UNIVERSITAT DE BARCELONA

MASTER'S THESIS

The multiplicity fraction in 324 open clusters from *Gaia*

Author:
Judit DONADA OLIU

Supervisors:
Dr. Friedrich ANDERS
Dr. Carme JORDI NEBOT

A thesis submitted in fulfillment of the requirements
for the Master's degree in Astrophysics, Particle Physics and Cosmology

June, 2022

UNIVERSITAT DE BARCELONA

IEEC
Institut d'Estudis
Espacials de Catalunya

UNIVERSITAT DE BARCELONA

Abstract

Master's degree in Astrophysics, Particle Physics and Cosmology

The multiplicity fraction in 324 open clusters from *Gaia*

by Judit DONADA OLIU

This Master's thesis is devoted to the estimation of the multiplicity fraction of high-mass ratio main-sequence multiple systems in Galactic open clusters. The main study consists in estimating in an automated fashion the unresolved multiplicity fraction through the location of the unresolved multiple systems in colour-magnitude diagrams, using *Gaia* DR2 and EDR3 data. Regarding this study, this thesis is a continuation of the previous preliminary work of my bachelor thesis (Donada, 2021), which explored the possibility of applying a Gaussian mixture model algorithm. Now we revise and improve this method, and also develop an alternative implementation using a Markov chain Monte Carlo method. The unresolved multiplicity fractions obtained through both methods are compared and characterized using custom realistic *Gaia*-like open cluster simulations performed with the *Gaia Object Generator* (Luri et al., 2014). Using these simulations we estimate the effective limiting mass ratio above which the Markov chain Monte Carlo method is able to detect the presence of a secondary companion (i.e., the high mass ratio range which our estimated unresolved multiplicity fraction comprises).

The simulations further enable us to correct for the presence of resolved multiple systems. So, as a second part of the study, we estimate the total high mass ratio multiplicity fraction of the open clusters' main sequences.

Finally, we compare our results to the ones estimated through `ASteCA` (Automated Stellar Cluster Analysis package).

The main result of this work is the largest homogeneous catalogue of multiplicity fractions in open clusters to date, including the unresolved and total multiplicity fractions of main-sequence systems with mass ratio larger than $0.6^{+0.05}_{-0.15}$ for 324 open clusters, estimated through the Markov chain Monte Carlo method. All studied open clusters are closer than 1.5 kpc, and have ages between 3.9 Myr and 4.3 Gyr. Their total multiplicity fractions, between 0.05 and 0.71, are found to increase with the mass of the primary star, display an overall decreasing trend with the open cluster age up until ages about 100-320 Myr, above which the trend increases; and do not depend on the open cluster position in the Galaxy.

Acknowledgements

I would first like to thank my advisors, Dr. Friedrich Anders and Dr. Carme Jordi Nebot, for their valuable contributions, guidance and dedication throughout all the development of this thesis.

I would also like to express my gratitude to Dr. Eduard Masana Fresno, for his assistance in the simulation aspect of the study. I am also very grateful to Dr. Teresa Antoja Castelltort for her helpful insight into the implementation of our model, as well as to all the rest of members of the *Gaia* team at the University of Barcelona.

Finally, I also wish to thank Dr. Gabriel Ignacio Perren for guiding me through the application of `ASteCA` code to our research.

Contents

Chapter 1

Introduction

Most of the stars are formed as part of a binary/multi-component system (Moe and Stefano, 2017). Binary stars lead to the formation of supernovae, blue straggler stars, chemically peculiar stars, etc.; and the characterization of the multiplicity fraction (f_b) in open clusters (OCs) and field stars is relevant to several branches of astrophysics (for reviews see e.g. Duchêne and Kraus, 2013; Offner et al., 2022). OCs typically host more than 30% of binaries (Borodina et al., 2019) and are highly suitable for studying them, because their members have, to first order, the same distance, age, initial chemical composition, and foreground extinction. Most of their binaries are unresolvable in images, and such photometric unresolved binaries (binaries hereafter) generally make up a significant fraction of OC photometric samples. Therefore, they must be taken into account to accurately estimate OC parameters, for instance their total mass and stellar mass function (e.g. Kroupa, 2001; Borodina et al., 2019; Rastello, Carraro, and Capuzzo-Dolcetta, 2020). They also influence the OC's dynamical evolution, causing the evaporation of single and low-mass systems and central segregation of massive stellar systems. Hence, f_b is likely to vary with respect to the primordial value through dynamical effects. Binaries also provide information of the outcome of star-forming processes in different environments and constraints on the initial OC state (e.g. Gieles, Sana, and Portegies Zwart, 2010; Li et al., 2020), and are responsible for high rates of stellar collisions and black-hole mergers (González et al., 2021; Aros et al., 2021; Banerjee, 2022).

OCs display significantly varying f_b, which both at primordial and late times are not very well constrained. Since typically Galactic OCs are disrupted on a time scale of a few hundred Myrs (Lamers et al., 2005), they tend to be relatively young. For young massive clusters in the local universe, the measured f_b are close to those of field stars (González et al., 2021).

For a binary system of total mass $M = M_1 + M_2$, the mass ratio is defined as $q = M_2/M_1$, 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 $0 \leq q \leq 1$. An unresolved binary system composed of two identical main-sequence (MS) stars ($q = 1$) has the same colour but twice the luminosity of an equivalent single star, and appears in the OC's colour-magnitude diagram (CMD) displaced vertically upwards by 0.753 mag with respect to the equivalent single star location irrespective of the wavelength bands used. The locus defined by such systems is called the equal-mass binary sequence. A system with two unequal MS 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 the single star but less than the one corresponding to the equal-mass binary system. Therefore, this system is displaced in the CMD both upwards (but not reaching 0.753 mag) and to the right relative to the MS position of the primary component. For q increasing from 0 to 1, the position of the binary system with respect to the MS position of the primary component always increases in magnitude, but for $0 < q < q_{\rm crit}$ moves towards redder colours, and for $q_{\rm crit} < q < 1$ towards the blue again (Maeder, 1974). The value of $q_{\rm crit}$ depends on the

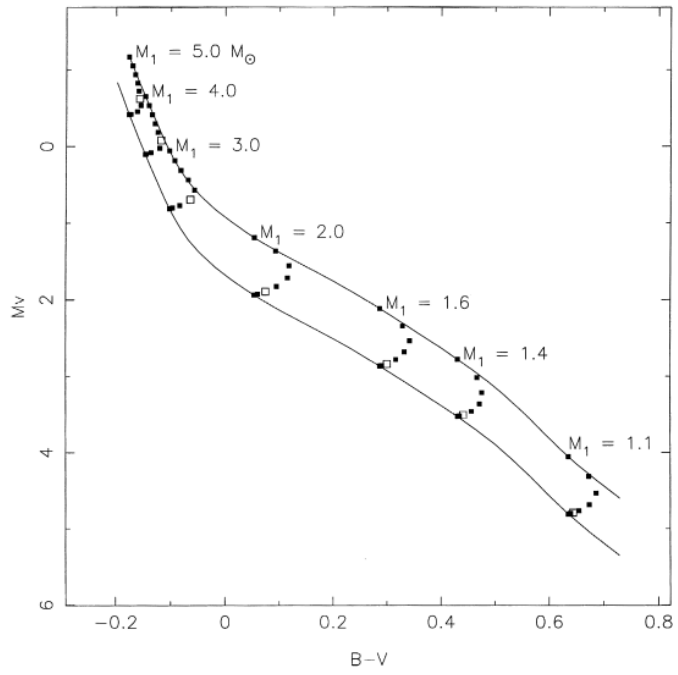


FIGURE 1: Theoretical zero-age single-star and equal-mass binary main sequences for a metallicity of $Z = 0.02$, represented in a M_V vs. $(B - V)$ colour–magnitude diagram (from Hurley and Tout, 1998). For a range of primary masses (M_1) are plotted the locus of binary points where q ranges from 1.0 to 0.0 in increments of 0.1. The point of $q = 0.5$ is an open square.

mass of the primary component, as can be seen in Fig. 1, what is caused by the fact that the shifts in magnitude and colours of a system of $q < 1$ depend on the magnitude of the primary star. 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 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. The larger M_1, the larger the high-q range for which binaries lie on the BS. And the lower M_1, the larger the low-q range for which the faint secondary makes almost no contribution and the binary system lies almost at the same position as the equivalent primary mass (see Fig. 1).

