

# The multiplicity fraction in 202 open clusters from *Gaia*

J. Donada<sup>1,2,3</sup>, F. Anders<sup>1,2,3</sup>, C. Jordi<sup>1,2,3</sup>, E. Masana<sup>1,2,3</sup>, M. Gieles<sup>1,2,3,4</sup>,  
G. Perren<sup>5</sup>, L. Balaguer-Núñez<sup>1,2,3</sup>, A. Castro-Ginard<sup>6</sup>, T. Cantat-Gaudin<sup>7</sup>, L. Casamiquela<sup>8</sup>

- <sup>1</sup> Departament de Física Quàntica i Astrofísica (FQA), Universitat de Barcelona (UB), Martí i Franquès, 1, 08028 Barcelona, Spain  
<sup>2</sup> Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona (UB), Martí i Franquès, 1, 08028 Barcelona, Spain  
<sup>3</sup> Institut d'Estudis Espacials de Catalunya (IEEC), Gran Capità, 2-4, 08034 Barcelona, Spain  
<sup>4</sup> Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys 23, E-08010 Barcelona, Spain  
<sup>5</sup> Instituto de Astrofísica de La Plata (IALP-CONICET), 1900 La Plata, Argentina  
<sup>6</sup> Leiden Observatory, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands  
<sup>7</sup> Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany  
<sup>8</sup> GEPI, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Paris Cité, 5 place Jules Janssen, 92190 Meudon, France

Received January 27, 2023; accepted ...

## ABSTRACT

In this study, we estimate the fraction of binaries with high mass ratios for 202 open clusters in the extended solar neighbourhood (closer than 1.5 kpc from the Sun). This is one of the largest homogeneous catalogues of multiplicity fractions in open clusters to date, including the unresolved and total (close-binary) multiplicity fractions of main-sequence systems with mass ratio larger than  $0.6^{+0.05}_{-0.15}$ . The unresolved multiplicity fractions are estimated applying a flexible mixture model to the observed *Gaia* colour-magnitude diagrams of the open clusters. Then we use custom *Gaia* simulations to account for the resolved systems and derive the total multiplicity fractions. The studied open clusters have ages between 6.6 Myr and 3.0 Gyr and total high-mass-ratio multiplicity fractions between 6% and 80%, with a median of 18%. The multiplicity fractions increase with the mass of the primary star, as expected. The average multiplicity fraction per cluster displays an overall decreasing trend with the open cluster age up to ages about 100 Myr, above which the trend increases. Our simulations show that most of this trend is caused by complex selection effects (introduced by the mass dependence of the multiplicity fraction and the magnitude limit of our sample). Furthermore, the multiplicity fraction is not significantly correlated with the clusters' position in the Galaxy. The spread in multiplicity fraction decreases significantly with the number of cluster members (used as a proxy for cluster mass). We also find that the multiplicity fraction decreases with metallicity, in line with recent studies using field stars.

**Key words.** Galaxy: open clusters, Galaxy: evolution, Galaxy: solar neighbourhood, methods: data analysis, statistical, stars: binaries general

## 1. Introduction

The star formation mechanism leads to the creation of a significant fraction of pairs of stars and higher-order hierarchies (e.g. Tokovinin & Moe 2020). Determining how frequent those multiple stellar systems are is not straightforward because of the observational difficulties in identifying close pairs and distant binary systems (e.g. Abt & Levy 1976).

Open star clusters (OCs) are highly suitable for studying stellar binary systems, because OC member stars have, to first order, the same distance, age, initial chemical composition, and foreground extinction. Although most of their binaries are unresolvable in images, they can be identified in the colour-magnitude diagram (CMD): they move above the main sequence (MS) depending on their mass ratio  $q^1$ , with the equal-mass binaries being  $-0.753$  mag brighter than their MS counterparts. The binary fraction in open clusters has therefore been an active field of study for decades (e.g. Mermilliod & Mayor 1989; Sandhu et al. 2003; Bica & Bonatto 2005; Sharma et al. 2008).

The characterisation of the multiplicity fraction in OCs and field stars is relevant to several branches of astrophysics (for re-

views see e.g. Duchêne & Kraus 2013; Moe & Stefano 2017; Offner et al. 2022). Correctly accounting for unresolved binary stars is essential in determining the total mass and stellar mass function of OCs (e.g. Kroupa 2001; Borodina et al. 2019; Rastello et al. 2020). They also influence the OC's dynamical evolution, providing relevant information about the outcome of star forming processes in different environments and constraints on the initial OC's state (e.g. Li et al. 2020). Furthermore, binaries are responsible for high rates of stellar collisions (Fregeau et al. 2004; González et al. 2021), black-hole mergers (Belczynski et al. 2002) and tracers of intermediate-mass black holes (Aros et al. 2021), thus proving relevant for high-energy astrophysics and gravitational-wave studies (Banerjee 2022).

Since typically Galactic OCs are disrupted on a time scale of a few hundred Myr (Lada & Lada 2003; Lamers et al. 2005), they tend to be relatively young (Wielen 1971; Anders et al. 2021). For young massive clusters in the local universe, the measured multiplicity fractions are close to those of field stars (González et al. 2021). In this paper we attempt to measure the high mass-ratio ( $q > 0.6$ ) multiplicity fraction for an unprecedented number of Galactic OCs located closer than 1.5 kpc to the Sun in an automated fashion, making use of the exquisite *Gaia* photometry (Riello et al. 2021). Our main input datasets are the recent *Gaia*

<sup>1</sup>  $q = M_2/M_1$ , where  $M_1$  is the mass of the primary component (i.e., the one with the larger initial mass),  $M_2$  is the mass of the secondary and  $0 \leq q \leq 1$ .

OC catalogues published by Tarricq et al. (2022, hereafter T22) and Cantat-Gaudin et al. (2020, hereafter CGa20).

The paper is structured as follows: In Sect. 2 we introduce the problem of estimating the multiplicity fraction in OCs. The treatment of the *Gaia* data is described in Sect. 3. In Sect. 4 we explain our method to measure the multiplicity fraction in each cluster. In the following Sect. 5 we validate the method (and determine the effective mass-ratio limit,  $q_{\text{lim}}$ , to which our method is sensitive) using custom OC simulations performed with the *Gaia Object Generator* (Luri et al. 2014). We present and discuss our results in Sect. 6 and conclude in Sect. 7.

## 2. Setting the problem

The multiplicity fraction  $f_b$  of a stellar population (or, obviating higher-order systems, the binary fraction), is defined as:

$$f_b = \frac{B + T + \dots}{S + B + T + \dots}, \quad (1)$$

where  $S$  is the number of single stars,  $B$  the number of binary systems,  $T$  the number of triple systems, and so on. The principal aim of this study is to estimate the  $f_b$  of MS unresolved multiple systems in OCs, using *Gaia*'s  $G$  vs. ( $BP - RP$ ) CMD.

An unresolved binary system composed of two identical main-sequence stars ( $q = 1$ ) has the same colour but twice the luminosity of an equivalent single star, and appears in the OC's CMD 0.753 mag brighter than the equivalent single star location (irrespective of the photometric band). A system with two unequal main-sequence components ( $q < 1$ ) has a combined colour that is redder than the colour of the primary component, and a combined luminosity greater than the one of a single star but less than the one corresponding to the equal-mass binary system. Therefore, such a system is displaced in the CMD both upwards (but not reaching 0.753 mag) and to the right relative to the main-sequence position of the primary component (see e.g. Fig. 1 in Milone et al. 2012). With  $q$  increasing from 0 to 1, the position of the binary system with respect to the main-sequence position of the primary component always decreases in magnitude, but for  $0 < q < q_{\text{crit}}$  moves towards redder colours, and for  $q_{\text{crit}} < q < 1$  towards the blue again (Maeder 1974; Hurley & Tout 1998). The value of  $q_{\text{crit}}$  depends on the mass of the primary component, caused by the fact that the shifts in magnitude and colours of a system of  $q < 1$  depend on the magnitude and colour difference of the two components. Hence, the binary sequences of a certain  $q < 1$  do not have a constant separation from the equal-mass binary sequence throughout all the MS. In most of the CMD MS range, there is not much separation between the  $q = 0$  and  $q = 0.5$  isochrones, so the low- $q$  unresolved binaries create an overdensity near the  $q = 0$  sequence (single-star main sequence, SS). On the other hand, the secondary sequence (binary sequence, BS) observed in most CMDs above the SS comprises not only equal-mass systems, but also high- $q$  unresolved binaries. Hence, the approximate location of a system on the binary sequence does not imply that it is an equal-mass binary (as first shown in Hurley & Tout 1998).

$q_{\text{crit}}$  does not exactly coincide with the smallest  $q$  for which a binary system lies on the binary sequence ( $q_{\text{lim}}$ ). From the observational perspective,  $q_{\text{lim}}$  is a more interesting parameter to consider: it is the mass ratio above which a binary system can be detected as such from photometry alone. Other effects like differential reddening, stellar rotation, and blends can potentially also displace a single star from its location in the MS towards redder and/or brighter positions mimicking multiplicity.

## 3. *Gaia* cluster member data

### 3.1. OC memberships and parameters

Since we are interested in an OC census that is representative for the extended solar neighbourhood, we pre-select all 368 local ( $d < 1.5$  kpc) OCs from the *Gaia* EDR3-based (Gaia Collaboration et al. 2021) catalogue of Tarricq et al. (2022). For OCs younger than 50 Myr (for which T22 do not provide improved membership lists), we supplement this dataset with the *Gaia* DR2-based catalogue (Gaia Collaboration et al. 2018) of CGa20, pre-selecting its 302 OCs younger than 50 Myr and closer than 1.5 kpc. In both cases, we use the homogeneously derived astrophysical parameters (age, distance, extinction) published in CGa20. We have, therefore, a preliminary sample of 670 local OCs, where the membership lists are taken from Tarricq et al. (2022) for OCs older than 50 Myr and from CGa20 for OCs younger than 50 Myr. 377 out of the 670 OCs have at least 30 MS members and a MS that extends over at least 1 mag in  $BP - RP$  and are considered for a detailed analysis (see below).

Both membership catalogues are based on astrometric data limited to  $G < 18$ .<sup>2</sup> Regarding the considered members of each OC, we have used all the members in the CGa20 catalogue that have membership probabilities  $\geq 70\%$  and have selected those of T22 with membership probabilities above the same threshold value. However, as the criteria for the membership probability assignment are not equal, establishing this common threshold does not make the selection exactly equivalent for both catalogues. After the further selections described in Sect. 4, a final sample of 202 OCs is obtained. Figure 1 shows their spatial and age distribution.

### 3.2. Selection of OC main-sequence members

We adopted a homogeneous selection of the main sequence of each OC in our sample. First of all, following CGa20 the absolute magnitude  $M_G$  and intrinsic colour index  $(BP - RP)_0$  were calculated for each cluster member (excluding the ones with missing  $BP$  and/or  $RP$  photometry):

$$M_G = G - \mu - 0.89 \cdot A_V, \quad (2)$$

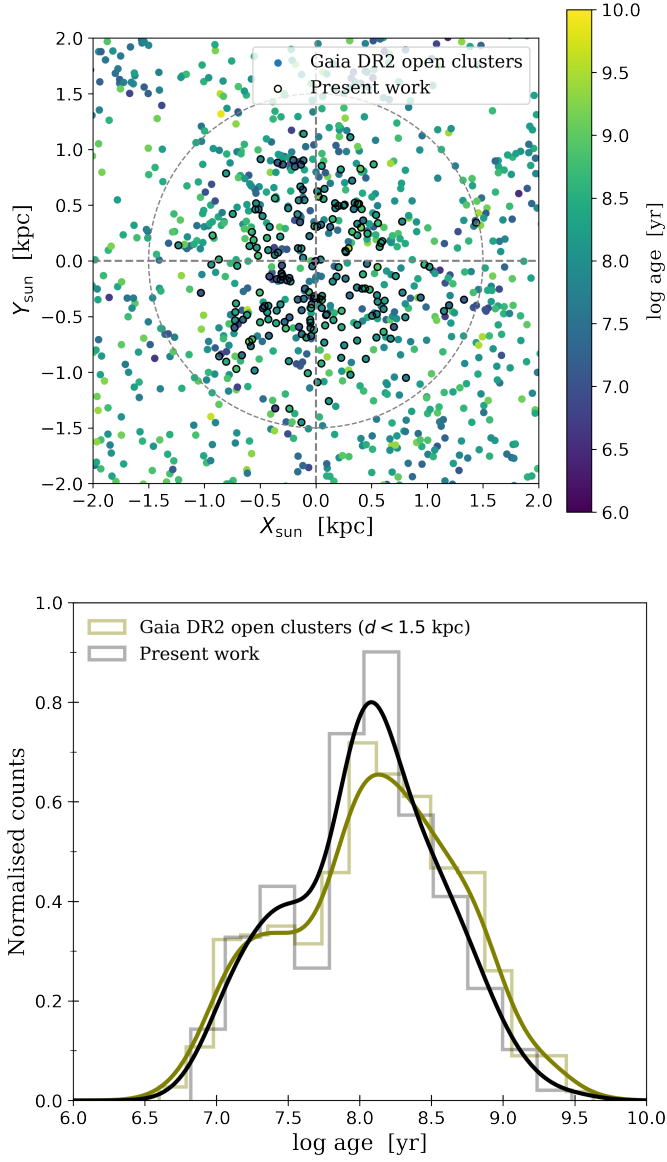
$$(BP - RP)_0 = (BP - RP) - \frac{0.89}{1.85} \cdot A_V, \quad (3)$$

where  $G$  is the *Gaia*  $G$ -band mean magnitude,  $\mu$  is the distance modulus from CGa20, and  $A_V$  is the OC's visual extinction (also from CGa20 and corrected by a factor in equation 2 to obtain the extinction in the  $G$ -band, and by a different factor in equation (3) to obtain the  $E_{(BP-RP)}$  colour excess). With these two quantities, we can now display the dereddened CMD for the OCs in our parent sample. We now derive restrictions (simple cuts in the dereddened CMD) for each member of a particular OC to be considered a MS member. The procedure (described below) is illustrated for two OCs of different ages in Fig. 2.

