expected values of $\sim$20 μas (cf. fig. 5 of Lindegren et al. 2021b). We report also in Table 1 this bias-corrected value, as $\varpi^{\rm L21}$.

As a final note, while the absolute value of the parallax has not a direct effect on the membership probabilities, which is mostly a differential computation, it would still be good to have an indication of the systematic error in the just derived parallax. Therefore, we can conservatively associate a maximal error of $|\varpi - \varpi^{\rm L21}| = 0.043$ mas, to the absolute parallax of M 37 derived from Gaia EDR3: $0.671 \pm 0.001 \pm 0.043$ mas, i.e. corresponding to a distance of 1.5 ± 0.1 kpc.

6 CATALOGUE OF M 37

As part of this work, we electronically release as Supporting Information on the Journal a catalogue containing the Gaia EDR3 `source ID` and P_ϖ (the membership probability calculated with the formalism presented in this work).

7 SUMMARY

In this paper, we presented a simple term, which involve parallaxes, to extend the classical method for computing cluster-membership probabilities based *only* on proper motions and spatial distributions. The proposed new formalism, therefore, takes into account the full-astrometric information to compute memberships. Although currently this method suites only data provided by the Gaia EDR3 catalogue, in principle this formalism could be adopted also to future other 5-parameters high-precision astrometric catalogues, or possibly to extensions of the *Gaia* astrometry to fainter magnitudes exploiting superior instruments capabilities (e.g. using *Hubble Space Telescope* observations as in Bedin & Fontanive 2018, 2020). We also note that employing relative instead of absolute parallaxes would not affect the membership probabilities as parallaxes enter only as a relative quantity in the calculations, nor would make any difference to add corrections for the systematic errors, such as those described in Lindegren et al. (2021b) for Gaia EDR3.

We successfully applied this formalism to the case of the close-by open cluster M 37, and release the derived membership probabilities. Results show that the new term allows us to better separate cluster members from field stars at all the magnitudes. We finally used the here-derived list of members to give a new estimate of the astrometric parameters of the cluster.

Future improvements of the method might combine the photometric information. Indeed, especially on wide open clusters with sparse densities, or in their outskirts in general, field objects might survive even tights membership probability selections, incidentally having same distance and motion of the clusters. Other future works might also include a term that would take into account the velocity along the line of sight of the sources (commonly referred to as *radial velocities* in spectroscopy), when available. As *Gaia* radial velocities have not a great precision (200–300 m s^{-1} at best, up to 2.5 km s^{-1}), nor they extend to sufficiently faint magnitudes (G in the range 4–13), we ignored this term in this paper, which is focused on the astrometric parameters only.

ACKNOWLEDGEMENTS

Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. The authors acknowledge support by MIUR under PRIN program #2017Z2HSMF and by PRIN-INAF 2019 under program #10-Bedin.

DATA AVAILABILITY

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).

REFERENCES

Balaguer-Núnez L., Tian K. P., Zhao J. L., 1998, A&AS, 133, 387
Baumgardt H., Dettbarn C., Wielen R., 2000, A&AS, 146, 251
Bedin L. R., Fontanive C., 2018, MNRAS, 481, 5339
Bedin L. R., Fontanive C., 2020, MNRAS, 494, 2068
Bellini A. et al., 2009, A&A, 493, 959
Cantat-Gaudin T. et al., 2018, A&A, 615, A49
Cantat-Gaudin T., Anders F., 2020, A&A, 633, A99
Castro-Ginard A., Jordi C., Luri X., Julbe F., Morvan M., Balaguer-Núñez L., Cantat-Gaudin T., 2018, A&A, 618, A59
Dias W. S., Alessi B. S., Moitinho A., Lépine J. R. D., 2002, A&A, 389, 871
Fabricius C. et al., 2021, A&A, 649, A5
Gagné J. et al., 2018, ApJ, 856, 23
Gaia Collaboration, 2016, A&A, 595, A1
Gaia Collaboration, 2021, A&A, 649, A1

Kozhurina-Platais V., Girard T. M., Platais I., van Altena W. F., Ianna P. A., Cannon R. D., 1995, AJ, 109, 672
Lindegren L. et al., 2021a, A&A, 649, A2
Lindegren L. et al., 2021b, A&A, 649, A4
Monteiro H., Dias W. S., Moitinho A., Cantat-Gaudin T., Lépine J. R. D., Carraro G., Paunzen E., 2020, MNRAS, 499, 1874
Nardiello D. et al., 2018, MNRAS, 481, 3382
Robichon N., Arenou F., Mermilliod J. C., Turon C., 1999, A&A, 345, 471
Sanders W. L., 1971, A&A, 14, 226
Scalco M. et al., 2021, MNRAS, 505, 3549
Soltis J., Casertano S., Riess A. G., 2021, ApJ, 908, L5
Tian K. P., Zhao J. L., Stetson P. B., 1998, A&AS, 131, 89
Vasilevskis S., Klemola A., Preston G., 1958, AJ, 63, 387
Yadav R. K. S. et al., 2008, A&A, 484, 609

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

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.