$q_{\rm crit}$ does not exactly coincide with the smallest q for which a binary system is seen to lie on the BS: $q_{\rm lim}$ (which depends on M_1 too). From the observational perspective, $q_{\rm lim}$ is the mass ratio above which an unresolved binary system can be detected as such (from photometry alone). In this study we attempt to measure the high mass-ratio ($q > q_{\rm lim}$) binary fraction for an unprecedented number of Galactic OCs in an automated fashion, making use of the exquisite *Gaia* photometry (Evans et al., 2018 for DR2 and Riello et al., 2021 for EDR3).

This work is structured as follows. The exact treatment of the data and our proposed model are described in Chapter 2. In Chapter 3 we explain the Gaussian Mixture Model (GMM) method for the unresolved f_b estimation, and in Chapter 4 the Markov Chain Monte Carlo (MCMC) implementation, and both are applied to *Gaia* data. In Chapter 5 we apply again the two previous methods but to simulated CMDs generated using the *Gaia Object Generator* (GOG), and we estimate the $q_{\rm lim}$ of our determined unresolved $f_b(q > q_{\rm lim})$ and compute the total $f_b^{tot}(q > q_{\rm lim})$ (of both resolved and unresolved systems) for the OCs in our sample. The results obtained for the f_b estimation and our catalogue are described in Chapter 6. In Chapter 7 our found f_b values are compared to the literature, and their dependence on mass, position and age is discussed. The application of the `ASteCA` code is summarized in the Appendix, and we highlight the main conclusions in Chapter 8.

Chapter 2

Data and modelling

For this study we use the recent *Gaia* OC catalogues published by Tarricq et al. (2022, hereafter T22) and Cantat-Gaudin et al. (2020, hereafter CGa20).

The T22 catalogue contains precise *Gaia* EDR3-derived memberships for 389 OCs located closer than 1.5 kpc from the Sun and older than 50 Myr. They are improved with respect to DR2-derived ones, extending them to OC's peripheral areas.

In CGa20, the OC's age, distance modulus, and extinction were computed from the observed *Gaia* DR2 parallaxes and G vs. $(BP - RP)$ CMDs by a multi-layer-perceptron neural network trained on a set of 347 OCs with well-determined parameters. The cluster membership lists were mostly taken from Cantat-Gaudin and Anders (2020) and Castro-Ginard et al. (2020).

In the first section of this Chapter we explain the selection of the studied OCs and their members. In the second one we describe the procedure applied to further select only those members in the MS of each OC and apply additional constrains on the studied OCs. In the last section we describe our model to estimate their unresolved f_b in *Gaia* 's G vs. $(BP - RP)$ CMD.

2.1 Open cluster memberships and parameters

Since we are interested in an OC census that is representative for the extended solar neighbourhood, we select all local ($d < 1.5$ kpc) OCs from T22. For OCs younger than 50 Myr (for which Tarricq et al. 2022 do not provide improved membership lists), we supplement this dataset selecting those OCs of $d < 1.5$ kpc from CGa20. In both cases, we use the homogeneously derived astrophysical parameters (age, distance, extinction) published in CGa20 (possible because all OCs studied in T22 are also present in CGa20). Figure 2 shows the spatial and age distribution of our studied final sample of 324 OCs (as explained in Chapter 4), for which we provide a catalogue of their f_b.

Both membership catalogues are limited to $G < 18$. Regarding the considered members of each OC, we have used all the members in CGa20 catalogue, which have membership probabilities $\geq 70\%$, and have selected those of T22 with membership probabilities above this same threshold value. However, as the criteria for the membership probability assignation are not equal, establishing this common threshold does not make the selection equivalent for both catalogues. Besides, the fact that CGa20 is based on DR2 astrometry and photometry instead of on EDR3 as T22, and that it comprehends a smaller spatial coverage, prevents our OC sample of being purely homogeneous.

2.2 Selection of open cluster's main-sequence members

As our f_b determination is restricted to only MS systems, we adopted a homogeneous selection of the MS members of each OC in our sample. First of all, the absolute magnitude M_G and intrinsic colour index $(BP - RP)_0$ were calculated for each

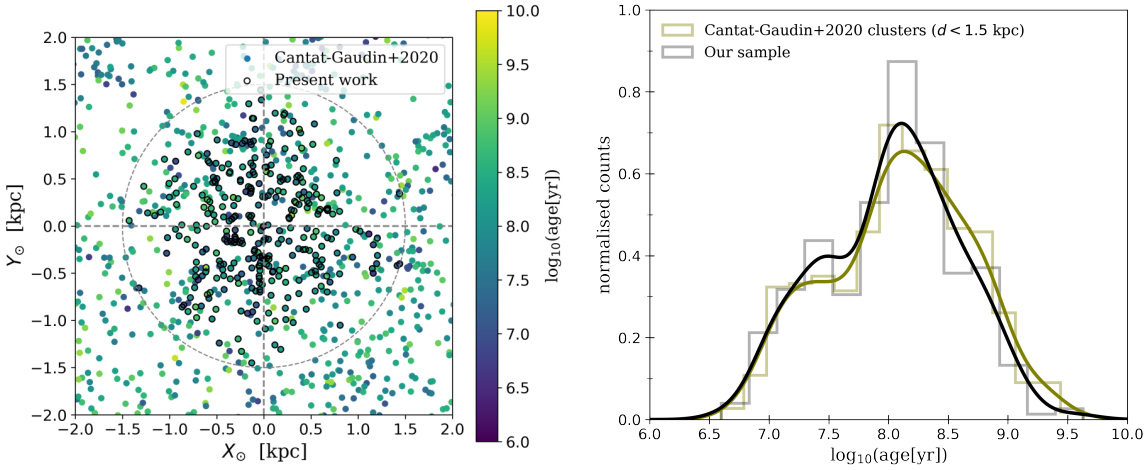


FIGURE 2: *Left*: Spatial distribution of the open clusters in CGa20 parent sample (points colour-coded by age) and in our final sample of 324 OCs (coloured points surrounded by a black contour) in heliocentric Cartesian coordinates (the Galactic Centre is towards positive $X_\odot$). The dashed grey circumference delimitates our sample's distance limit of 1.5 kpc. *Right*: Age distribution of the OCs closer than 1.5 kpc in CGa20 parent sample (olive histogram) and in our final sample (grey histogram). Both histograms have areas that integrate to one, and their corresponding curves are kernel-estimated probability density functions.

cluster member (excluding the ones with missing BP and/or RP photometry):

$$M_G = G - \mathrm{DM} - 0.89 \cdot \mathrm{A_V}, \tag{1}$$

$$(BP - RP)_0 = BP - RP - \frac{0.89}{1.85} \cdot A_V, \tag{2}$$

where G is the *Gaia* G-band mean magnitude, DM is the distance modulus, and A_V is the OC's visual extinction (corrected by a factor in equation (1) to obtain the extinction in the G-band, and by a different factor in equation (2) to obtain the $E_{(BP-RP)}$ colour excess). With these two quantities, we can 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 in Fig. 3.

First, two straight parallel lines which lie above and below the MSs of all clusters are drawn in the CMD (black lines in Fig. 3). Using them we define a first rough selection that excludes red-clump (RC) stars, white dwarfs, and some extreme outliers:

$$2.9 \cdot (BP - RP)_0 - 1.4 < M_G < 2.9 \cdot (BP - RP)_0 + 3.4 \tag{3}$$

Another restriction is required to exclude members already evolving towards giant stars. In order to estimate the intrinsic colour index of the MS turn-off (MSTO), we fitted an interpolation polynomial $f(\log_{10}(\mathrm{age[yr]}))$ to the $(BP - RP)_0$ dependence on the age for the bluest member of the MSs of 15 different solar-metallicity PARSEC 1.2S isochrones (Bressan et al., 2012) with ages $\log_{10}(\mathrm{age[yr]}) \in [6.6, 9.8]$. So, for each OC, the dereddened colour of the MSTO was obtained as a function of its age as $(BP - RP)_0^{TO} = f(\log_{10}(\mathrm{age[yr]}))$.