As our multiplicity fraction determination involves only MS stars, two straight parallel lines which lie above and below the MSs of all clusters can be drawn in the CMD. This results in a first rough selection that excludes red-clump stars, white dwarfs, and some extreme outliers (black lines in Fig. 2):

$$2.9 \cdot (BP - RP)_0 - 1.4 < M_G < 2.9 \cdot (BP - RP)_0 + 3.4 \quad (4)$$

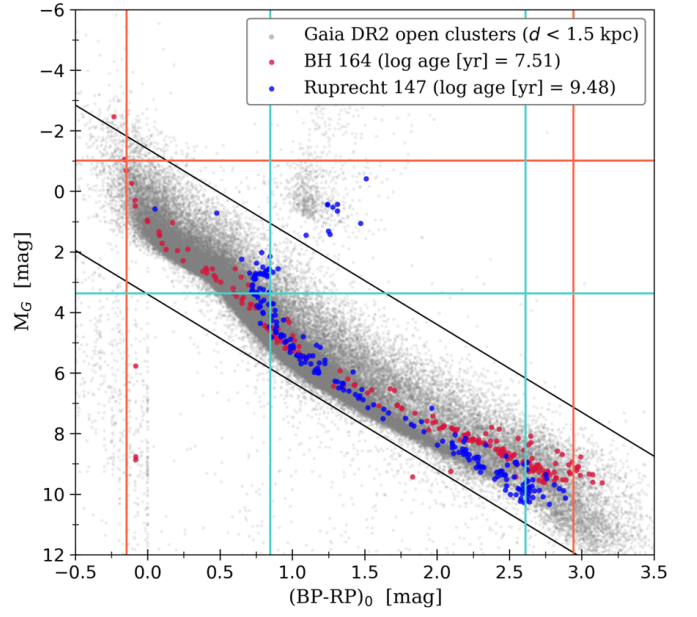
<sup>2</sup> Penoyre et al. (2022) showed how the proper motions and parallax of unresolved binaries can be biased depending on the period. This can sometimes result in true member stars being misclassified as field stars.



**Fig. 1.** *Top panel:* spatial distribution of the open clusters in the *Gaia* DR2-based catalogue of CGa20 (points colour-coded by age) and of the 202 OCs in our final sample (points encircled by a black circumference) in heliocentric Cartesian coordinates. The Galactic Centre is towards the right. *Bottom panel:* Normalised age distributions for the local OCs in CGa20 (738 OCs with  $d < 1.5$  kpc, olive histogram and kernel-density estimates); 202 of them constitute our final sample of OCs whose membership list is taken from either T22 or CGa20 depending on the age) and for our final sample of 202 OCs (black histogram and kernel-density estimates).

Another restriction is required to exclude members which are already evolving towards giant stars. In order to estimate the intrinsic colour index of the MS turn-off (MSTO), we fitted an interpolation polynomial  $f(\log \tau)$  to the  $(BP - RP)_0$  dependence on the age ( $\tau$ ) for the bluest point of the MSs of 15 different PARSEC 1.2S isochrones (Bressan et al. 2012) of solar metallicity and with ages in years in the range  $\log \tau \in [6.6, 9.8]$ . So, for each OC, the dereddened colour of the MSTO was obtained as a function of its age as  $(BP - RP)_0^{TO} = f(\log \tau)$ .

Then, MSTO and blue-straggler stars were excluded by only considering members at least 0.2 mag redder than the MSTO



**Fig. 2.** Two examples showing the selection of main-sequence members from the dereddened CMD. The black lines indicate the exclusion mainly of giants and white dwarfs. The red and blue lines indicate the cuts to exclude MS turn-off stars and to avoid incompleteness at the red end for the BH 164 and Ruprecht 147 (respectively).

(i.e., members with  $M < M_{TO}$  were selected):

$$(BP - RP)_0 > (BP - RP)_0^{TO} + 0.2 \quad (5)$$

and also imposing the extra condition that their magnitude was fainter than the one of the bluest member selected in the previous step minus 0.75 mag:

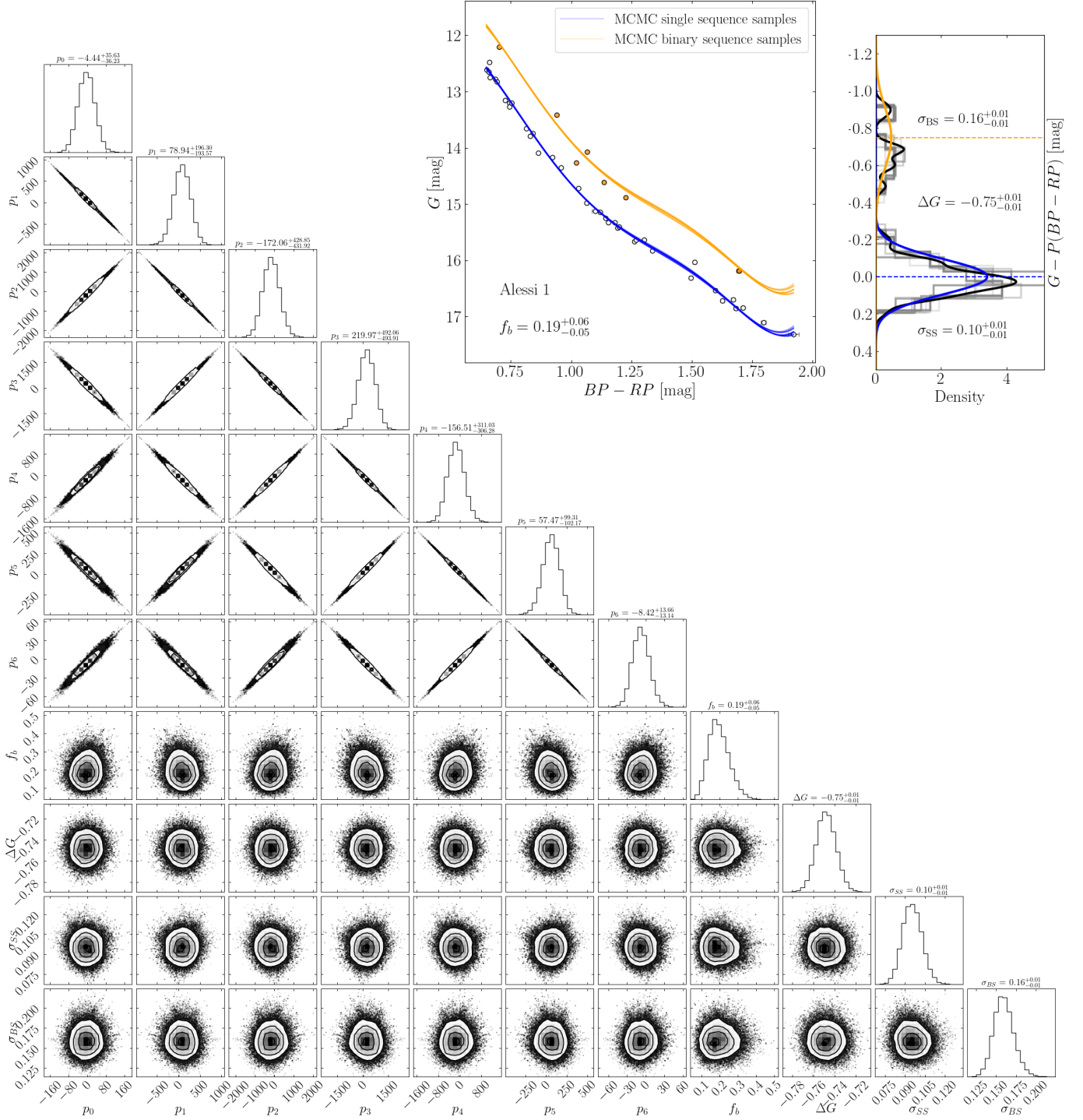
$$M_G \geq M_G((BP - RP)_{0,\text{bluest member}}) - 0.75 \quad (6)$$

Such conditions were found to succeed in excluding from the study not only the MSTO members, but also those extreme MS members for which the binary sequence approaches the single-star main sequence and eventually intersects it. For these cases our mixture-model algorithm (see Sect. 4) would not work because it is based on assuming a roughly constant separation between both sequences.

Finally, the redder MS members of each OC were excluded to avoid an overestimation of the multiplicity fraction due to the magnitude limit ( $G < 18$ ) in the parent cluster membership catalogues. We therefore selected only stars 0.2 mag bluer than the reddest star:

$$(BP - RP)_0 \leq (BP - RP)_{0,\text{reddest member}} - 0.2. \quad (7)$$

The subsequent study is carried out only for those OCs having a number of retained MS members larger or equal than 30, and, at the same time, a MS length in  $(BP - RP)_0$  longer than or equal to 1 mag (computed as the rest of the maximum and minimum values of  $(BP - RP)_0$  within the selected members). We now have an input sample of 377 OCs closer than 1.5 kpc: 250 OCs older than 50 Myr with EDR3 memberships from T22 and 127 OCs younger than 50 Myr with DR2 memberships from CGa20.



**Fig. 3.** MCMC fit for Alessi 1, which has 46 selected MS members. *Lower left panels:* corner plot of the posterior. The first 7 parameters correspond to the polynomial coefficients  $p_i$ , while the last four describe the unresolved multiplicity fraction  $f_b$ , the separation in  $G$  magnitude between the binary sequence and the single-star sequence  $\Delta G$ , the intrinsic  $G$ -magnitude spread of the single-star sequence  $\sigma_{SS}$ , and the one of the binary sequence  $\sigma_{BS}$ . The main outcome of the fit is the unresolved multiplicity fraction  $f_b$ , all the other parameters are considered nuisance parameters. *Top right panels:* CMD overplotted by 30 random samples of the posterior SS and BS, where the members  $> 3\sigma_{SS}$  brighter than the fitted SS polynomial are coloured in orange (left). The corresponding residuals with respect to the polynomial fit are also shown as histograms and kernel-density estimates, together with the resulting Gaussian mixture fit (right).

#### 4. Fitting the single-star and binary sequences

Studying stellar multiplicity in a statistically robust manner is a highly non-trivial task (e.g. Duquennoy & Mayor 1991; Belloni et al. 2017). The mass ratio of binaries, for instance, is intercon-

nected with the primary mass, the orbital period, the eccentricity, and the system metallicity (Offner et al. 2022).

Following the path laid out in Hogg et al. (2010), we can write down a mixture model for the distribution of cluster members in the CMD. If the functional form of the single-star main

sequence (SS) in the *Gaia* CMD is perfectly known (i.e. we know the functional form  $G_{SS} = f(BP - RP)_{SS} + \text{scatter}$ ), the cluster CMD (at least the part sufficiently redder than the MSTO) can be described as a simple mixture model of two populations that are described by the same functional form (separated by a constant offset  $\Delta G = G_{BS} - G_{SS} \approx -0.75$  mag) plus their respective intrinsic scatter in  $G$  magnitude. One population mostly accounts for single stars, resolved binaries and low- $q$  unresolved binaries, seen in the CMD as a widened single-star sequence approximately bottom bounded by the single-star sequence locus. The other population mostly accounts for unresolved binaries of  $q_{\text{lim}} < q \leq 1$ , seen in the CMD as a widened binary sequence, BS, approximately bound by the equal-mass binary sequence locus. In a Gaussian approximation for their scatter in  $G$  magnitude, the unresolved  $f_b$  is then just the weight of the binary sequence Gaussian:  $f_b(q > q_{\text{lim}}) = \frac{w_{BS}}{w_{SS} + w_{BS}} = w_{BS}$  (if both weights are normalised:  $w_{SS} + w_{BS} = 1$ ). Assuming these distributions to be Gaussians is a first-order approach.

In fact, this is the integrated high- $q$  unresolved  $f_b$  in the MSs of OCs, because two Gaussian distributions in  $G$  magnitude are considered for all the MS systems. And, as each Gaussian has a fixed standard deviation, the  $q$  range each of them comprises varies as a function of  $M_1$ ; and hence the  $q_{\text{lim}}$  below which multiple systems are regarded as simple systems varies slightly as a function of  $BP - RP$  (Hurley & Tout 1998). We, however, provide a single mean  $q_{\text{lim}}$  value for each OC, having integrated over the colour index and the other dimensions introduced by  $q$  dependencies. We also assume negligible contamination by field star interlopers, a valid assumption for the exquisite *Gaia* OC catalogues in the vast majority of cases; and triples and higher-order systems cannot be distinguished from binaries by our method (being included in the binary sequence population), so the estimated  $f_b(q > q_{\text{lim}})$  includes their contribution.

The likelihood for the CMD distribution below the turn-off can be described, therefore, by an arbitrary polynomial  $P(BP - RP)$  with a small intrinsic  $G$ -magnitude scatter for the SS,  $\sigma_{SS}$ , and a parallel one (perhaps with a slightly larger intrinsic scatter  $\sigma_{BS}$ ) for the BS. If we further assume the observational errors in  $BP - RP$  colour and  $G$  magnitude to be uncorrelated, the likelihood can be written as

$$\ln \mathcal{L} = \sum_{i=1}^N \ln \left[ \frac{(1 - f_b)}{\sqrt{2\pi(\sigma_{\text{eff},i}^2 + \sigma_{SS}^2)}} \cdot e^{-\frac{(M_{G_i} - P(BP-RP)_i)^2}{2(\sigma_{\text{eff},i}^2 + \sigma_{SS}^2)}} + \frac{f_b}{\sqrt{2\pi(\sigma_{\text{eff},i}^2 + \sigma_{BS}^2)}} \cdot e^{-\frac{(M_{G_i} - P(BP-RP)_i + \Delta G)^2}{2(\sigma_{\text{eff},i}^2 + \sigma_{BS}^2)}} \right], \quad (8)$$

where, in analogy to Chapter 7 of Hogg et al. (2010), we have introduced the effective uncertainties  $\sigma_{\text{eff},i}$  (the combined uncertainty projected onto the polynomial fit), defined as

$$\sigma_{\text{eff},i}^2 := \frac{1}{1 + m_i} (\sigma_{G,i}^2 + m_i \cdot \sigma_{(BP-RP)_i}^2) \quad \text{with} \quad m_i := \left( \frac{\partial P(x)}{\partial x} \right)^2 \Big|_{x=(BP-RP)_i}. \quad (9)$$

The likelihood in equation (8) is thus a function of  $p + 5$  parameters, being  $p$  the order of the fitted polynomial  $P$  ( $p + 1$  nuisance parameters for the polynomial itself, the intrinsic width of the SS,  $\sigma_{SS}$ , the intrinsic width of the BS,  $\sigma_{BS}$ , the vertical offset between the two sequences  $\Delta G$ , and the cluster's high- $q$  unresolved multiplicity fraction  $f_b$ ).