Then, MSTO stars were excluded by only considering members at least 0.2 mag redder than the MSTO (i.e., members with $M < M_{TO}$ were selected):

$$(BP - RP)_0 > (BP - RP)_0^{TO} + 0.2 \tag{4}$$

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:

$$M_G \geq M_G((BP - RP)_{0,\text{bluest member}}) - 0.75 \tag{5}$$

Such conditions (retaining members towards the right of the OC's bluest vertical line and below the OC's horizontal line in Fig. 3) were generally found to succeed in excluding from the study not only the MSTO members, but also those extreme MS members for which the BS approaches the SS and eventually intersects it. For these cases our mixture-model algorithm 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 f_b overestimation due to the magnitude limit ($G < 18$) in the parent cluster membership catalogues. We therefore select only members more than 0.2 mag bluer than the reddest one:

$$(BP - RP)_0 \leq (BP - RP)_{0,\text{reddest member}} - 0.2. \tag{6}$$

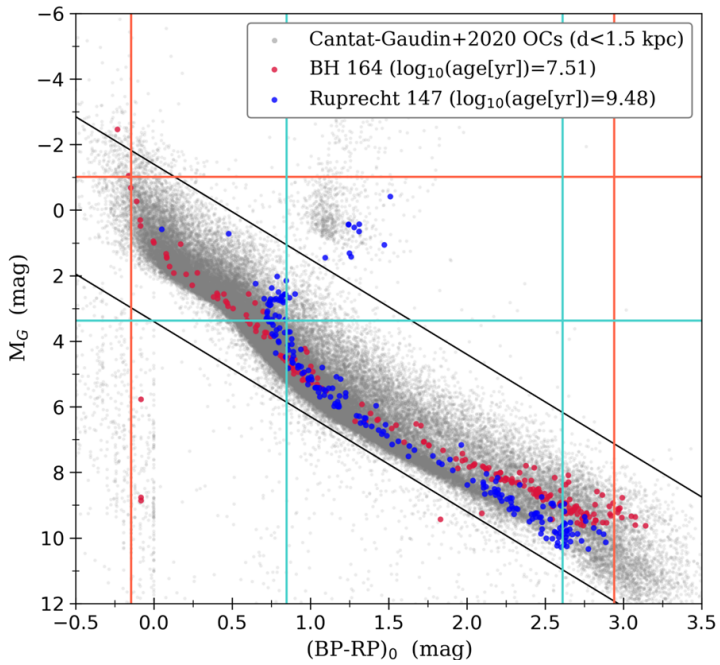


FIGURE 3: Example for two open clusters of different ages (BH 164 from CGa20 and Ruprecht 147 from T22) showing the selection of their main-sequence members from the dereddened *Gaia* CMD.

The subsequent study is carried out only for those OCs having a number of retained MS members larger or equal than 30 (as required in T22), and, at the same time, a MS extension in $(BP - RP)_0$ larger or equal than 1 mag (computed as the rest of the maximum and minimum $(BP - RP)_0$ values of the selected members). Through these conditions we disregarded a considerable proportion of the OCs for

which our model cannot be properly applied, because their CMDs are not enough populated or too sparse.

We now have a sample of 246 OCs with EDR3 memberships from T22 and 128 OCs with DR2 memberships from CGa20 for which the MS members have been selected.

2.3 Modelling for the unresolved binary fraction inference in the *Gaia* CMD

The multiplicity fraction f_b of a stellar population, hereafter referred to as the binary fraction, is defined as:

$$f_b = \frac{B + T + ...}{S + B + T + ...}, \tag{7}$$

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. Studying stellar multiplicity in a statistically robust manner is a highly non-trivial task (e.g. Duquennoy and Mayor 1991; Belloni et al. 2017). The mass ratio of binaries, for instance, is interconnected with the primary mass, the orbital period, the eccentricity, and the system metallicity.

We model the CMD as a mixture distribution of single stars and unresolved binaries, considering that the scatter of binaries between the $q = 0$ and $q = 1$ isochrones arises from two Gaussian distributions in G magnitude approximately centred on the SS and BS, one of them mostly accounting for single stars, resolved binaries and low-q unresolved binaries and the other for unresolved binaries of $q_{\rm lim} < q \leq 1$. Assuming these distributions to be Gaussians is a first-order approach, not physically motivated. Then, our estimated unresolved f_b of high-q systems is just the weight of the binaries' Gaussian: $f_b(q > q_{\rm lim}) = \frac{w_{BS}}{w_{SS}+w_{BS}} = w_{BS}$ (if both weights are normalised: $w_{SS} + w_{BS} = 1$).

In fact, this is the integrated high-q unresolved f_b in the MSs of OCs, because two common Gaussian distributions of a fixed standard deviation are considered for all the MS systems. And, as the q range they comprise depends on M_1, the $q_{\rm lim}$ below which multiple systems are regarded as simple systems varies as a function of $BP - RP$. We, however, provide a single mean $q_{\rm lim}$ value for each cluster, 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, so the estimated $f_b(q > q_{\rm lim})$ includes their contribution.

In the two following chapters we carry out two different implementations of this modelling, using GMM and MCMC methods. Our model has the advantages that the fit can be performed directly in observable space (the observed *Gaia* CMD), and that the fitted function of the SS (and BS too in the case of MCMC) does not depend on stellar models at all.

Chapter 3

Gaussian Mixture Model method

Our implementation of the GMM algorithm, which gives a probabilistic 2-Gaussian model describing the one-dimensional distribution of the OC members' distance to the fitted SS polynomial, is described, as well as the application to empirical *Gaia* data.

3.1 Fitting the single-star main sequence

A fourth degree polynomial $p_{SS}((BP-RP)_0)$ was fitted to each OC's CMD of selected MS members through the least-square method with a Huber loss function (Huber, 1964). The value of soft margin between inlier and outlier residuals was set to 0.05, what enabled to fit the polynomial following the single-star MS locus (regarding multiple systems as outliers).

The study was only carried out further for those OCs whose SS fitted polynomial was monotonically decreasing over the entire CMD domain (e.g. left panel of Fig. 4).

3.2 Gaussian Mixture Model for unresolved binaries

The SS fitted polynomial was used to calculate the difference between the absolute magnitude of each OC member and the one corresponding to the single-star MS locus fitted polynomial for its same intrinsic colour index:

$$\Delta G = M_G - p_{SS}((BP-RP)_0), \tag{8}$$

and a ΔG histogram was represented for each OC. The final selection of the considered MS members consisted in removing the members having $\Delta G \notin [-1.5, 1.0]$, considered far enough from both the SS and the BS so as to be regarded as outliers. Still, all the studied OCs turned out to have more than 30 retained MS members.