Since we are interested in sampling the posterior PDF of this parameter space, we also need to impose some priors. To leave maximal possible freedom to the fitting algorithm, we do not impose priors on the polynomial coefficients and only quantify our a-priori knowledge of stellar evolution. In particular, we assume Gaussian priors for the vertical offset  $\Delta G$  and the logarithm of the width of the SS and the BS, respectively:

$$p_1(\Delta G) = \mathcal{N}(-0.75, 0.05) \quad (10)$$

$$p_2(\log_{10} \sigma_{BS}) = \mathcal{N}(-1, 0.2) \quad (11)$$

$$p_3(\log_{10} \sigma_{SS}) = \mathcal{N}(-0.8, 0.2) \quad (12)$$

$$p_{\text{full}}(\Delta G, \log_{10} \sigma_{SS}, \log_{10} \sigma_{BS}) = p_1 \cdot p_2 \cdot p_3 \quad (13)$$

For each OC, we performed the Markov chain Monte Carlo (MCMC) fits using the python package *emcee* (Foreman-Mackey et al. 2013), using `n_walkers` = 32, `n_steps` = 10000 and `burnin` = 5000, and for  $p=6$ . With these conditions, the median and 16<sup>th</sup> and 84<sup>th</sup> percentiles of the  $p + 5$  likelihood parameters could be estimated for all 127 OCs from CGa20, and for 248 out of the 250 OCs from T22. In order to select only OCs for which the fitted polynomials and the Gaussian mixture model are accurate descriptions of the main-sequence CMD, we require the following (relatively strict) conditions:

1.  $\sigma_{SS} \leq 0.2$ ,
2.  $\sigma_{BS} \leq 0.25$ ,
3.  $|\Delta G + 0.75| \leq 0.05$ ,
4.  $\sigma_{f_b} \leq 0.25$  ( $\sigma_{f_b}$  is the mean of  $f_b$ 's 16<sup>th</sup> and 84<sup>th</sup> percentiles),
5. CMD is visually well fit (according to a visual inspection by the first three authors).

The result of these cuts is a final sample of 202 OCs whose unresolved  $f_b(q > q_{\text{lim}})$  has been estimated using MCMC, 146 ( $\sim 72\%$ ) from T22 and 56 ( $\sim 28\%$ ) from CGa20.

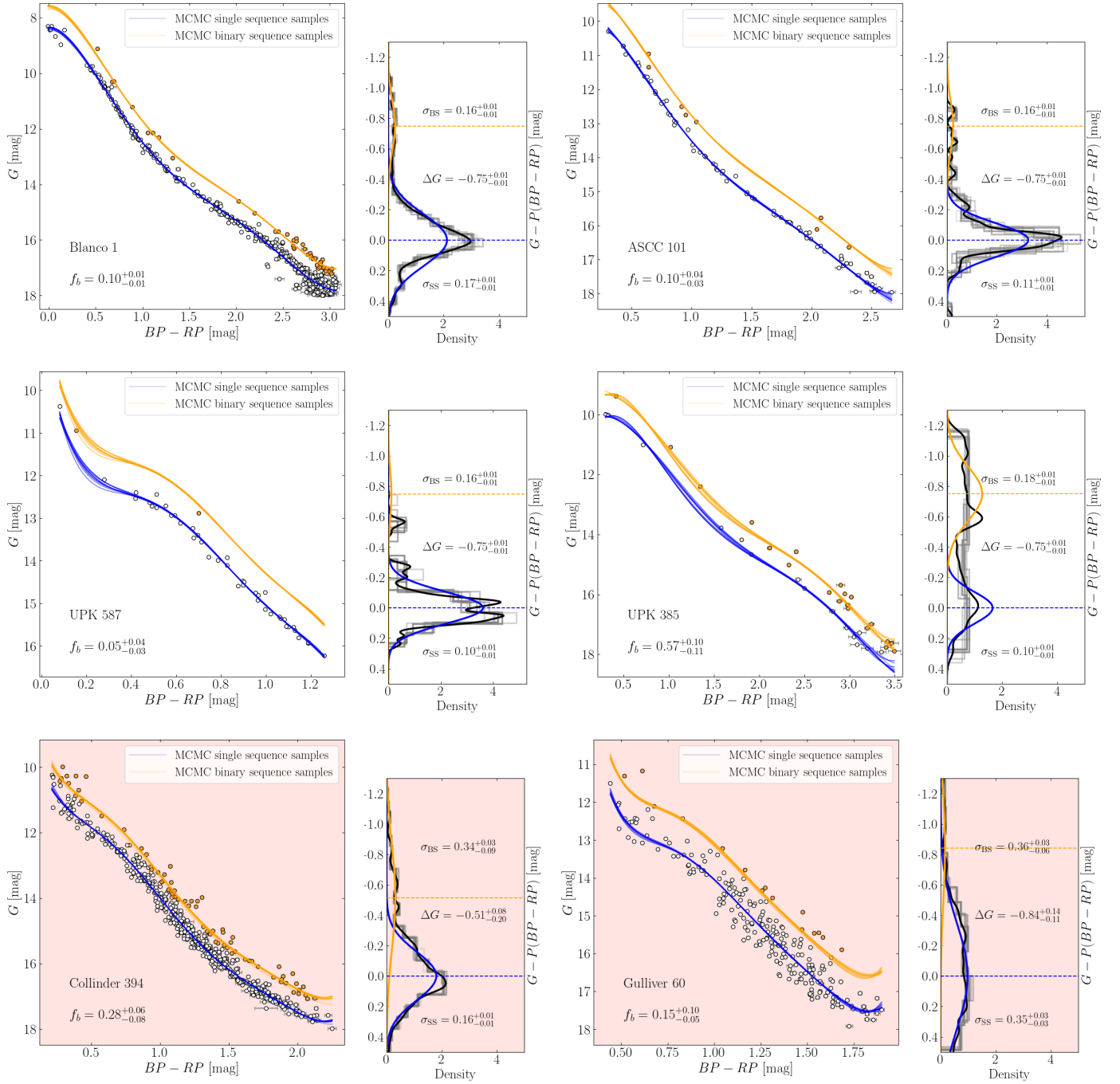
Figure 3 shows an illustrative example for a cluster with a moderately populated MS, Alessi 1. We see that in this case (as in most other well-fit OCs; see Fig. 4) apart from the polynomial coefficients there are little correlations between the fit parameters, in particular our main desirable parameter,  $f_b$ , is not strongly correlated with other parameters of the model, and the marginal posterior gives a sensible result ( $f_b = 0.19^{+0.06}_{-0.05}$ ). We note that the significant correlations between the polynomial coefficients in the top left part of Fig. 3 are not a problem (provided that the single-star sequence is accurately fit by the polynomial), since we are primarily interested in the estimate of  $f_b$ .

Our method has three advantages: 1. the fit can be performed directly in observable space (the observed *Gaia* CMD), 2. it correctly incorporates uncertainties in both dimensions, and 3. the form of the fitted SS and BS functions is very flexible and does not depend on stellar models at all. In the vast majority of the cases, the adjusted polynomials give an accurate description of the observed single-star main sequences. The characterisation of the mass ratio threshold  $q_{\text{lim}}$  below which multiple systems are regarded as simple systems by our method is discussed in Sect. 5.3.

## 5. Application to simulated clusters

In this section we describe the *Gaia Object Generator* (GOG; Luri et al. 2014) and how we use it to generate realistic simulations of the OCs with well-determined  $f_b$ . We apply our MCMC method for deriving the unresolved  $f_b(q > q_{\text{lim}})$  to 10 realisations of the simulated CMDs of 324 OCs. We are thus able to estimate  $q_{\text{lim}}$  for the application of MCMC to both the simulated





**Fig. 4.** Examples of the MCMC fits to the main sequences of several OCs (in the same style as Fig. 3 top right panels), illustrating the diversity of the binary sequences and the resulting diversity in the quality of the MCMC fits. The two examples marked by a light red background are excluded from the final catalogue because they do not fulfill the conditions described in Sect. 4.

and observed CMDs. Then this  $q_{\text{lim}}$  value is used to estimate the resolved multiplicity fraction in the same  $q_{\text{lim}} < q \leq 1$  range as the unresolved one; and also the corresponding total multiplicity fraction  $f_b^{\text{tot}}(q > q_{\text{lim}})$  (of both resolved and unresolved systems), for all the 202 OCs in our sample.

### 5.1. Simulating clusters with the Gaia Object Generator

The *Gaia Object Generator* (GOG; Luri et al. 2014) is a simulation tool that was developed to provide synthetic data that statistically reproduces the *Gaia* mission data (Gaia Collaboration

et al. 2016). For a given population of celestial objects, it applies the spacecraft and payload models to simulate the main *Gaia* observables (astrometry, photometry and spectroscopy) with realistic error models. Its complexity is increased after each *Gaia* data release, so that the simulated objects are in reasonable agreement with the data. GOG is typically used in conjunction with the *Gaia* Universe Model simulations (first described in Robin et al. 2012). The Universe Model relies on state-of-the-art descriptions of the characteristics of *Gaia* sources and on realistic scenarios for their formation, evolution, and dynamics. Most in-

interestingly for our purpose, it includes a module that simulates multiple star systems (Arenou 2011).

For each OC we generate a synthetic population of single stars drawn from PARSEC 1.2S isochrones (Bressan et al. 2012; Marigo et al. 2017) of solar metallicity with the age given by CGa20. The input stars are all single, generated according to the single-star IMF from Kroupa (2001) and Kroupa (2002). Now, the multiple-star module may change each input single star, according to some physically-motivated probability, into a system with the given star as primary and a lower-mass star of the same age as the secondary. Triples and higher order systems are not present in our simulations. The details of this process are described in Arenou (2011); in the next paragraphs we explain only the main assumptions.

The selection of single stars and primaries that will be part of a system is done so that they follow the luminosity function of primaries in the solar neighbourhood. For MS stars, the considered probability that an input single star gives birth to a system is given by the following function for the multiplicity fraction depending on the primary mass of the MS star, which is considered to fit well this dependence in the whole MS mass range:

$$f_b(M_1) = 83.88 \cdot \tanh(0.688M_1 + 0.079).$$

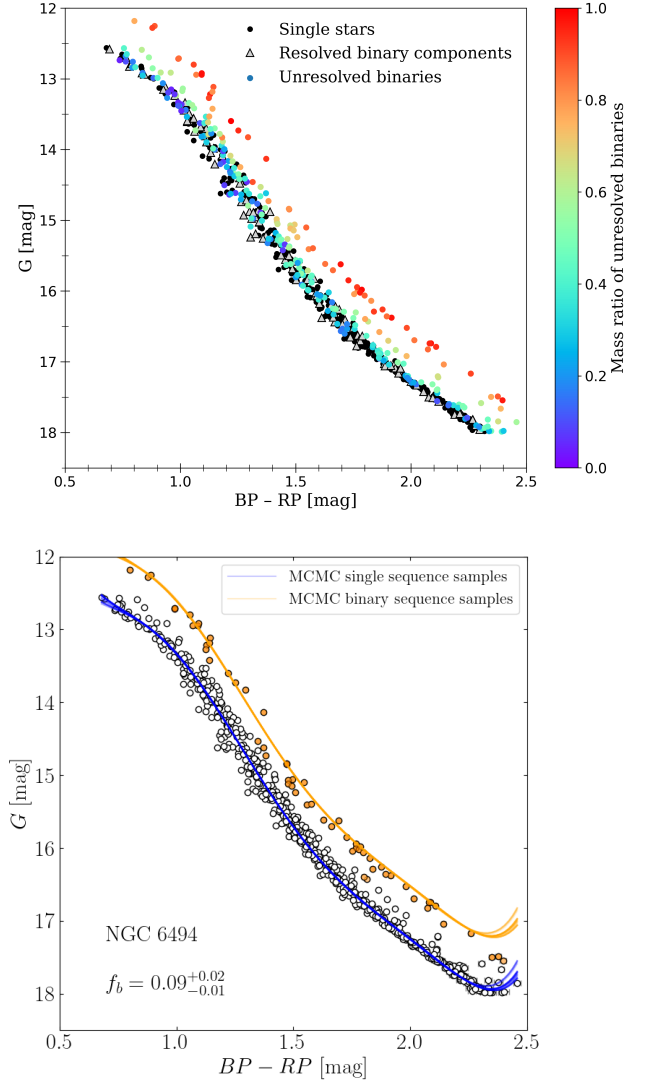
This function is monotonically increasing with  $M_1$ , and is roughly compatible (by eye) with the several classes of dynamical decay models from Sterzik & Durisen (2004) or random pairing of Thies & Kroupa (2007). There is also abundant observational evidence for the increase of  $f_b$  with  $M_1$  (e.g. Kaczmarek et al. 2011; Sana et al. 2012; Fuhrmann et al. 2017).

For each generated multiple system, the mass of the secondary is drawn from the mass-ratio distribution  $f(q)$ , which is modelled as a probability density function that is linear by segment and depends on the spectral type of the primary and on the binary period. For the distribution of semi-major axis  $a$ , a Gaussian distribution in  $\log(a)$  is assumed, with mean and standard deviation depending on  $M_1$ . From this random generation of  $a$ , the orbital period  $P$  is drawn using Kepler's third law. The eccentricities are assumed to be uniformly distributed within the interval  $[0, 2E[e]]$ , where  $E[e]$  is the average eccentricity (which depends on  $P$  and the primary's spectral type). The other orbital parameters are then drawn randomly: the periastron date  $T$  is chosen uniformly between  $[0, P]$ , the argument of the periastron  $\omega_2$  uniformly in  $[0, 2\pi]$ , the position angle of the node  $\Omega$  uniformly in  $[0, 2\pi]$ , and the inclination is chosen randomly in  $\cos(i)$ . A Roche model is used to avoid generating physically unrealistic systems of too small separations (for more details see Arenou 2011).

Finally, GOG decides whether a multiple system can be resolved by *Gaia* (so that the components appear all along the same isochrone in distinct positions) or not (so that their flux is joined and the system appears above the single-star MS as a single point in the CMD). To do so, GOG takes into account the angular separation of the components of the system (we have adopted 500 mas, following Fabricius et al. 2021). Finally, GOG also adds magnitude-dependent photometric uncertainties.