Our method for the f_b estimation is based on assuming that all selected ΔG values of the histogram come from a mixture of two Gaussian distributions with unknown parameters: one chiefly associated with the SS and the other with the BS. Hence, the problem is reduced to a single dimension: regardless of the colour index of the member, what determines if it belongs to the SS or the BS is its distance from the single-star MS locus. Depending on the characteristic features of each OC's CMD, however, the resulting ΔG histogram can be attributed most likely to a mixture of a different number of Gaussians. We used the fit method of `GaussianMixture` option of `sklearn.mixture` package to find the best mixture of the desired number of Gaussians through an expectation-maximization approach. This GMM method provides the mean μ, standard deviation σ and weight w of the fitted Gaussians, as well as the Bayesian information criterion (BIC) yield by the fit. To ensure that our modelling was consistent with the experimental data, instead of blindly fitting a mixture of two Gaussians to each histogram, we compared the BIC yield by the

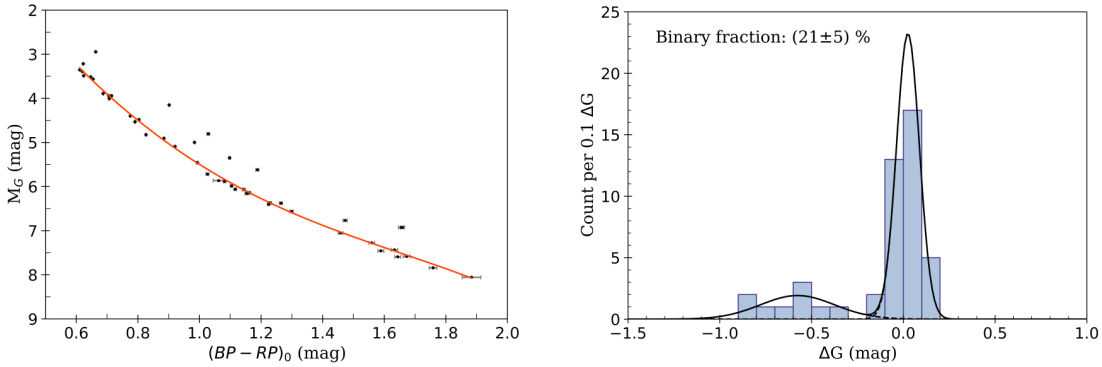


FIGURE 4: *Left*: Dereddened CMD of the 46 selected main-sequence members of Alessi 1 ($\log_{10}$(age[yr])=9.16). Its observational data is from T22 catalogue, and has hardly noticeable errorbars. The orange curve is the polynomial fitted to its single-star sequence. *Right*: Corresponding ΔG histogram with the singles' sequence and binary sequence Gaussians fitted through a two-component Gaussian Mixture Model, yielding an estimated unresolved binary fraction (weight of the binaries' Gaussian) of $21 \pm 5\%$.

fitting of one, two and three Gaussians and selected only those OCs for which the BIC value was minimum for the 2-component GMM (implying that a 2-component model is generally preferred).

The highest weight Gaussian (w_{SS}) corresponds to the one fitted tracing the SS, while the one with the smallest weight (w_{BS}) corresponds to the BS (less populated); and their weights are normalized.

3.3 Fit quality cuts

The last conditions that the studied OCs had to meet were imposed on the two Gaussians fitted through the GMM: σ_{SS}, $\sigma_{BS} < 0.3$ and $w_{SS} > 0.7$. They were derived through visual inspection to restrict the sample to those OCs that display a bimodal ΔG histogram where each of the fitted Gaussians contributes non-negligibly to a single well-distinguished ΔG peak (either the one associated with the SS or the BS one), so that $f_{b,\mathrm{GMM}}(q > q_{\mathrm{lim}}^{\mathrm{GMM}}) = w_{BS}$ and our model can be properly applied (e.g. right panel of Fig. 4).

The resulting final sample of OCs for which f_b is estimated through GMM has 84 OCs in total: 67 from T22 and 17 from CGa20.

3.4 Binary fraction estimation

As explained in Sect. 2.3, $f_{b,\mathrm{GMM}}(q > q_{\mathrm{lim}}^{\mathrm{GMM}}) = w_{BS}$. To estimate its uncertainty, we consider that the probability of any system out of the N selected MS members being a binary system follows a binomial distribution with a probability of success w_{BS}. Then, the Wald formula (Agresti and Coull, 1998) is applied:

$$\delta\left[f_{b,\mathrm{GMM}}\right] = z\sqrt{\frac{w_{BS}(1-w_{BS})}{N}}, \tag{9}$$

with $z = 0.99$ (given for a 68% confidence level). This uncertainty is underestimated, as it only involves a statistical estimation of the error derived from the Gaussian fitted to the binaries. It does not take into account intrinsic photometry errors or uncertainties associated with the method applied to fit the Gaussians (regarding the accuracy of the MS members selection and SS polynomial fitting).

We estimate $q_{\rm lim}^{\rm GMM,sim}$ of the simulated CMDs (Sect. 5.2.2), but in Sect. 6.1 we discuss why it cannot be equalled to $q_{\rm lim}^{\rm GMM,obs}$ of the observed CMD.

Chapter 4

Markov Chain Monte Carlo method

Following the path laid out in Hogg, Bovy, and Lang (2010), we can write down explicitly a mixture model for the distribution of cluster members in the CMD. Assuming negligible contamination by field star interlopers, the likelihood for the CMD distribution below the MSTO can be described by an arbitrary polynomial $P(BP-RP)$ with a small intrinsic scatter for the SS, and a parallel one (perhaps with a slightly larger intrinsic scatter) for the BS. So, if the functional form of the single-star MS in the *Gaia* CMD is perfectly known (i.e. we know the functional form $G_{SS} = f(BP-RP)_{SS}$ + scatter), the cluster CMD (at least the part sufficiently redder than the MSTO) can be described as a simple mixture model of SS and BS populations described by the same functional form (separated by a constant offset $\Delta G = G_{BS} - G_{SS} \simeq -0.75$ mag) plus some intrinsic scatter. In a Gaussian approximation for the scatter, again $f_b(q > q_{\mathrm{lim}}) = w_{BS}$ (as $w_{SS} + w_{BS} = 1$).

If we further assume the observational errors in colour and G magnitude to be uncorrelated, the likelihood can be written as

$$\begin{aligned}\ln \mathcal{L} = \sum_{i=1}^{N} \ln \Bigg[& \frac{(1-f_b)}{\sqrt{2\pi(\sigma_{\mathrm{eff},i}^2 + \sigma_{SS}^2)}} \cdot \mathrm{e}^{-\frac{(M_{G_i} - P(BP-RP)_i)^2}{2(\sigma_{\mathrm{eff},i}^2 + \sigma_{SS}^2)}} \\ & + \frac{f_b}{\sqrt{2\pi(\sigma_{\mathrm{eff},i}^2 + \sigma_{BS}^2)}} \cdot \mathrm{e}^{-\frac{(M_{G_i} - P(BP-RP)_i + \Delta G)^2}{2(\sigma_{\mathrm{eff},i}^2 + \sigma_{BS}^2)}} \Bigg],\end{aligned} \tag{10}$$

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

$$\begin{aligned}\sigma_{\mathrm{eff},i}^2 &:= \frac{1}{1+m_i}(\sigma_{G,i}^2 + m_i \cdot \sigma_{(BP-RP)_i}^2) \\ \text{with} \quad m_i &:= P'(x)^2|_{x=(BP-RP)_i}.\end{aligned} \tag{11}$$

The likelihood in equation (10) is thus a function of $p+5$ parameters, being p the order of the fitted polynomial: $p+1$ nuisance parameters for the polynomial itself, the intrinsic width of the SS σ_{SS}, the intrinsic width of the BS σ_{BS}, the vertical offset between the two sequences ΔG, and the cluster's binary fraction f_b.

Since we are interested in sampling the posterior probability distribution (PPD) 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:

$$p_1(\Delta G) = \mathcal{N}(-0.75, 0.05) \tag{12}$$
$$p_2(\log_{10}\sigma_{BS}) = \mathcal{N}(-1, 0.2) \tag{13}$$
$$p_3(\log_{10}\sigma_{SS}) = \mathcal{N}(-0.8, 0.2) \tag{14}$$
$$p_{\rm full}(\Delta G, \log_{10}\sigma_{SS}, \log_{10}\sigma_{BS}) = p_1 \cdot p_2 \cdot p_3 \tag{15}$$

For each OC, we performed the MCMC fits using the python package `emcee` (Foreman-Mackey et al., 2013), using `n_walkers` = 32, `n_steps` = 15000 and `burnin` = 2000, and for p=6. With these conditions, the median and 16$^{\rm th}$ and 84$^{\rm th}$ percentiles of the $p + 5$ likelihood parameters could be estimated for all 128 OCs from CGa20 that verify the conditions on the MS population that enable their f_b to be estimated (as explained in Sect. 2.2), and for 243 out of the 246 OCs from T22 that verify them. Then, the results for these 371 OCs were double-checked by visual inspection and classified as reliable, possibly doubtful, or problematic; resulting in a reliability parameter between $[0, 1]$ for each OC. For the vast majority of OCs, the adjusted polynomials give an accurate description of the observed SS. In order to disregard those for which the fitted polynomials are not accurate in some range or having dubious curvature, we select only as the final sample of OCs studied through MCMC those with reliability parameter greater than 0.5. The result is a final sample of 324 OCs whose f_b has been estimated using MCMC, 228 ($\sim$ 70%) from T22 and 96 ($\sim$ 30%) from CGa20, which includes all the 84 OCs for which f_b has also been estimated through GMM. Out of these 324 OCs, 300 ($\sim$ 93%) have a reliability parameter greater than 0.75.

Figure 5 shows an illustrative example for an OC with a moderately populated MS, Alessi 1 (which has 46 selected MS members). We see that in this case (as in most others; see Fig. 6) there are little correlations between the fit parameters. Our main desirable parameter $f_b = f_{b,\rm MCMC}(q > q_{\rm lim}^{\rm MCMC})$ is not strongly correlated with other parameters of the model (as can be seen in the marginalised two-dimensional projections of the PPD in the off-diagonal panels of Fig. 5, that show the covariances and display approximately round shapes), and the marginal posterior probability distribution gives a sensible result ($f_b = 0.19^{+0.06}_{-0.05}$), compatible with the GMM approach (Chapter 3).

In Sect. 6.2 we find an estimation of $q_{\rm lim}^{\rm MCMC}$ following the method described in Sect. 5.2.2.

With respect to the GMM implementation, MCMC has the advantages that it simultaneously fits the SS polynomial, the offset defining the parallel BS polynomial, the intrinsic widths of both sequences and their scatter distribution weights (which give f_b), correctly incorporating uncertainties in both dimensions. Hence, the form of the fitted isochrones is more flexible.

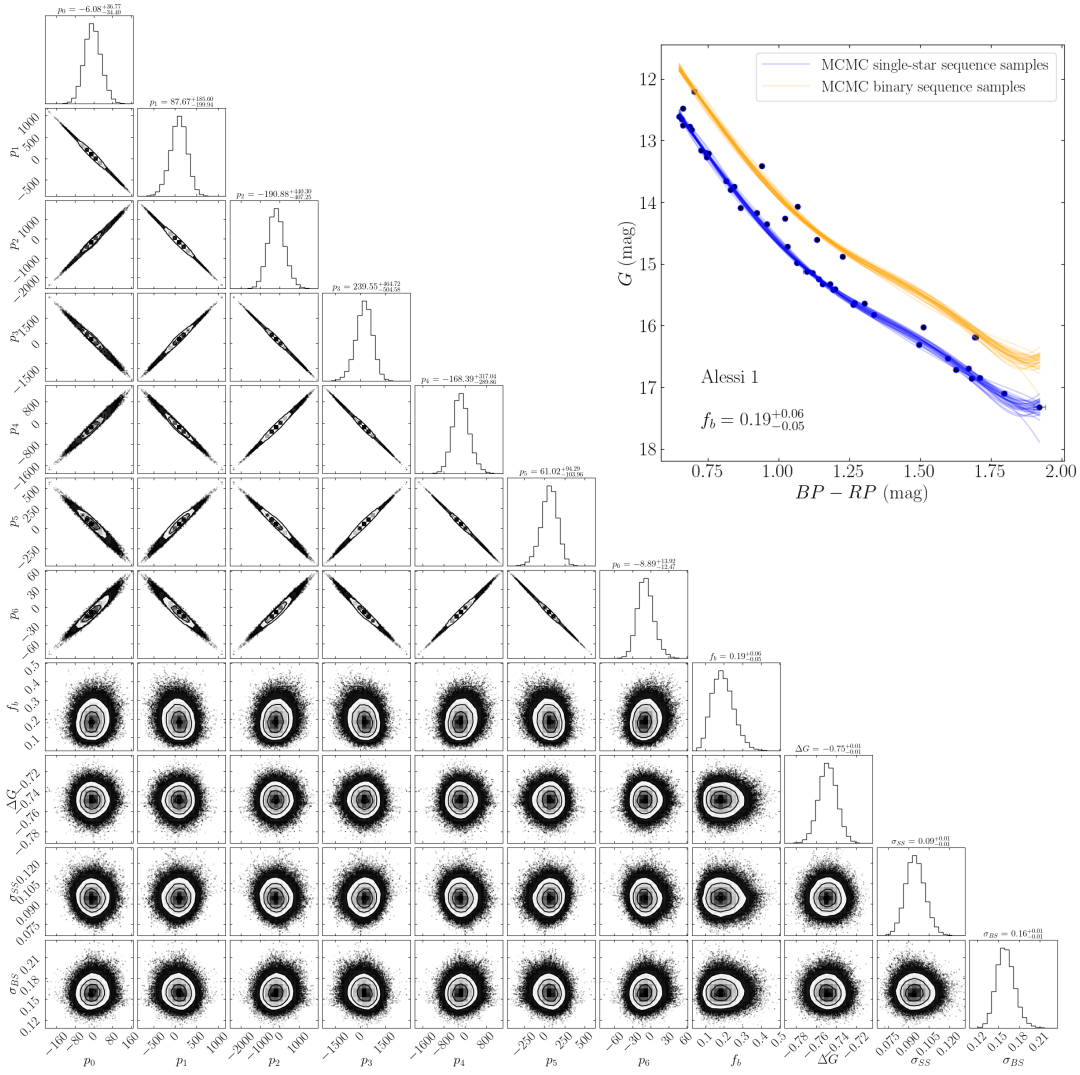


FIGURE 5: MCMC fit for Alessi 1 observational data (from T22). *Lower left panels*: `corner` plot of the posterior probability distribution for the full parameter set. The first 7 parameters correspond to the polynomial coefficients p_i, while the last four describe the unresolved binary fraction f_b, the separation between the binary sequence and the single-star sequence ΔG, the intrinsic G-magnitude spread of the single-star sequence σ_{SS}, and the one of the binary sequence σ_{BS}. The main outcome of the fit is the binary fraction, all other parameter are considered nuisance parameters. *Top right panel*: *Gaia* 's CMD overplotted by 30 random samples of the posterior probability of the single-star sequence and binary sequence 6^{th} order polynomials.