## 5.2. Custom OC simulations

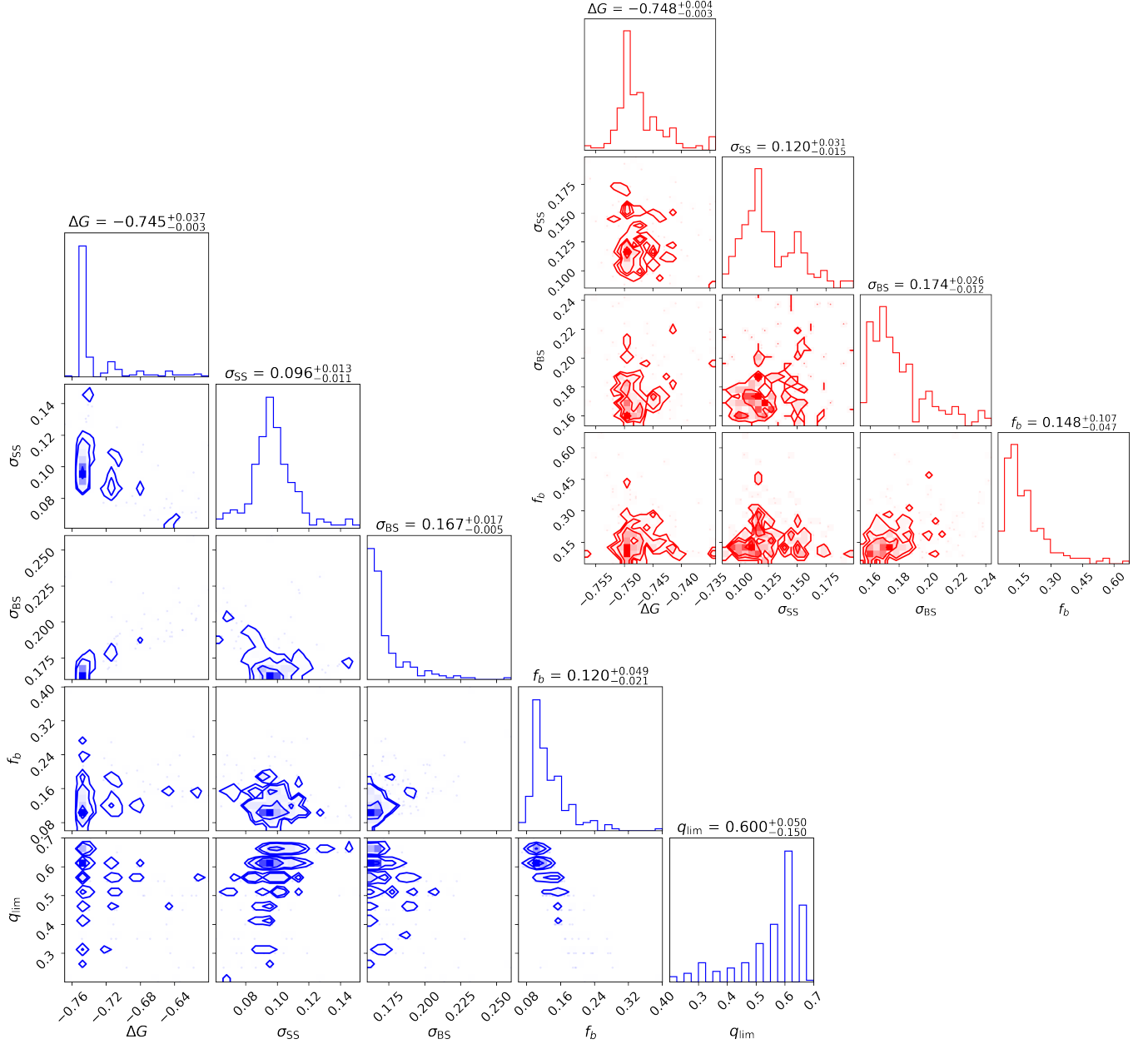
For 324 OCs (corresponding to a preliminary sample of okay-fit OCs) we have run custom simulations with the multiple-star module and GOG, applying them to each PARSEC single-star population of mass 10,000  $M_\odot$  with solar metallicity and the same age, distance, and extinction as the corresponding observed OC. The resulting synthetic population, now containing



**Fig. 5.** One realisation of the simulated CMD of NGC 6494 (log age/yr = 8.58, 682 selected main-sequence systems). *Top panel:* Colour-coded by mass ratio. The black dots are single systems, unresolved binary systems are colour-coded by their mass ratio  $q$ . Light grey triangles with black contours overlapping with single stars are resolved components of binary systems. The errorbars are smaller than the markers. *Bottom panel:* Result of the MCMC mixture-model fit, in the same style as Fig. 4. The simulated  $f_b$  for systems with  $q > 0.6$  for this realisation is 13%. The mean unresolved binary fraction over the 10 CMD realisations is  $(14 \pm 7)\%$ .

single stars and binaries, is returned by GOG as a catalogue of the members (or systems, if unresolved) with "true" and "observed" parameters. Photometric errors are taken into account, while other possible effects (most importantly differential extinction and rotation) that may contribute to the widening of the MS are not considered.

For each of these returned synthetic populations, we first select the MS members following the same procedure as for the observed OCs (see Sect. 3.2). To have the same number of members in the MS range as for the corresponding observed OC, we apply a random selection of MS members without replacement among all MS simulated members. A uniform probability in the random selection is adequate because the simulated population as seen by *Gaia* has already been populated according to an IMF



**Fig. 6.** Summary corner plots of the MCMC output parameters  $\Delta G$ ,  $\sigma_{ss}$ ,  $\sigma_{BS}$ , and  $f_b$  for the cleaned results corresponding to the 219 GOG OC simulations (blue plots) and for the 202 well-fit *Gaia* OCs (red plots). In the case of the simulations, also  $q_{lim}$  is shown.

for single stars, and then companions have been added. By doing so, the generated synthetic CMDs are approximate realisations of the corresponding OC. They automatically verify the conditions of having at least 30 MS members and a MS extension of 1 mag in colour index (see Fig. 5 for an example). We generate  $N_{sim} = 10$  realisations of each observed OC using the same simulated data but applying  $N_{sim}$  different times the random selection of the observed number of selected MS members. This allows to take into account statistically the effect that the random selection of a reduced number of members can have in the inferred  $f_b$ .

### 5.3. MCMC fitting to the simulated OCs

We perform the equivalent estimations of  $f_b$  through the mixture-model fitting described in Sect. 4, but now for a sample of 324 simulated OCs, 10 times each. We select those simulated OCs for which their CMDs have been well fit (according to the same criteria as in Sect. 4) for at least six out of the  $N_{sim} = 10$  realisations, and retain 219 (68%) of the simulated sample (144/219 are in common to the final sample of 202 observed OCs). The high- $q$  unresolved multiplicity fraction,  $f_b^{sim}$ , is calculated as the mean of the 10 values of  $f_b^{sim}(q > q_{lim}^{sim})$ 's median obtained through MCMC; and its nominal uncertainty  $\delta[f_b^{sim}]$  as the mean over the 10 realisations of the mean of its 16<sup>th</sup> and 84<sup>th</sup> percentiles. We also compute the standard deviation of  $f_b^{sim}$  over the 10 real-



isations ( $\sigma_{f_b^{\text{sim}}}$ ); both values give similar results and decrease for OCs with more MS members as expected.

We can also derive additional information: as GOG first simulates the systems, and later decides whether they can be resolved by *Gaia* or not, we know which of our selected MS members are single stars, which are two resolved components of the same binary system, and which correspond to unresolved binary systems. Therefore, we can compute the mean of the theoretical unresolved  $f_b$  over the 10 realisations:  $\overline{f_b^{\text{sim,theo}}}(q > 0)$ . We also compute the mean theoretical  $f_b$  of unresolved multiple systems having a  $q$  equal or greater than 11 different values:  $\overline{f_b^{\text{sim,theo}}}(q \geq q_{\text{lim}})$  (with  $q_{\text{lim}}$  in the range [0.2, 0.7] in steps of 0.05). These values are then compared to  $\overline{f_b^{\text{sim}}}$ , derived from the MCMC fits, to estimate  $q_{\text{lim}}^{\text{sim}}$ :  $q_{\text{lim}}^{\text{sim}}$  is assumed to be equal to the  $q_{\text{min}}$  for which  $\overline{f_b^{\text{sim,theo}}}(q \geq q_{\text{min}})$  is the closest possible to  $\overline{f_b^{\text{sim}}}$ .

For the sample with good MCMC fits, we determine a median value of  $q_{\text{lim}}^{\text{sim}} = 0.6^{+0.05}_{-0.15}$  (see Fig. 6). Although  $q_{\text{lim}}$  is not the same for all clusters, we can confidently say that for a random observed and well-fit OC, we are determining the multiplicity fraction of systems with  $q > 0.6^{+0.05}_{-0.15}$ .

#### 5.4. Estimation of the total multiplicity fraction: correction for resolved systems

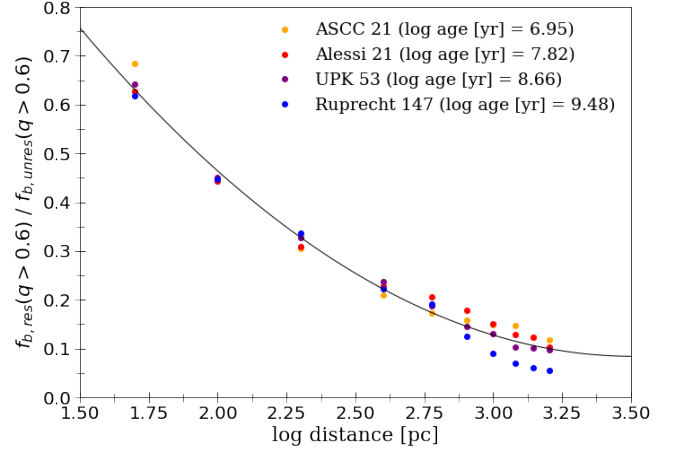
One of the most nearby OCs is the Pleiades cluster. It was recently studied in detail by Torres et al. (2021) using *Gaia* and long-term spectroscopic observations of thousands of stars. This work illustrates the complexity that one faces when trying to estimate the true  $f_b$  of an OC, even in such a very nearby case. They find, after applying corrections for undetected binaries, a binary frequency (for periods up to 104 days) of  $(25 \pm 3)\%$ . When including known astrometric binaries, this estimate increases to more than 57%.

In our study, we are limited to photometric observations and inherit the *Gaia* magnitude limit from the membership catalogues. Our method provides an estimation of the unresolved  $f_b$  of systems with  $q > q_{\text{lim}}$  based on their CMD positions, where  $q_{\text{lim}}$  is in principle characteristic of each OC. We therefore have to resort to simulations to correctly account for resolved systems, in order to be able to estimate the total  $f_b$ :  $f_b^{\text{tot}} = f_{b,\text{unres}} + f_{b,\text{res}}(d, \text{age})$ . As our unresolved  $f_b(q > q_{\text{lim}}) = f_{b,\text{unres}}^{\text{measured}}$  is limited to systems of  $q > q_{\text{lim}}$ , we also estimate  $f_b^{\text{tot}}$  of both unresolved and resolved binaries of  $q > q_{\text{lim}}$ , which can be calculated as follows:

$$f_b^{\text{tot}}(q > q_{\text{lim}}) = f_{b,\text{unres}}^{\text{measured}} \left[ 1 + \frac{f_{b,\text{res}}^{\text{sim}}(q > q_{\text{lim}}; d, \text{age})}{f_{b,\text{unres}}^{\text{sim}}(q > q_{\text{lim}}; d, \text{age})} \right] \quad (14)$$

We estimate it using the ratio between the resolved and unresolved  $f_b$  of the simulations instead of just adding up the measured unresolved  $f_b$  with the simulated resolved  $f_b$  in order to lessen the dependence on the total number of simulated binaries: what matters is the relative proportion between resolved and unresolved systems, not their absolute numbers.

We use the GOG simulations to estimate the dependence of the resolved over unresolved multiplicity fractions as a function of the distance and age. We simulate four OCs with ages covering the age range of the studied sample, placing them at 10 distances in the studied distance range, and compute their  $f_{b,\text{res}}^{\text{sim}}$  over  $f_{b,\text{unres}}^{\text{sim}}$  ratio for each distance, with a certain  $q_{\text{lim}}$  value. The result is represented in Fig. 7 for a value  $q_{\text{lim}} = 0.6$  (justified



**Fig. 7.** Ratio of the fraction of resolved binaries over the fraction of unresolved binaries, both of systems with  $q > 0.6$  ( $\frac{f_{b,\text{res}}^{\text{sim}}(q > q_{\text{lim}}=0.6; d)}{f_{b,\text{unres}}^{\text{sim}}(q > q_{\text{lim}}=0.6; d)}$ ), as a function of logarithmic distance, for four open clusters simulated with GOG. The black solid curve is the quadratic fit to the four open clusters that we use to compute the total multiplicity fraction.

and discussed in Sect. 5.3). As expected, the closer the cluster, the higher the fraction of resolved binaries. There is no strong dependence on the OC's age, so we fit a quadratic function to the dependence with the logarithm of the distance common for the four ages (black solid line in Fig. 7). This function is used to interpolate the value of the ratio  $\frac{f_{b,\text{res}}^{\text{sim}}(q > q_{\text{lim}}; d)}{f_{b,\text{unres}}^{\text{sim}}(q > q_{\text{lim}}; d)}$  for each OC in our sample, and used to estimate its  $f_b^{\text{tot}}(q > q_{\text{lim}})$ . We do not estimate the total  $f_b$  of all binary systems because we would need to add the unresolved and resolved binaries with  $q < q_{\text{lim}}$ , which requires a modelling of the  $q$  distribution and depends on the specific properties of each binary system of the OC.

## 6. Results and discussion

The results of our mixture-model fitting are included in a catalogue accessible via CDS. It contains the sample of 202 OCs well-fit by our mixture model (146 from T22 and 56 from CGa20). We thus present one of the largest homogeneous catalogues of estimated multiplicity fractions in OCs to date. For each OC, the catalogue provides the values and uncertainties of the unresolved multiplicity fraction deduced with MCMC and the estimated total multiplicity fraction (of both resolved and unresolved systems) of MS systems with  $q > 0.6$ . The catalogue also contains the intrinsic colour range of the selected MS systems and its translation into masses; as well as the values of the other fit parameters. The most populated OC has 1158 MS members, the least populated one 30. Only 6 well-fit OCs have at least 500 MS members. The median mass of the least massive MS member selected is  $0.71 M_{\odot}$ , and the median mass of the most massive MS member selected is  $2.03 M_{\odot}$ .

The values of  $f_b$  (before taking into account resolved binaries) that we find for the OC sample lie between 5% and 67%, with a median value of 15%. Their distribution (compatible with a log-normal distribution) is shown in the rightmost panel of Fig. 6. The typical uncertainties for  $f_b$  amount to 4%, and 95% of the OCs have uncertainties smaller than 10%.