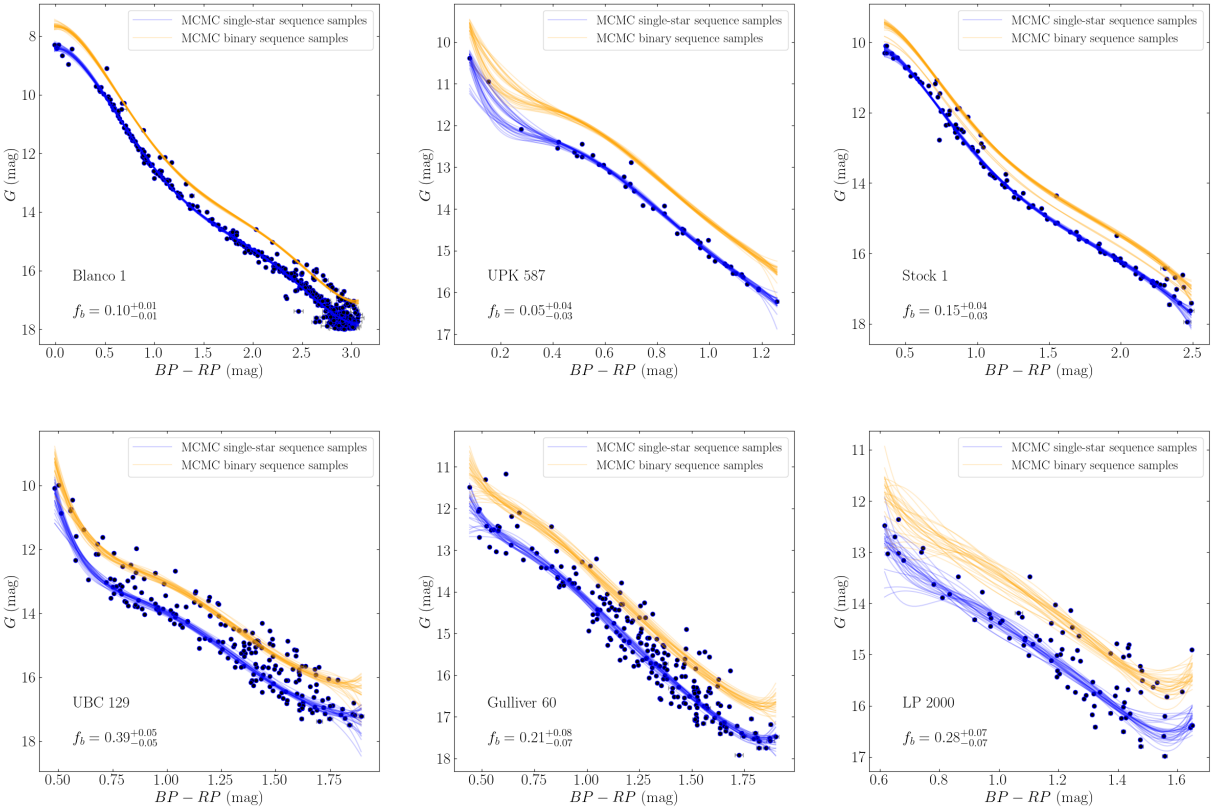


FIGURE 6: Examples of the MCMC fits to the single-star and binary sequences of several open clusters in our final sample (in the same style as Fig. 5 top right panel), illustrating the diversity of the binary sequences and the resulting diversity in the quality of the MCMC fits (assessed visually and translated into a reliability parameter). For most systems the errorbars are smaller than their markers.

Chapter 5

Simulated open clusters

We describe the *Gaia Object Generator* and how we use it to generate realistic simulations of 84 studied OCs. We apply both GMM and MCMC to several realisations of their simulated CMDs. For MCMC we are thus able to estimate $q_{\rm lim}^{\rm MCMC}$ for the observed CMDs, used to estimate the resolved multiplicity fraction in the same $q_{\rm lim}^{\rm MCMC} < q \leq 1$ range as the unresolved one; and also the corresponding total multiplicity fraction $f_b^{tot}(q > q_{\rm lim}^{\rm MCMC})$.

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. For a given population of celestial objects, it applies the spacecraft and payload models to simulate the main *Gaia* observables (astrometry and photometry) with realistic error models. Its complexity is increased after each *Gaia* data release, so that the full 3D spatial distributions and statistics of observables (spectral characteristics to derive photometry and spectroscopy, and motions) of these 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) that have been developed to describe the stellar content of the Milky Way. 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. It also takes into account interstellar extinction based on a 3D model. And, most interestingly for our purpose, it includes a module that simulates multiple star systems (Arenou, 2011) which we use here in conjunction with GOG.

For each OC we want to simulate, 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), and the multiple-star module now considers each input star and may change it with 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, therefore, 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 binary fraction depending on the primary mass of the MS star, which is considered to fit well this dependence in the whole MS mass range:

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

This function is monotonically increasing with M_1, and is roughly compatible (by eye) with the several classes of dynamical decay models from Sterzik and Durisen (2004) or random pairing of Thies and Kroupa (2007). There is also abundant observational evidence for the increase of f_b with M_1 (Kaczmarek, Olczak, and Pfalzner, 2011; 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 linear by segment and depends on the spectral type of the primary and on the binary period. The module also checks that the pairing is realistic in terms of the evolutionary stage of the stars.

For the distribution of semi-major axis a, a Gaussian distribution in $\log(a)$ is assumed, with different 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 eccentricity is simulated uniformly 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 ω_2 uniformly in $[0, 2\pi]$, the position angle of the node Ω 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.

Finally, GOG decides whether a multiple system can be resolved by *Gaia* (so that the components appear all along the same isochrones 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 separation of the components of the system in relation to the telescope resolution, and the limiting apparent magnitude that is detectable. Then, it adds photometric uncertainties, which depend on the magnitude and the calibration.

5.2 Unresolved binary fraction estimation for simulated open clusters

5.2.1 Custom simulations for each open cluster in our sample

For the sample of 84 observed OCs studied with both GMM and MCMC methods, 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 as observed by *Gaia* DR2, now containing single stars and binaries, is returned by GOG as a catalogue of the members (or systems, if unresolved) with "true" and "observed" parameters. Interstellar extinction and 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.

From each of these returned synthetic populations we generate N_{sim} realisations of its corresponding OC, each time drawing the same number of MS members as observed in the real OC. In order to do so, first the selection of the MS members of each synthetic population as observed by *Gaia* is carried out exactly as for the observed OCs (see Sect. 2.2). The resulting simulated populations, however, are still not equivalent to the observed OCs because of the initial mass of the population input in the simulation: in the equivalently selected MS range, the simulated

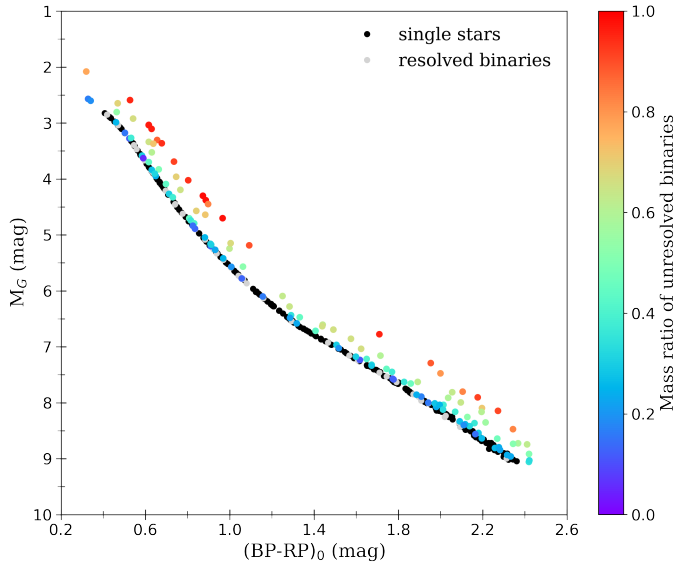


FIGURE 7: Simulated CMD of NGC 2281 ($\log_{10}$(age[yr])=8.79), having 381 main-sequence systems (like the number of selected main-sequence systems in the observed CMD). The black dots are single stars, and the unresolved binary systems are colour-coded by their mass ratio q. The light grey points overlapping with single stars are resolved components of binary systems. The errorbars are smaller than the markers.

populations have many more members than the OCs. This is solved by applying a random selection of MS members without replacement, picking in total the same number of MS members that is found in the selected MS range of the observed CMD of the same OC. 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 for single stars, and then companions have been added. By doing so, the synthetic populations as observed by *Gaia* are in reasonable approximation realisations of possible *Gaia* observations of the corresponding OC in the range where the f_b is studied; and they automatically verify the conditions of having at least 30 MS members and a MS extension of 1 mag in colour index (see Fig. 7 for an example).

For a studied OC, however, instead of generating and studying a single CMD realisation, we generate N_{sim} realisations of the 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.

5.2.2 Application of GMM and MCMC methods to the simulated open clusters

In this section we perform the equivalent estimations of f_b through GMM and MCMC as described in Chapters 3 and 4, but now for a sample of 84 simulated OCs, each of them corresponding to one of the observed OCs which have been studied with both GMM and MCMC.

For the GMM method we have $N_{sim} = 30$, and the analysis is applied to each of the 30 simulated CMDs corresponding to an observed OC as described in Chapter 3, except for the three following considerations. First, the polynomial fitted to the

OC's SS is of eighth degree instead of fourth. Second, the number of fitted Gaussians is imposed to two, regardless of whether the BIC displays a minimum value for two components or not, in order to be able to carry out the study for all 84 OCs. Third, as the CMDs of the simulated OCs generally show less dispersion than the observed ones, the GMM is applied using an initial guess for the mean of the binaries' Gaussian of $\mu_{BS} = -0.75$. It was visually considered to yield slightly better fits of the Gaussians in the ΔG histogram, but the presence or absence of this initial guess does not significantly alter the derived binary fractions (see Sect. 6.1).