A comparison with the simulations is illustrated in Fig. 6. For the simulated OCs, we obtain values of  $\overline{f_b^{\text{sim}}} \simeq (12^{+5}_{-2})\%$ , with typical errorbars comparable to but often slightly lower than for the observed clusters. The simulations also allow us to determine the typical minimum mass ratio to which our method is sensitive:  $q_{\text{lim}} = 0.6^{+0.05}_{-0.15}$ .

Our  $q_{\text{lim}}^{\text{sim}}$  of 0.6 is similar to the value found in the recent study of Jadhav et al. (2021): for their considered magnitude range ( $M_G \in [1, 10]$  mag; comparable to our range of selected MS members), they find  $q = 0.6$  to be the cut-off that ensures that the single MS stars are  $3\sigma$  away from the thus-defined BS.

We also estimate the total ( $q > 0.6$ ) multiplicity fraction, including resolved systems, for each OC (as explained in Sect. 5.4). For 89% of all the 202 OCs, this correction of additionally taking into account resolved systems (apart from the unresolved binaries already accounted for in the unresolved multiplicity fraction) is smaller than 5%. The total high-mass-ratio multiplicity fraction covers values from 6% to 80%, approximately following a log-normal distribution with a peak around 14% and a median of 18%. Only 9% of the OCs in the sample have an  $f_b^{\text{tot}}(q > 0.6) > 0.35$ . The median total multiplicity fraction uncertainty is  $\sim 5\%$ , and varies strongly from one OC to another.

### 6.1. Comparison to other open cluster studies

The multiplicity fractions in many nearby OCs have been studied before (e.g. Sharma et al. 2008; Sollima et al. 2010; Cordoni et al. 2018; Cohen et al. 2020); however, large homogeneous studies are still rare. The results of different studies, moreover, are very seldom directly comparable. Apart from the data and OC membership determinations used in each study, the main differences arise from the mass range covered by the studied OC members and the minimum  $q$  to which the estimated  $f_b$  is sensitive ( $q_{\text{lim}}$ ).

Besides, the estimated  $f_b$  also depends on the assumptions of the modelling applied and the treatment of outliers. As discussed in Dal Tio et al. (2021), the assumption that all binaries are unresolved is still a well-accepted approach to model binaries in CMD-fitting works. This approach generally provides  $f_b$  values very similar to the ones found for more detailed prescriptions for the binaries for distant OCs. Since the fraction of resolved binaries decreases with distance (see Fig. 7), the closer the studied OC, the less appropriate it is to blindly use the CMD-fitted  $f_b$  assuming all binaries to be unresolved.

In this subsection we compare the results of our  $f_b$  estimation with the literature for the OCs that we have in common with four of the most recent studies of  $f_b$  in OCs. As they assume all binaries to be unresolved, we compare our likewise estimated  $f_b$  with their results, instead of our calculated  $f_b^{\text{tot}}$  that takes into account the resolved binaries too. Niu et al. (2020), Jadhav et al. (2021) and Li & Shao (2022) rely on the CMD of  $G$  vs. ( $BP - RP$ ), as in our study, while Li et al. (2022) use  $V$  vs. ( $V - I$ ). The comparisons are summarised in Fig. 8.

Niu et al. (2020) used *Gaia* DR2 photometry and LAMOST spectroscopy (Zhao et al. 2012) to study the fundamental parameters of 12 well-populated OCs. From their fits they derived synthetic CMDs and inferred the  $f_b$  of MS stars for systems with  $0 < q < 1$ . They found the  $q$  distribution in their OCs to be flat, in accordance with the finding of Torres et al. (2021) for the Pleiades. They also provide the limiting values  $m_{\text{min}}$  (minimum mass estimated by the faintest MS star in the OC) and  $m_{\text{max}}$  (estimated by the brightest MS star in the OC) of the mass range

for which  $f_b$  has been estimated. The top left panel of Fig. 8 shows the comparison to our study for the three OCs we have in common. None of the studied mass ranges coincides and, as is shown in Fig. 8, our estimated  $f_b$  are smaller than the ones obtained by Niu et al. (2020) in all cases. This is consistent with the fact that these authors both study mass ranges up to higher masses (while  $m_{\text{min}}$  does not differ by much) and are sensitive to a much wider  $q$  range, thus including more (low mass-ratio) binaries. To account for the latter effect, in each of the panels in Fig. 8 the black dashed line shows the expected relation under the assumption of a flat mass-ratio distribution.

Furthermore, Niu et al. (2020) recalculate the  $f_b$  for the Pleiades and NGC 2099 in mass ranges coincident with the ones of other published studies. We also perform our  $f_b$  calculation in these two same mass ranges ( $[0.6, 1.0] M_{\odot}$  for the Pleiades and  $[1.06, 1.63] M_{\odot}$  for NGC 2099), so that we can compare our estimated  $f_b$  for each OC with two different values computed in the same mass range but for different  $q$  ranges.

For the Pleiades (Melotte 22), our inferred  $f_b(q > 0.6) = 0.086^{+0.012}_{-0.011}$  is still smaller than the one of Niu et al. (2020),  $f_b(q < 1) = 0.20 \pm 0.03$ , being compatible only within  $3\sigma$ . This is mostly attributed to the fact that their minimum detected  $q$  is lower, aside from intrinsic discrepancies. Our inferred  $f_b$  is also much smaller than the one inferred by the pre-*Gaia* study of Pinfield et al. (2003),  $f_b(0.5 \leq q \leq 1) = 0.23^{+0.06}_{-0.05}$ , despite being computed for a  $q$  range compatible with ours.

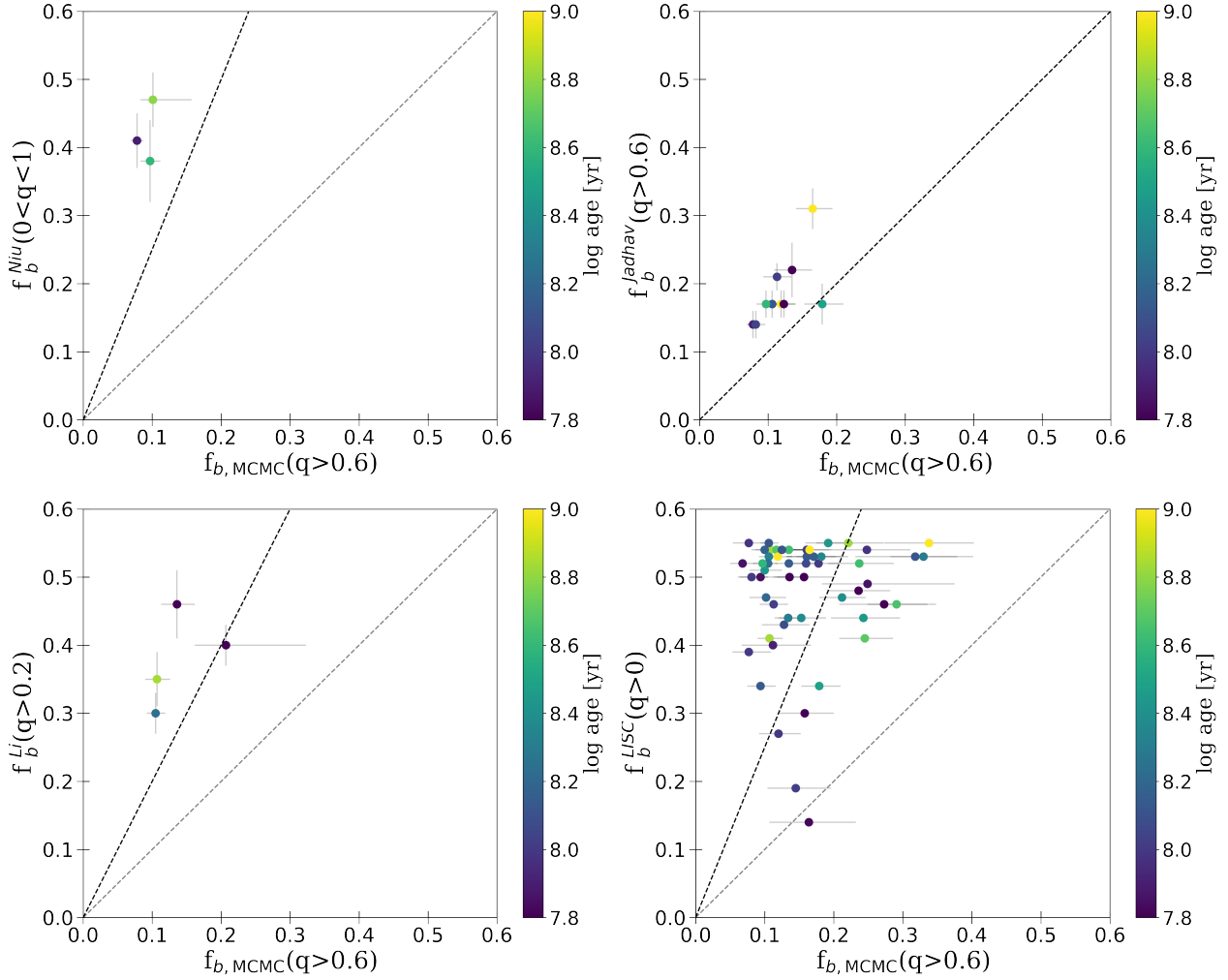
For NGC 2099, our inferred  $f_b(q > 0.6) = 0.11^{+0.03}_{-0.03}$  is also smaller than the one of Niu et al. (2020),  $f_b(q < 1) = 0.23 \pm 0.02$ , being compatible only within  $3\sigma$ . Compared to the work of Cordoni et al. (2018), however, our value and theirs ( $f_b(0.7 \leq q \leq 1) = 0.085$ ) are computed for very similar  $q$  ranges and are compatible within  $1\sigma$ .

The top right panel of Fig. 8 shows the comparison to Jadhav et al. (2021), for the ten OCs we have in common. These authors used interpolated PARSEC isochrones to calculate the magnitudes of the primary star, secondary star and the combined unresolved binary for various  $q$  values, and selected unresolved MS binaries of  $q > 0.6$  to compute the unresolved  $f_b$  in 23 OCs using *Gaia* DR2 data. They take into account, therefore, unresolved binaries in the same  $q$  range as we do. They do not provide, however, the mass range for which  $f_b$  is estimated, only the total number of MS members for each OC.

Their considered number of members is larger than ours for all OCs, although for some OCs they are comparable and for others our sample represents only a 30% of theirs. For 9 OCs our estimated  $f_b$  is smaller than theirs; while for only one OC (NGC 6793) ours is larger (although they are both compatible within  $1\sigma$ ). These discrepancies can be real if the studied mass range is different for some of the OCs, or can otherwise arise (at least partially) from the fact that they refer to different OC members.

Li & Shao (2022) apply a Bayesian framework that models the observed CMD of an OC as a mixture distribution of single stars, unresolved binaries and field stars, and then measure the fraction of unresolved binaries with  $q > 0.2$  among the member stars. They apply this method to 10 OCs with *Gaia* EDR3 photometric data. Figure 8 (bottom left panel) shows the comparison to our study for the four OCs we have in common. Their  $f_b$  is larger than ours for all OCs, which (as in the case of the comparison to Niu et al. 2020) is mainly due to the fact that they reach a lower  $q$  threshold.

Finally, Li et al. (2022) determine the primordial  $f_b$  fitting the CMD morphologies using the *Powerful CMD* code. They provide a catalogue, named LI team's Star Cluster (LISC), for



**Fig. 8.** Unresolved multiplicity fraction estimates from four literature studies as a function of our estimated unresoluted multiplicity fraction  $f_b(q > 0.6)$ , for the common open clusters between the two studies in each case. As a guidance for comparison, the equality line is drawn as a dashed grey line in each panel, and the black dashed lines show the expected relation when a flat mass-ratio distribution is assumed. *Top left:* Niu et al. (2020). *Top right:* Jadhav et al. (2021). *Bottom left:* Li & Shao (2022). *Bottom right:* Li et al. (2022).

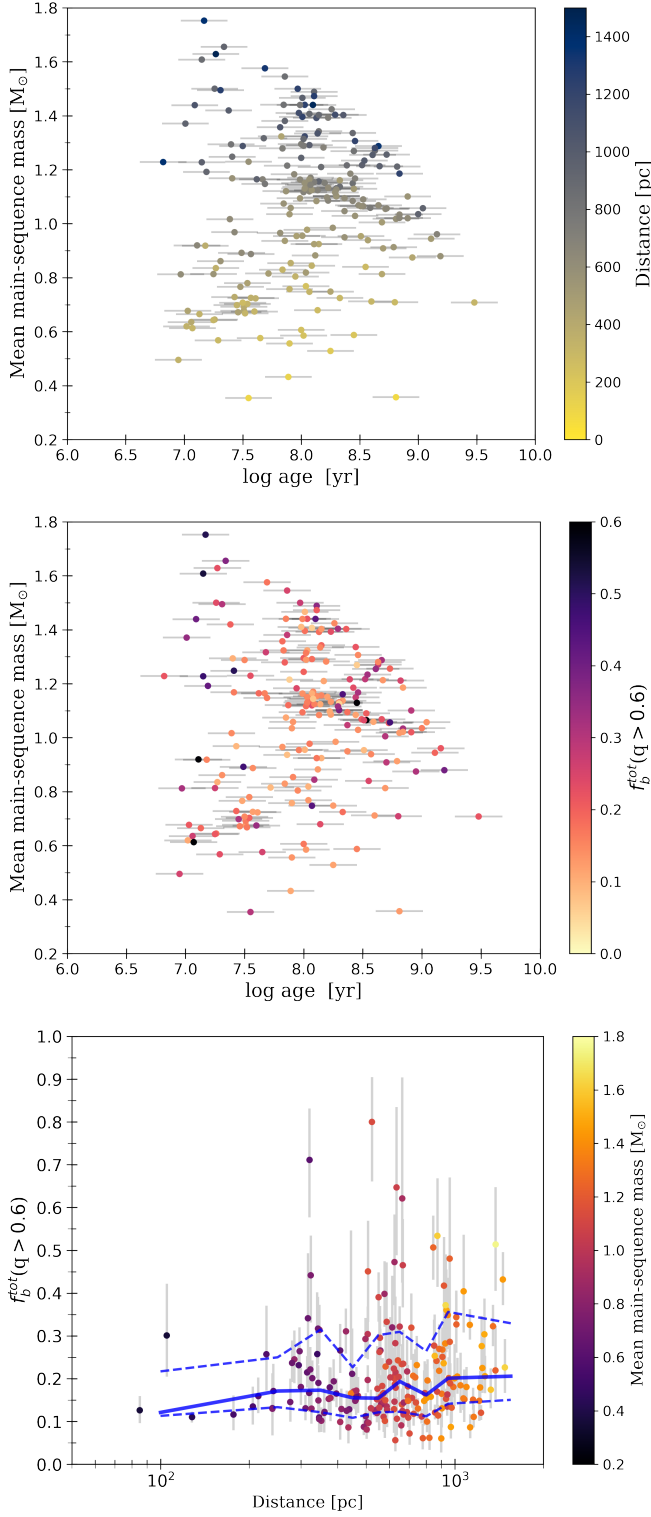
309 OCs for which they consider the fit to be good, and for 288 OCs for which it is not as good, without specifying either the mass or  $q$  ranges considered or the  $f_b$  uncertainty. Our study and their good-fit one have 53 OCs in common. For all but one, their estimated  $f_b$  is larger than ours and mostly incompatible with our values (see Fig. 8, bottom right panel). The study of Li et al. (2022) has the particularity of accumulating  $f_b$  values between 0.5 and 0.55, which hints that their  $f_b$  distribution might be mainly the result of an imposed prior and/or poorly constrained fits.