Therefore, as 30 GMM fittings and their derived Gaussian parameters and $f^{sim}_{b,\mathrm{GMM}}$ estimation are performed for each OC, a statistical treatment of these results is carried out. We consider the $f^{sim}_{b,\mathrm{GMM}}$ of the simulated OC to be the mean of the 30 $f^{sim}_{b,\mathrm{GMM}}$ calculated ($\overline{f^{sim}_{b,\mathrm{GMM}}}$), and its uncertainty the mean of the 30 $f^{sim}_{b,\mathrm{GMM}}$ Wald-estimated errors ($\overline{\delta[f^{sim}_{b,\mathrm{GMM}}]}$); and we also compute the standard deviation of $f^{sim}_{b,\mathrm{GMM}}$ over the 30 realisations of the CMD ($\sigma_{f^{sim}_{b,\mathrm{GMM}}}$).

For the MCMC method, considerably more time consuming for each OC than the GMM, we have $N_{sim} = 10$, and the analysis is carried out exactly as explained in Chapter 4 for the observed OCs. We consider the binary fraction of the simulated OC, $\overline{f^{sim}_{b,\mathrm{MCMC}}}$, to be the mean of the 10 values of $f^{sim}_{b,\mathrm{MCMC}}$'s median, and its nominal uncertainty $\overline{\delta[f^{sim}_{b,\mathrm{MCMC}}]}$ as the mean over the 10 realisations of the mean of its 16$^{\mathrm{th}}$ and 84$^{\mathrm{th}}$ percentiles; and we also compute the standard deviation of $f^{sim}_{b,\mathrm{MCMC}}$ over the 10 realisations ($\sigma_{f^{sim}_{b,\mathrm{MCMC}}}$).

From the simulated OCs, we can derive additional information apart from the application of both methods to estimate f_b for a certain number of CMD realisations of an OC as observed by *Gaia*. As the 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 N_{sim} realisations: $\overline{f^{sim,theo}_b}(q > 0)$. In order to use this information to characterize GMM and MCMC methods for f_b estimation, for each OC we compute the mean theoretical f_b of unresolved multiple systems having a q equal or greater than 11 different values: $\overline{f^{sim,theo}_b}(q \geq q_{\mathrm{min}})$ (with q_{min} in the range [0.2, 0.7] in steps of 0.05), for the $N_{sim} = 30$ realisations. These values are then compared to $\overline{f^{sim}_{b,\mathrm{GMM}}}$ and to $\overline{f^{sim}_{b,\mathrm{MCMC}}}$, so that the $q^{\mathrm{GMM,sim}}_{\mathrm{lim}}$ and $q^{\mathrm{MCMC,sim}}_{\mathrm{lim}}$ values above which each method takes into account binaries in f_b estimation are estimated for each simulated OC as the q_{min} for which $\overline{f^{sim,theo}_b}(q \geq q_{\mathrm{min}})$ is the closest possible to the mean f_b applying the corresponding method to the simulated CMDs ($\overline{f^{sim}_{b,\mathrm{GMM}}}$ or $\overline{f^{sim}_{b,\mathrm{MCMC}}}$).

5.3 Estimation of the total binary fraction: correction for resolved binaries

One of the most nearby OCs is the Pleiades cluster. It was recently studied in detail by Torres, Latham, and Quinn (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. GMM and MCMC methods provide an estimation of the unresolved f_b of systems with $q > q_{\rm lim}$ based on their CMD positions, where $q_{\rm lim}$ is in principle characteristic of each OC and method. We therefore have to resort to simulations to correctly account for resolved systems, in order to be able to estimate the total f_b: $f_b^{tot} = f_b^{\rm unresolved} + f_b^{\rm resolved}(d, age)$. As our unresolved f_b is limited to systems of $q > q_{\rm lim}$, we also estimate f_b^{tot} of unresolved and resolved binaries of $q > q_{\rm lim}$, which can be calculated as follows:

$$f_b^{tot}(q > q_{\rm lim}) = f_{b,\rm unres}^{\rm measured} \left[1 + \frac{f_{b,\rm res}^{\rm sim}(q > q_{\rm lim};\ d, age)}{f_{b,\rm unres}^{\rm sim}(q > q_{\rm lim};\ d, age)} \right] \tag{16}$$

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 how many binaries in a given OC are actually resolved by *Gaia* and thus do not appear in the BS of the *Gaia* CMD. This estimate depends on the treatment of wide binaries in the *Gaia* Universe Model; and a resolution must be set. Following Fabricius et al. (2021), we have chosen 500 mas, so that binaries with greater separations are considered to be resolved by *Gaia*. 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 resolved f_b over unresolved f_b ratio for each distance, with a certain $q_{\rm lim}$ value. The result is represented in Fig. 8 for a value $q_{\rm lim} = 0.6$ (justified and discussed in Sect. 6.2). There is not a strong dependence on the OC's age, so we decide to fit a quadratic function to the dependence with the logarithm of the distance common for all four OCs (black solid line, its functional form is specified in Fig. 8 legend). This is used to interpolate the value of the ratio for each OC in our sample in order to estimate its $f_b^{tot}(q > q_{\rm lim})$.

We do not estimate the total f_b of all binary systems because we would need to add the unresolved and resolved binaries of $q < q_{\rm lim}$, what requires a modelling of the q distribution and depends on the specific properties of each binary system of the OC. Such study is out of the scope of this work.

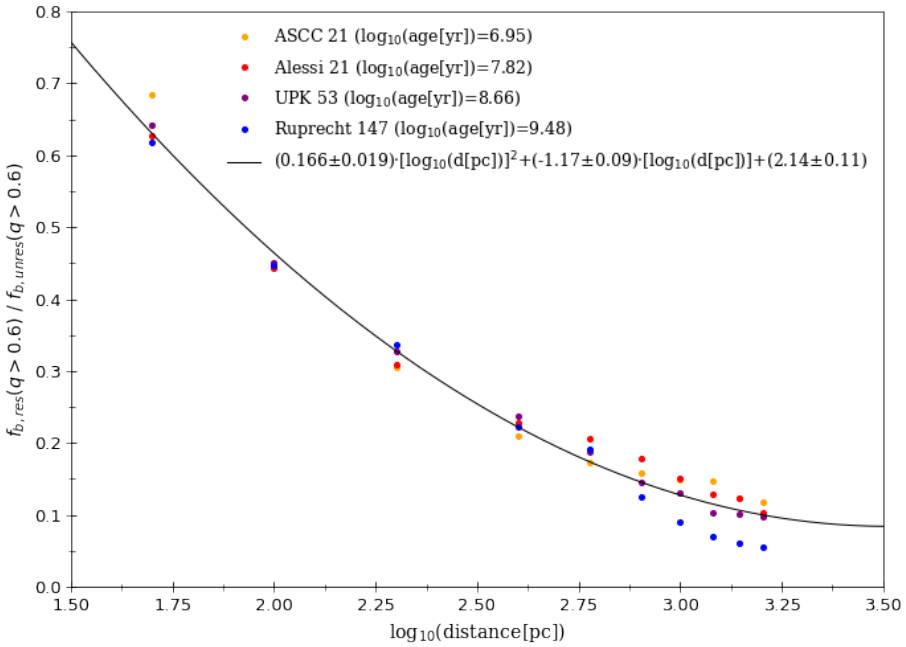


FIGURE 8: Representation of the theoretical ratio of the fraction of resolved binaries of $q > 0.6$ over the fraction of unresolved binaries of $q > 0.6$ $\left(\frac{f^{\mathrm{sim}}_{b,\mathrm{res}}(q>q_{\mathrm{lim}}=0.6;\, d)}{f^{\mathrm{sim}}_{b,\mathrm{unres}}(q>q_{\mathrm{lim}}=0.6;\, d)} \right)$, as a function of the decimal logarithm of the distance, for four open clusters simulated with GOG. The black solid curve is a quadratic fit to all four open clusters.

Chapter 6

Results

We describe and compare the estimations of the unresolved binary fraction through GMM and MCMC; and also characterise the total one estimated for MCMC. We finally describe our catalogue of the high-q MS binary fraction estimation in 324 OCs using MCMC.