## 6.2. Dependence of the multiplicity fraction on the distance and stellar mass

Figure 9 (top panel) demonstrates that for our distance-limited sample, the cluster parameters age, distance, and mean sampled mass are heavily entangled. We therefore expect that sample selection effects dominate most of the possible trends of the inferred multiplicity fraction with any cluster parameter (e.g. cluster mass, age, position in the Galaxy). The well-known dependence of the multiplicity fraction on stellar mass (see Sect. 2.2. of Offner et al. 2022 and references therein), for example, is heavily imprinted in our sample, as we show below.

The bottom panel of Fig. 9 shows the dependency of the estimated total multiplicity fraction with  $q > 0.6$  on distance. Firstly, we see that the number of studied OCs increases with distance as expected (due the larger volume covered), with only 4 OCs out of the 202 well-fit ones located closer than 200 pc. Beyond this distance, there is a wide dispersion of  $f_b^{\text{tot}}$  values at any distance. Still, for both the unresolved and total multiplicity fractions, the running median displays a general tendency to increase with distance, because for more distant OCs we tend to see only the upper ends of their MSs, which have more massive stars and are thus expected to display a higher  $f_b^{\text{tot}}$ . The increase of  $f_b^{\text{tot}}$  with the mass of the primary star is in full agreement with abundant observational evidence (see e.g. Offner et al. 2022, Figs. 1 & 4). This effect is seen in the bottom panel of Fig. 9, where the colourbar corresponds to the mean MS mass of each OC, computed as a function of the limits of our studied mass range ( $m_{\min}$  and  $m_{\max}$ ) and Kroupa (2001) power-law IMF ( $\xi(m) \propto m^\alpha$  with  $\alpha = 2.3$  for  $m \geq 0.5 M_\odot$ ):

$$\langle m \rangle = \frac{\int_{m_{\min}}^{m_{\max}} m \xi(m) dm}{\int_{m_{\min}}^{m_{\max}} \xi(m) dm} = \frac{1 - \alpha}{2 - \alpha} \cdot \frac{m_{\max}^{2-\alpha} - m_{\min}^{2-\alpha}}{m_{\max}^{1-\alpha} - m_{\min}^{1-\alpha}} \quad (15)$$

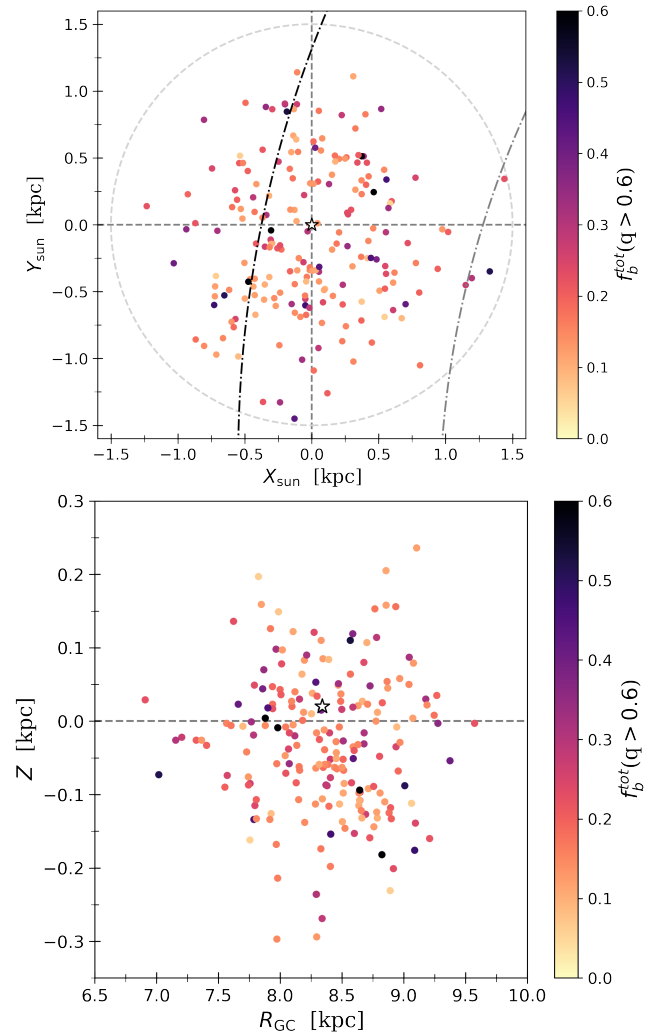


**Fig. 9.** Correlations between age, distance, mean main-sequence mass, and multiplicity fraction for the 202 studied open clusters. *Top panels:* Mean main-sequence mass of the open cluster as a function of logarithmic age, colour-coded by their distance (top panel) and by their total multiplicity fraction of systems with  $q > 0.6$  (middle panel). *Bottom panel:* Total multiplicity fraction of systems with  $q > 0.6$  as a function of logarithmic distance, colour-coded by the mean main-sequence mass of each open cluster. The central thick blue line is the running median, and the dashed blue lines are the 16<sup>th</sup> and 84<sup>th</sup> running percentiles.

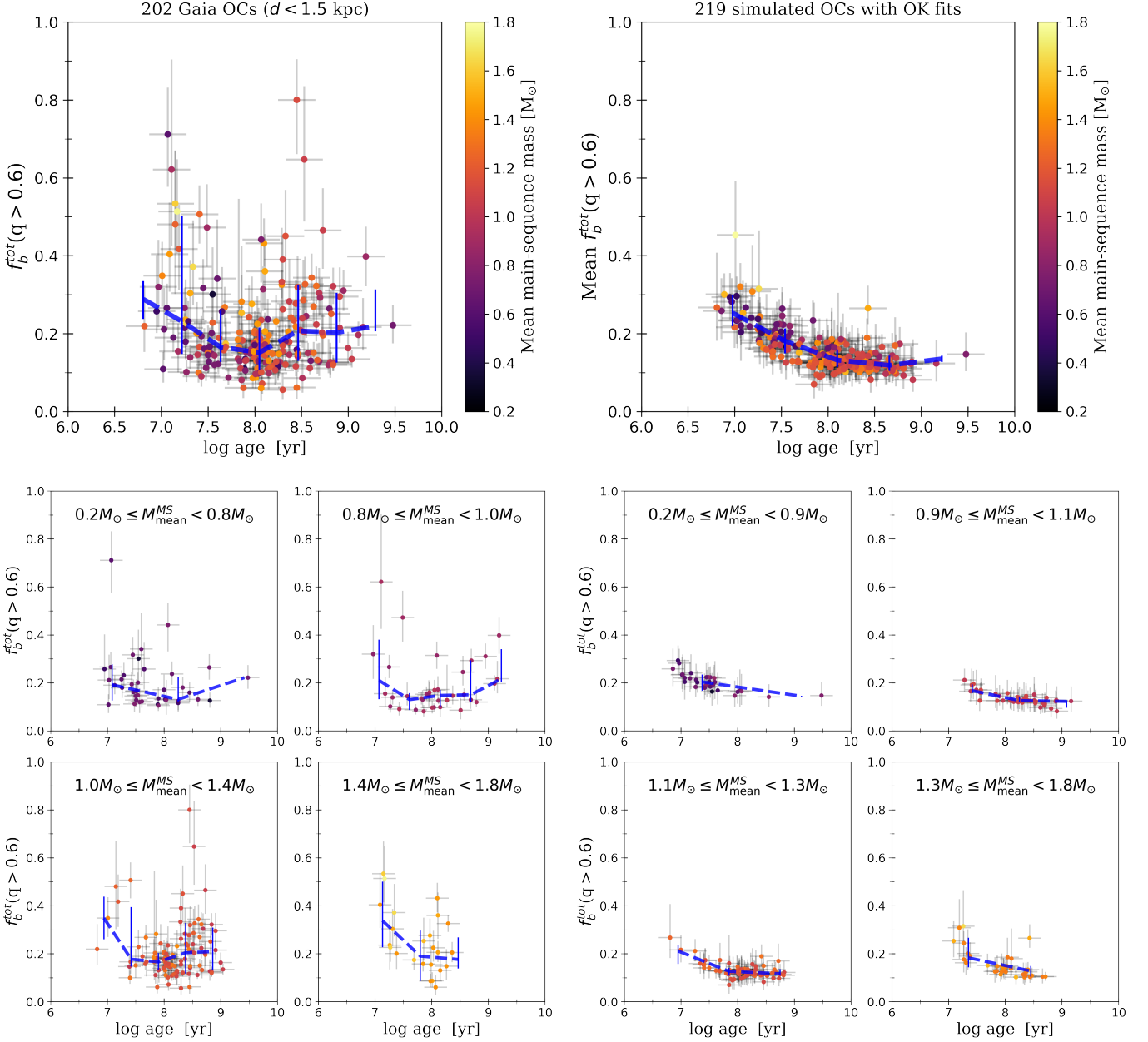
We find that for increasing distance, the observed  $m_{\text{max}}$  does not vary significantly, while  $m_{\text{min}}$  increases. For more distant OCs, only the upper MS ends are included in the magnitude-limited membership lists, thus preferring more luminous and massive stars (yielding thus a larger  $\langle m \rangle$ ) and consequently a higher  $f_b^{\text{tot}}$ . Hence, the tendency of  $f_b^{\text{tot}}$  increasing with distance in the bottom panel of Fig. 9 reflects the fact that the limiting apparent magnitude causes the observable portion of the OC’s MS to be dependent on the distance; and that the  $f_b^{\text{tot}}$  depends on the mass of the stars.

### 6.3. Dependence of the multiplicity fraction on position

Figure 10 displays the three-dimensional spatial distribution of our OC sample colour-coded by  $f_b^{\text{tot}}(q > 0.6)$ . The top panel shows the projection onto the Galactic plane (similar to Fig. 1). The OCs are approximately evenly scattered around the Sun, with a few of them being located close to the Local and the



**Fig. 10.** Locations of the 202 open clusters in our catalogue, colour-coded by the total high mass-ratio multiplicity fraction. *Top panel:* top-down view on the Galactic plane (within a distance of 1.5 kpc, indicated by the dashed circumference). The Galactic centre is located towards positive  $X_{\text{sun}}$ . The dash-dotted curves are the centres of the Local (black) and the Sagittarius–Carina arms (grey), as defined in Reid et al. (2019). *Bottom panel:* Height above the Galactic plane versus Galactocentric radius. The Sun’s position is indicated by an asterisk in both panels.



**Fig. 11.** Total multiplicity fraction of systems with  $q > 0.6$  as a function of logarithmic age, colour-coded by the mean main-sequence mass of each open cluster. *Top left panel:* For the 202 open clusters in our sample. The dashed blue line is the running median, and the errorbars are calculated as the 16<sup>th</sup> and 84<sup>th</sup> percentiles. *Top right panel:* The same for the 219 well-fit simulated open clusters. *Bottom panels:* The same as above, but now split into four bins of mean-sequence mass, both for the observed *Gaia* OCs (left) and for the simulated clusters (right).

Sagittarius–Carina spiral arms. Apart from the distance dependence discussed above, no apparent correlation between the total multiplicity fraction and the Galactic position is seen. The bottom panel shows the OC’s height above the Galactic plane as a function of the distance from the Galactic centre. All the OCs fall inside the solar neighbourhood scale height for the thin disc,  $z_d \approx 300$  pc. As before, no clear correlation is seen between  $f_b^{\text{tot}}$  and the position in the Galactocentric frame.

#### 6.4. Dependence of the multiplicity fraction on age

Figure 11 shows the dependence of  $f_b^{\text{tot}}(q > 0.6)$  on age, both for the *Gaia* sample and for its simulated counterpart. The dis-

tribution for the observed OCs (top left panel) is dominated by a large  $f_b^{\text{tot}}$  dispersion at all ages, which is not present in the simulated sample (top right panel) and therefore indicates a missing element of randomness in the GOG simulations. The running median decreases with age for the younger OCs up to approximately 100 Myr (log age/yr  $\approx 8.0$ ). For OCs older than log age/yr  $\approx 8.0$  the total multiplicity fraction increases slightly with age. The age range for which we find an inversion of the trend is compatible with Thompson et al. (2021) who found a flattening of the unresolved multiplicity fraction trend with age at  $\sim 200$  Myr in a very small sample of 8 OCs.

The mean  $f_b^{\text{tot}}(q > 0.6)$  vs. age distribution for the simulated OCs in the top right panel of Fig. 11 (where the mean  $f_b^{\text{tot}}(q >$



0.6) has been calculated with equation (14) for  $f_{b,\text{unres}}^{\text{measured}} = \overline{f_b^{\text{sim}}}$  reproduces at least the decreasing trend for ages smaller than 100 Myr. Therefore, this trend is not a physical one, but mainly caused by the interplay of the mass dependence of the multiplicity fraction with the complex selection function of our OC sample. Each of the panels in Fig. 11 is colour-coded by the mean main-sequence mass of the portion of the MS considered by our fit. When dividing the sample into four bins of mean main-sequence mass  $\langle m \rangle$  (lower panels of Fig. 11), the age trends seen in the full sample persist.

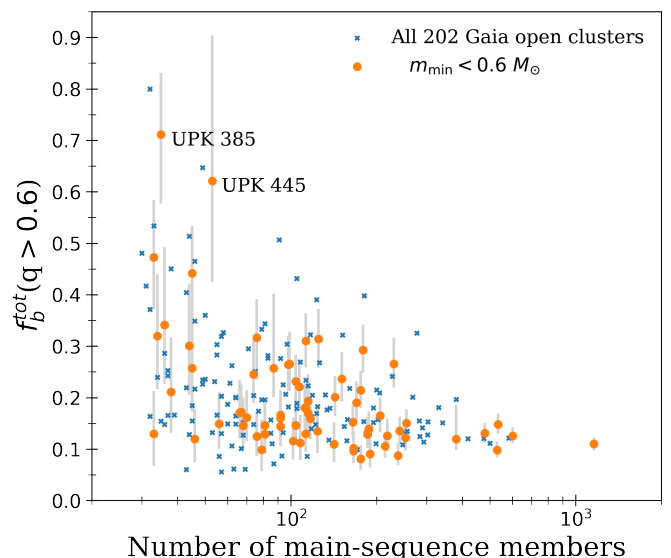
To better understand the trends seen in Fig. 11, it is also helpful to return to the top panel of Fig. 9: for a certain age, the further the OC, the shorter its upper MS portion observed, which is thus more massive (increasing  $\langle m \rangle$ ); whereas the closer the OC is, the observed MS extends from the upper end towards the less massive and more abundant stars that dominate (decreasing  $\langle m \rangle$ ). Focusing on the further distances instead, the older the OC, the smaller  $\langle m \rangle$ . This reflects that the MSTO is shifted towards redder  $BP - RP$  for older OCs, thus comprising less massive stars, and this is the reason why the top right corner of the top panel of Fig. 9 is empty.

From a dynamical-evolution point of view, a clear intrinsic dependence of the  $f_b^{\text{tot}}$  on OC age is also unexpected. For example, Hurley et al. (2007) showed with  $N$ -body simulations that the overall multiplicity fraction of a cluster "almost always remains close to the primordial value, except at late times when a cluster is near dissolution". Our simple GOG simulations (see Sect. 5) also show that it is not necessary to invoke efficient multiplicity disruption during the early evolution of a cluster to explain the observed trend of  $f_b^{\text{tot}}$  decreasing with age during the first  $\sim 100$  Myr.

### 6.5. Dependence of the multiplicity fraction on the number of cluster members

The physical quantities governing the dynamical evolution of a star cluster (modelled as a gravitational  $N$ -body system with stellar evolution) are the total mass, the initial mass function, the primordial multiplicity fraction, the central density, and the strength and variability of the external tidal field. Precise knowledge of most of these quantities is missing. However, it has been shown that clusters with similar remaining mass fractions (present-day mass divided by initial mass) have depleted their stellar mass functions by similar amounts of low-mass stars (Baumgardt & Makino 2003; Trenti et al. 2010). Ebrahimi et al. (2022) have recently reinforced this result: they performed detailed synthetic CMD fitting (à la Sollima et al. 2012) for a sample of 15 nearby OCs and found a significant correlation between the present-day mass-function slope and the ratio of age to half-mass relaxation time.

In the absence of homogeneous estimates of the remaining mass fraction or the present-day mass function (our sample covers a too large distance range for this to be feasible), the number of MS members per OC can give a very rough estimate of how close an OC is to dissolution. Again, however, this number depends on the OC distance, age, and especially the completeness of the underlying membership catalogue, which can even vary from cluster to cluster (Cantat-Gaudin et al. 2023). In Fig. 12, we highlight OCs whose MSs extend to masses lower than  $0.6 M_{\odot}$  and show their total multiplicity fractions as a function of the number of main-sequence members. While it is difficult to detect a clear trend, we do find that the upper right part of the diagram is practically empty: for example, OCs with more than 200 mem-



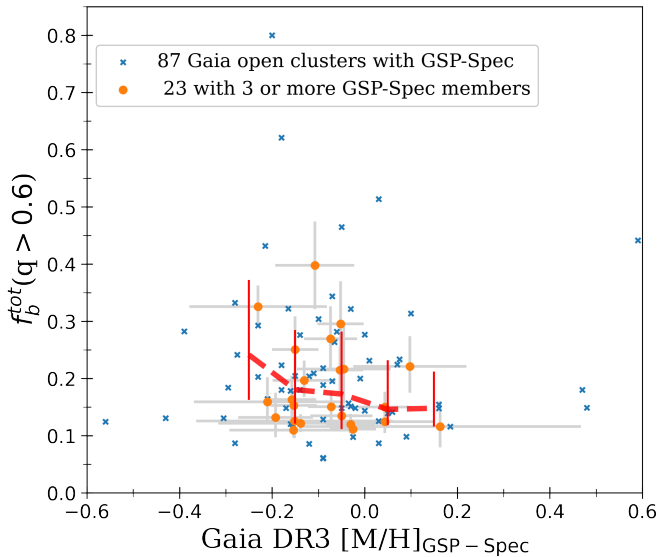
**Fig. 12.** Dependence of the total multiplicity fraction of systems with  $q > 0.6$  on the number of selected main-sequence members of each OC. OCs with extended low-mass main sequences ( $m_{\text{min}} < 0.6 M_{\odot}$ ) are highlighted as big orange circles with errorbars. The names of the two OCs with extended low-mass main sequences and high  $f_b^{\text{tot}}$  (and thus possible candidates for close-to-disruption OCs) are annotated.

bers all have  $f_b^{\text{tot}} < 0.5$  (or even 0.3 for the ones with  $m_{\text{min}} < 0.6 M_{\odot}$ ), while for the least populated OCs we find cases with much larger total multiplicity fractions. The most interesting cases are the OCs with  $m_{\text{min}} < 0.6 M_{\odot}$  and highly elevated multiplicity fractions, which could be candidates for being in their very final stage of evolution (UPK 385:  $d = 321$  pc, age = 12 Myr, see Fig. 4; UPK 445:  $d = 663$  pc, age = 13 Myr; both from CGa20). A low initial velocity dispersion or a low initial central density in those clusters, or fewer interactions with the environment, could also lead to the observed high binary fractions (Sollima 2008).

### 6.6. Dependence of the multiplicity fraction on metallicity

Already the star-formation simulations by Machida (2008) and Machida et al. (2009) suggested that low-metallicity gas clouds have a larger probability of fragmentation and thus a higher multiplicity frequency than metal-rich ones. Recent field-star analyses confirmed that the multiplicity fraction depends on the metallicity of the stellar system, at least for close binaries (Badenes et al. 2018; El-Badry & Rix 2019; Moe et al. 2019). For example, Badenes et al. (2018) found, using APOGEE multi-epoch spectroscopy (Majewski et al. 2017), that metal-poor ( $[\text{Fe}/\text{H}] \leq -0.5$ ) stars have multiplicity fractions 2 – 3 times higher than metal-rich ( $[\text{Fe}/\text{H}] \geq 0.0$ ) stars. The multiplicity fraction may further depend on the  $[\alpha/\text{Fe}]$  abundance ratio (Mazzola et al. 2020).

Here we try to test these findings using the GSP-Spec metallicities (Recio-Blanco et al. 2022) derived from the *Gaia* DR3 data (Gaia Collaboration et al. 2022c) for our parent sample, the OC members of T22 and CGa20. We follow the recommendations of Gaia Collaboration et al. (2022b) and select a "high-quality" subset (see their Appendix B) with an additional effective temperature cut of  $T_{\text{eff}} < 8000$  K. For 87 OCs in our sample GSP-Spec metallicities are available, but only for 23 of them we have metallicities for at least three member stars.



**Fig. 13.** Dependence of the open cluster’s total multiplicity fraction of systems with  $q > 0.6$  on metallicity (taken from Recio-Blanco et al. 2022). The 23 OCs with more than three individual GSP-Spec  $[M/H]$  measurements in *Gaia* DR3 are highlighted as big orange circles with errorbars. The dashed red line is the running median for the 87 OCs in our sample for which GSP-Spec metallicities are available, and the errorbars are calculated as the 16<sup>th</sup> and 84<sup>th</sup> percentiles.

Figure 13 shows the result of this exercise. While we cannot clearly confirm that the total high-mass-ratio multiplicity fraction strongly depends on metallicity, due to the limited range in  $[M/H]$  sampled by our local OC sample, we see on average larger  $f_b^{\text{tot}}$  at sub-solar metallicities compared to super-solar metallicities, in agreement with the field-star literature (Badenes et al. 2018; El-Badry & Rix 2019; Moe et al. 2019), confirming that our study is mostly sensitive to close binaries. In addition, our results indicate that the spread in  $f_b^{\text{tot}}$  is larger at  $[M/H] < 0$  than at  $[M/H] > 0$ , although in the super-solar regime our number statistics is low.

## 7. Conclusions

The study of the multiplicity fraction  $f_b$  in OCs requires membership determination and binary identification. Both have been made possible thanks to the unprecedented amount of high precision data for parallaxes, proper motions, and photometry provided by the successive data releases of the *Gaia* mission. In this paper we have successfully studied 202 *Gaia*-confirmed OCs closer than 1.5 kpc to the Sun, 146 from Tarricq et al. (2022) and 56 from Cantat-Gaudin et al. (2020) catalogues, retaining only their members having a membership probability larger than 0.7. Therefore, our conclusions are constrained by the membership completeness and selection biases in these works, as well as the incompleteness induced by the fit criteria of the present study.

The study of  $f_b$  in clusters is usually restricted to the unresolved binaries. Their identification can be done individually through photometry (as done in this work and most of the above cited papers, e.g. Niu et al. 2020; Borodina et al. 2021; Li & Shao 2022), but also through astrometry (Belokurov et al. 2020; Penoyre et al. 2022; Gaia Collaboration et al. 2022a) and radial-velocity monitoring (Kobulnicky et al. 2014; Bodensteiner et al.

2021; Yalyalieva et al. 2022). Given the number of stars in each OC and the velocity precision necessary to detect most binary systems, spectroscopic surveys typically take many years to complete. Such individual studies are not feasible, therefore, to be carried out for large OC samples.

This is why the CMD, only requiring imaging in two filters, is still most often used for OCs as a fundamental diagnostic tool. A simple classification is usually done dividing the CMD into two regions, one associated with the SS locus (including single stars and low- $q$  systems) and the other with the equal-mass BS (including the higher- $q$  systems), as done in de Grijs et al. (2013), Cordoni et al. (2018), or Jadhav et al. (2021). The degeneracy of metallicity and reddening, however, affects the isochrone fitting precision. An alternative is to create simulated CMDs using isochrones and binary prescriptions, and find the best fit (giving the nominal parameters) minimizing the distances between synthetic and observed data, as done in Sollima et al. (2010); Perren et al. (2015); Niu et al. (2020) or Li & Shao (2022).

Our approach, alternatively, models the observed  $G$  vs.  $(BP - RP)$  CMD as a mixture distribution of single stars and unresolved binaries, considering that two Gaussian distributions centred on the SS and BS loci introduce the observed scatter in  $G$  magnitude; one of them mostly accounting for the simple and low- $q$  systems, and the other for the high- $q$  binaries (and potentially higher-order systems). We are not concerned with estimating all the OC’s fundamental parameters, only the unresolved  $f_b$ , which is estimated as the weight of the BS’s Gaussian component; without need for estimating the  $q$  of each unresolved binary system. However, we do need to characterise the  $q$  range the BS’s Gaussian (and, thus,  $f_b$ ) takes into account, which by construction involves the higher  $q$  values down to  $q_{\text{lim}} = 0.6^{+0.05}_{-0.15}$ . We estimated this value by applying the method to a representative sample of simulated CMDs.

Applying strict criteria for a cluster to be well-fit, our MCMC implementation of the proposed mixture model yields good results for 202 (54%) of the 377 nearby OCs with at least 30 MS members and a MS that extends over at least 1 mag in  $BP - RP$ , confirmed by visual inspection. It yields values of  $f_b(q > 0.6)$  in the range  $[0.05, 0.67]$  with a median nominal uncertainty of 0.04 (not taking into account systematics). We also provide estimates of the total high-mass-ratio multiplicity fraction,  $f_b^{\text{tot}}(q > 0.6)$ , which takes into account the usually neglected portion of resolved binaries.

The main advantage of our modelling is that it does not rely on theoretical isochrone fitting. The versatility of the polynomial function fitted through MCMC enables us to study OCs with varied characteristics and degrees of differential extinction, thus enabling a homogeneous study of a large sample. The use of realistic *Gaia*-like simulations allows us to estimate both  $q_{\text{lim}}$  and the missing contribution from resolved binaries.

The main drawbacks are: 1. All the OCs’ selected MS members are regarded as either simple or binary systems, not considering the possible presence of field star contamination or higher-order multiple systems; 2. The uncertainties intrinsic to the modelling are sometimes difficult to quantify. The uncertainty percentiles of the estimated  $q_{\text{lim}} = 0.6^{+0.05}_{-0.15}$  could be considered too large for some purposes; 3. Our method only provides the multiplicity fraction integrated over the main-sequence mass range per OC, rendering a direct comparison to other studies difficult and resulting in complex selection effects (discussed in Sect. 6.2). For example, our unresolved  $f_b$  estimations are coherent with the values obtained by Jadhav et al. (2021), but much harder to be compared to other studies, which do not specify the con-

sidered mass and/or  $q$  range (see Sect. 6.1). Overall, the  $f_b$  estimation is by no means trivial, and it depends inevitably to some degree on modelling.

Summarising our main results from Sect. 6, we find:

1. The total high-mass-ratio multiplicity fraction taking into account also resolved binaries,  $f_b^{\text{tot}}(q > 0.6)$ , covers values from 6% to 80%, approximately following a log-normal distribution with a peak around 14% and a median of 18%. Only 9% of the OCs in the sample have an  $f_b^{\text{tot}}(q > 0.6) > 0.35$ .
2.  $f_b^{\text{tot}}$  increases with the mass of the primary star, in agreement with observational evidence from field stars and OCs (Bouvier et al. 1997; Deacon & Kraus 2020).
3. No apparent correlation appears between  $f_b^{\text{tot}}$  and the position in or perpendicular to the Galactic plane.
4. We observe a great dispersion for  $f_b^{\text{tot}}$  at all ages. There is, however, a decreasing trend with age until approximately 100 Myr, followed by a slight increase for older clusters.
5. Our custom simulations using GOG show that part of the trend seen with age is caused by the complex selection effects (introduced by the mass dependence of the multiplicity fraction and the magnitude limit of our sample).
6. The dispersion in  $f_b^{\text{tot}}$  seems to decrease significantly with the number of members (used as a proxy for cluster mass). We suggest that the highest values of  $f_b^{\text{tot}}$  (for UPK 385 and UPK 445) may indicate that those objects are very close to dissolution (in line with simulations by e.g. Hurley et al. 2007). Other possibilities such as a low initial velocity dispersion in those clusters, exist (Sollima 2008).
7. The multiplicity fraction decreases with metallicity, in line with recent studies using close binaries in the field (El-Badry & Rix 2019; Moe et al. 2019).
8. The data and code used to produce this paper are available at [https://github.com/fjaellet/oc\\_binary\\_fraction](https://github.com/fjaellet/oc_binary_fraction)

The study of multiplicity is a vibrant field, especially in OCs and especially since the advent of *Gaia*. The recent review by Offner et al. (2022) nicely illustrates how multiplicity depends on a number of physical parameters, and that the determination of the overall multiplicity fraction is entangled with the determination of separation/period, eccentricity, and mass-ratio distributions, as well as with the chemical properties of stars. Future work is therefore necessary to corroborate our results with revised and more complete membership lists using the latest *Gaia* data (including also radial velocities when possible; *Gaia* Collaboration et al. 2022a), and to extend it to the determination of the mass-ratio distributions (ideally as a function of primary mass and not averaged over a whole OC).

For nearby well-populated OCs, past works have already shown that it is possible to constrain the mass-ratio distribution (e.g. Li & Shao 2022). The recent study of Albrow & Ulusele (2022), for example, shows that in M67 the mass-ratio distribution rises for  $q > 0.3$ , in line with an earlier study for NGC 188 (Cohen et al. 2020), but slightly at odds with the close-to-flat  $q$  distribution found by Torres et al. (2021) for the Pleiades.

Another remaining question is whether the multiplicity fractions in OCs are measurably different from the multiplicity fractions in the field (except from the well-known effects of dynamical mass segregation; e.g. Geller & Mathieu 2012; Sheikhi et al. 2016; Tarricq et al. 2022; Casamiquela et al. 2022; Evans & Oh 2022). Parker & Goodwin (2012) showed that stochastic interactions within an OC can significantly vary the intermediate-separation multiplicity fraction by a factor of 2. Parker & Reggiani (2013) concluded, however, that the shape of the observed

$q$  distribution is not significantly altered by internal interactions in (Orion-Nebula-type) clusters after its birth, but rather an imprint of the star formation process. How the process of multiple-star formation depends on the density (but also metallicity and magnetic fields) in the birth environment, remains an open question (Offner et al. 2022).

**Acknowledgements.** We thank Nigel Hambly (Edinburgh) who inspired the initial idea for this project by presenting a *Gaia* DR2 python tutorial at the Royal Astronomical Society's National Astronomy Meeting 2019 and Teresa Antoja (Barcelona) for suggesting to fit the main sequence and the multiplicity fraction simultaneously. We also thank Yoann Tarricq (Bordeaux) for making his OC membership catalogue available to us before publication. Finally, we thank the anonymous referee for constructive comments that helped improve the paper. This work has made use of data from the European Space Agency (ESA) mission *Gaia* ([www.cosmos.esa.int/gaia](http://www.cosmos.esa.int/gaia)), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, [www.cosmos.esa.int/web/gaia/dpac/consortium](http://www.cosmos.esa.int/web/gaia/dpac/consortium)). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. This work was partially funded by the Spanish MICIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe" by the "European Union" through grant RTI2018-095076-B-C21 and PID2021-122842OB-C21, and the Institute of Cosmos Sciences University of Barcelona (ICCUB, Unidad de Excelencia 'María de Maeztu') through grant CEX2019-000918-M. FA acknowledges financial support from MICIN (Spain) through the Juan de la Cierva-Incorporación programme under contract IJC2019-04862-I. MG acknowledges support from grants EUR2020-112157 and PID2021-125485NB-C22. The preparation of this work has made use of TOPCAT (Taylor 2005), NASA's Astrophysics Data System Bibliographic Services, as well as the open-source Python packages *astropy* (Astropy Collaboration et al. 2018), *NumPy* (van der Walt et al. 2011), *corner* (Foreman-Mackey et al. 2016), and *emcee* (Foreman-Mackey et al. 2013). The figures in this paper were produced with *matplotlib* (Hunter 2007).

## References

- Abt, H. A. & Levy, S. G. 1976, *ApJS*, 30, 273  
 Albrow, M. D. & Ulusele, I. H. 2022, *MNRAS*, 515, 730  
 Anders, F., Cantat-Gaudin, T., Quadrino-Lodosa, I., et al. 2021, *A&A*, 645, L2  
 Arenou, F. 2011, in *American Institute of Physics Conference Series*, Vol. 1346, International Workshop on Double and Multiple Stars: Dynamics, Physics, and Instrumentation, ed. J. A. Docobo, V. S. Tamazian, & Y. Y. Balega, 107–121  
 Aros, F. I., Sippel, A. C., Mastrobuono-Battisti, A., et al. 2021, *MNRAS*, 508, 4385  
 Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, 156, 123  
 Badenes, C., Mazzola, C., Thompson, T. A., et al. 2018, *ApJ*, 854, 147  
 Banerjee, S. 2022, *Phys. Rev. D*, 105, 023004  
 Baumgardt, H. & Makino, J. 2003, *MNRAS*, 340, 227  
 Belczynski, K., Kalogera, V., & Bulik, T. 2002, *ApJ*, 572, 407  
 Belloni, D., Askar, A., Giersz, M., Kroupa, P., & Rocha-Pinto, H. J. 2017, *MNRAS*, 471, 2812  
 Belokurov, V., Penoyre, Z., Oh, S., et al. 2020, *MNRAS*, 496, 1922  
 Bica, E. & Bonatto, C. 2005, *A&A*, 431, 943  
 Bodensteiner, J., Sana, H., Wang, C., et al. 2021, *A&A*, 652, A70  
 Borodina, O. I., Carraro, G., Seleznev, A. F., & Danilov, V. M. 2021, *ApJ*, 908, 60  
 Borodina, O. I., Seleznev, A. F., Carraro, G., & Danilov, V. M. 2019, *ApJ*, 874, 127  
 Bouvier, J., Rigaut, F., & Nadeau, D. 1997, *A&A*, 323, 139  
 Bressan, A., Marigo, P., Girardi, L., et al. 2012, *MNRAS*, 427, 127  
 Cantat-Gaudin, T., Anders, F., Castro-Ginard, A., et al. 2020, *A&A*, 640, A1  
 Cantat-Gaudin, T., Fouesneau, M., Rix, H.-W., et al. 2023, *A&A*, 669, A55  
 Casamiquela, L., Olivares, J., Tarricq, Y., et al. 2022, *A&A*, 664, A31  
 Cohen, R. E., Geller, A. M., & von Hippel, T. 2020, *AJ*, 159, 11  
 Cordoni, G., Milone, A. P., Marino, A. F., et al. 2018, *ApJ*, 869, 139  
 Dal Tio, P., Mazzi, A., Girardi, L., et al. 2021, *MNRAS*, 506, 5681  
 de Grijs, R., Li, C., Zheng, Y., et al. 2013, *ApJ*, 765, 4  
 Deacon, N. R. & Kraus, A. L. 2020, *MNRAS*, 496, 5176  
 Duchêne, G. & Kraus, A. 2013, *ARA&A*, 51, 269  
 Duquennoy, A. & Mayor, M. 1991, *A&A*, 248, 485  
 Ebrahimi, H., Sollima, A., & Haghi, H. 2022, *MNRAS*  
 El-Badry, K. & Rix, H.-W. 2019, *MNRAS*, 482, L139  
 Evans, N. W. & Oh, S. 2022, *MNRAS*, 512, 3846  
 Fabricius, C., Luri, X., Arenou, F., et al. 2021, *A&A*, 649, A5

- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306
- Foreman-Mackey, D., Voutsden, W., Price-Whelan, A., et al. 2016, *corner.py*: *corner.py* v1.0.2
- Fregeau, J. M., Cheung, P., Portegies Zwart, S. F., & Rasio, F. A. 2004, *MNRAS*, 352, 1
- Fuhrmann, K., Chini, R., Kaderhandt, L., & Chen, Z. 2017, *MNRAS*, 464, 2610
- Gaia Collaboration, Arenou, F., Babusiaux, C., et al. 2022a, *A&A*, in press, arXiv:2206.05595
- Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, *A&A*, 616, A1
- Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2021, *A&A*, 649, A1
- Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, *A&A*, 595, A1
- Gaia Collaboration, Recio-Blanco, A., Kordopatis, G., et al. 2022b, *A&A*, in press, arXiv:2206.05534
- Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2022c, *A&A*, in press, arXiv:2208.00211
- Geller, A. M. & Mathieu, R. D. 2012, *AJ*, 144, 54
- González, E., Kremer, K., Chatterjee, S., et al. 2021, *ApJ*, 908, L29
- Hogg, D. W., Bovy, J., & Lang, D. 2010, arXiv e-prints, arXiv:1008.4686
- Hunter, J. D. 2007, *Computing in Science and Engineering*, 9, 90
- Hurley, J. & Tout, C. A. 1998, *MNRAS*, 300, 977
- Hurley, J. R., Aarseth, S. J., & Shara, M. M. 2007, *ApJ*, 665, 707
- Jadhav, V. V., Roy, K., Joshi, N., & Subramaniam, A. 2021, *AJ*, 162, 264
- Kaczmarek, T., Olczak, C., & Pfalzner, S. 2011, *A&A*, 528, A144
- Kobulnicky, H. A., Kiminki, D. C., Lundquist, M. J., et al. 2014, *ApJS*, 213, 34
- Kroupa, P. 2001, *MNRAS*, 322, 231
- Kroupa, P. 2002, *Science*, 295, 82
- Lada, C. J. & Lada, E. A. 2003, *ARA&A*, 41, 57
- Lamers, H. J. G. L. M., Gieles, M., Bastian, N., et al. 2005, *A&A*, 441, 117
- Li, L. & Shao, Z. 2022, *ApJ*, 930, 44
- Li, L., Shao, Z., Li, Z.-Z., et al. 2020, *ApJ*, 901, 49
- Li, Z., Deng, Y., Chi, H., et al. 2022, *ApJS*, 259, 19
- Luri, X., Palmer, M., Arenou, F., et al. 2014, *A&A*, 566, A119
- Machida, M. N. 2008, *ApJ*, 682, L1
- Machida, M. N., Omukai, K., Matsumoto, T., & Inutsuka, S.-I. 2009, *MNRAS*, 399, 1255
- Maeder, A. 1974, *A&A*, 32, 177
- Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, *AJ*, 154, 94
- Marigo, P., Girardi, L., Bressan, A., et al. 2017, *ApJ*, 835, 77
- Mazzola, C. N., Badenes, C., Moe, M., et al. 2020, *MNRAS*, 499, 1607
- Mermilliod, J. C. & Mayor, M. 1989, *A&A*, 219, 125
- Milone, A. P., Piotto, G., Bedin, L. R., et al. 2012, *A&A*, 540, A16
- Moe, M., Kratter, K. M., & Badenes, C. 2019, *ApJ*, 875, 61
- Moe, M. & Stefano, R. D. 2017, *The Astrophysical Journal Supplement Series*, 230, 15
- Niu, H., Wang, J., & Fu, J. 2020, *ApJ*, 903, 93
- Offner, S. S. R., Moe, M., Kratter, K. M., et al. 2022, *Proceedings of Protostars and Planets VII*, in press, arXiv:2203.10066
- Parker, R. J. & Goodwin, S. P. 2012, *MNRAS*, 424, 272
- Parker, R. J. & Reggiani, M. M. 2013, *MNRAS*, 432, 2378
- Penoyre, Z., Belokurov, V., & Evans, N. W. 2022, *MNRAS*, 513, 5270
- Perren, G. I., Vázquez, R. A., & Piatti, A. E. 2015, *A&A*, 576, A6
- Pinfield, D. J., Dobbie, P. D., Jameson, R. F., et al. 2003, *MNRAS*, 342, 1241
- Rastello, S., Carraro, G., & Capuzzo-Dolcetta, R. 2020, *ApJ*, 896, 152
- Recio-Blanco, A., de Laverny, P., Palicio, P. A., et al. 2022, *A&A*, in press, arXiv:2206.05541
- Reid, M. J., Menten, K. M., Brunthaler, A., et al. 2019, *ApJ*, 885, 131
- Riello, M., De Angeli, F., Evans, D. W., et al. 2021, *A&A*, 649, A3
- Robin, A. C., Luri, X., Reylé, C., et al. 2012, *A&A*, 543, A100
- Sana, H., de Mink, S. E., de Koter, A., et al. 2012, *Science*, 337, 444
- Sandhu, T. S., Pandey, A. K., & Sagar, R. 2003, *A&A*, 408, 515
- Sharma, S., Pandey, A. K., Ogura, K., et al. 2008, *AJ*, 135, 1934
- Sheikhi, N., Hasheminia, M., Khalaj, P., et al. 2016, *MNRAS*, 457, 1028
- Sollima, A. 2008, *MNRAS*, 388, 307
- Sollima, A., Bellazzini, M., & Lee, J. W. 2012, *ApJ*, 755, 156
- Sollima, A., Carballo-Bello, J. A., Beccari, G., et al. 2010, *MNRAS*, 401, 577
- Sterzik, M. F. & Durisen, R. H. 2004, in *Revista Mexicana de Astronomía y Astrofísica Conference Series*, Vol. 21, *Revista Mexicana de Astronomía y Astrofísica Conference Series*, ed. C. Allen & C. Scarfe, 58–62
- Tarricq, Y., Soubiran, C., Casamiquela, L., et al. 2022, *A&A*, 659, A59
- Taylor, M. B. 2005, *Astronomical Society of the Pacific Conference Series*, Vol. 347, TOPCAT & STIL: Starlink Table/VOTable Processing Software, ed. P. Shopbell, M. Britton, & R. Ebert, 29
- Thies, I. & Kroupa, P. 2007, *ApJ*, 671, 767
- Thompson, B. A., Frinchaboy, P. M., Spoo, T., & Donor, J. 2021, *AJ*, 161, 160
- Tokovinin, A. & Moe, M. 2020, *MNRAS*, 491, 5158
- Torres, G., Latham, D. W., & Quinn, S. N. 2021, *ApJ*, 921, 117
- Trenti, M., Vesperini, E., & Pasquato, M. 2010, *ApJ*, 708, 1598
- van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *Computing in Science and Engineering*, 13, 22
- Wielen, R. 1971, *Ap&SS*, 13, 300
- Yalyalieva, L., Carraro, G., Glushkova, E., Munari, U., & Ochner, P. 2022, *MNRAS*, 513, 5299
- Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, *Research in Astronomy and Astrophysics*, 12, 723