6.1 Binary fractions estimated through GMM

The sample of 84 observed OCs studied through GMM suffers a limitation in the studied f_b range arising from the applicability of this method: it works properly for OCs with $w_{SS} > 0.7$, implying $f_{b,\mathrm{GMM}} = w_{BS} < 30\%$ (see Chapter 3). The sample comprises values of $f_{b,\mathrm{GMM}}$ between 3% and 30%, with errors estimated with the Wald formula ranging from 1% to 8%. The most populated OC has 1051 MS members, the least populated one 32; and there is a 2% of OCs with more than 500 MS members. The median of the least massive MS member selected is $0.71M_{\odot}$, and of the most massive member is $1.98M_{\odot}$.

The corresponding determinations of f_b as the mean over 30 realisations of each simulated OC ($\overline{f^{sim}_{b,\mathrm{GMM}}}$) are also found to be smaller than 30% for all OCs, ranging from 17% to 29%, with mean Wald-estimated errors ranging from 1% to 7%. On the other hand, $\sigma_{f^{sim}_{b,\mathrm{GMM}}}$ ranges from 2% to 10%. Its relation to the mean of the Wald-estimated errors is represented in Fig. 9, which shows that the latter accounts well for the typical variation of $f^{sim}_{b,\mathrm{GMM}}$ over the different realisations, being just slightly underestimated. Both estimations, however, only account for the statistical error in the Gaussian fitting process, and are hence still underestimated if the uncertainty in the MS selection and SS polynomial fitting were to be considered. For the observed OCs, nonetheless, no standard deviation of $f_{b,\mathrm{GMM}}$ can be calculated because GMM is applied to the single observed CMD, so we conclude that the Wald formula provides a reasonable estimation of $f_{b,\mathrm{GMM}}$'s statistical error. Fig. 9 also displays the expected tendency of the typical variation in f_b over different GMM fittings being smaller for more populated OCs, whose SS and BS are better traced and have distributions less dependent on particular random selections of MS members than for the least populated OCs.

As a basic check of the GMM performance, we corroborate that for 78/84 (93%) OCs $\overline{f^{sim}_{b,\mathrm{GMM}}} < \overline{f^{sim,theo}_{b}}(q > 0)$, as expected because the SS Gaussian inevitably accounts for a fraction of low-q unresolved binaries and hence we derive $\overline{f^{sim}_{b,\mathrm{GMM}}}(q > q^{\mathrm{GMM}}_{\mathrm{lim}})$, for which $\overline{f^{sim,theo}_{b}}(q > 0)$ is an upper limit. Regarding the rest of OCs, 5/6 have a $\overline{f^{sim}_{b,\mathrm{GMM}}}$ which is compatible within 1σ with being smaller than $\overline{f^{sim,theo}_{b}}(q > 0)$.

The distribution of $q^{\mathrm{GMM,sim}}_{\mathrm{lim}}$ values for the 84 simulated OCs, for which $\overline{f^{sim,theo}_{b}}(q \geq q^{\mathrm{GMM,sim}}_{\mathrm{lim}})$ is the closest possible to $\overline{f^{sim}_{b,\mathrm{GMM}}}$ for each OC, has values ranging from 0.2 to 0.45, being both the mean and the median 0.3, and the 10^{th} and 90^{th} percentiles 0.2 and 0.4.

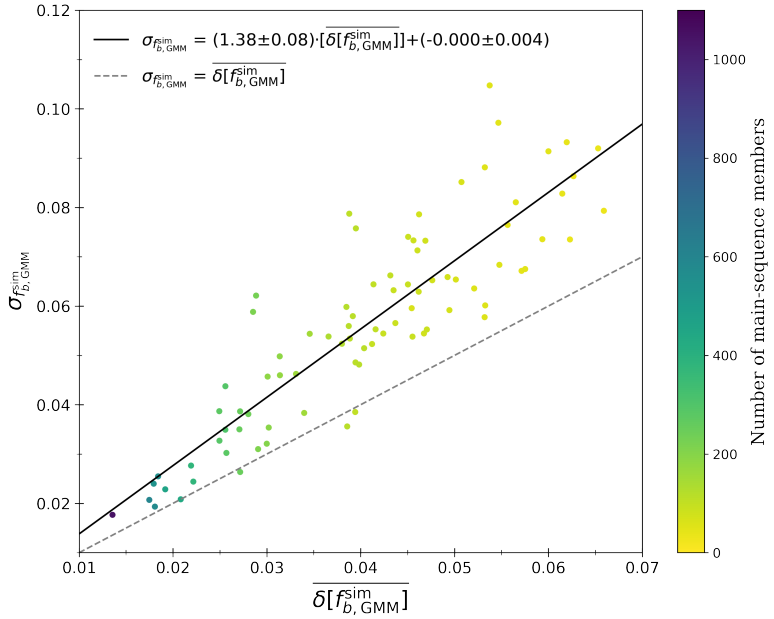


FIGURE 9: Standard deviation of the estimated $f^{sim}_{b,\mathrm{GMM}}$ over the 30 CMD realisations of each simulated OC to which GMM is applied ($\sigma_{f^{sim}_{b,\mathrm{GMM}}}$) against the mean over the same realisations of the $f^{sim}_{b,\mathrm{GMM}}$ errors estimated with the Wald formula; colour-coded by the number of selected MS members. The black solid line is a linear fit, and the dashed grey one the identity line.

Regarding the assessment of the assumptions made, one is that a 2-component GMM has been forced for all realisations of all OCs. It is found that the most frequent value of the BIC minimum over the 30 realisations is actually at two Gaussians for 63% of the OCs in the sample. The assumption of an initial guess for the mean of the binaries' Gaussian is found not to be critical because for all 84 simulated OCs their derived $\overline{f^{sim}_{b,\mathrm{GMM}}}$ using this initial guess are compatible within 1σ with the ones derived not using it.

Table 1 shows the means of the estimated f_b and of the Gaussians' parameters and their Wald-estimated errors over the 84 OCs, applying GMM to the observed and simulated CMDs. For 71/84 (85%) OCs, $\overline{f^{sim}_{b,\mathrm{GMM}}} > f_{b,\mathrm{GMM}}$ (see distributions in left panel of Fig. 10). This is consistent with the fact that for the observed and simulated CMDs $\overline{\sigma_{BS}}$ is compatible (see Table 1), while the simulated CMDs have on average a much smaller SS Gaussian width ($\overline{\sigma^{sim}_{SS}}$=0.037$\pm$0.016 against $\overline{\sigma^{obs}_{SS}}$=0.15$\pm$0.05), implying that the relative weight of the binaries' Gaussian (and thus the f_b) increases. This is due to the fact that the simulated CMDs have a better delineated SS, with intrinsically less dispersion than the observed CMDs' one, so that the Gaussian fitted to the SS is narrower and does not contribute as much for the low-q systems as the one of the observed CMDs. This is also reflected in the mean separation between the mean positions of both Gaussians, which is $\overline{\Delta G}$=-0.65$\pm$0.14 for the observed CMDs and nearly the half for the simulated CMDs: $\overline{\Delta G}$=-0.31$\pm$0.03 (see distributions in Fig. 10, right panel). So, as the simulated CMDs have on average the binaries' Gaussian centred on lower q than the observed CMDs, and at the same time their SS Gaussian width is smaller, this results in their BS Gaussian accounting for a wider spectrum of q values (thus estimating an unresolved f_b that

GMM mean parameters	Observed OCs	Simulated OCs
$\overline{f_{b,\mathrm{GMM}}}$	0.16 ± 0.03	0.24 ± 0.04
$\overline{\sigma_{f^{sim}_{b,\mathrm{GMM}}}}$	-	0.06
$\overline{\mu_{SS}}$	0.04±0.02	0.011±0.005
$\overline{\mu_{BS}}$	-0.61±0.14	-0.30±0.03
$\overline{\sigma_{SS}}$	0.15±0.05	0.037±0.016
$\overline{\sigma_{BS}}$	0.23±0.06	0.250±0.019
$\overline{\Delta G}$	-0.65±0.14	-0.31±0.03

TABLE 1: Means over the 84 OCs studied through GMM of the estimated unresolved binary fraction and its Wald-estimated error, and of the single-sequence and binary-sequence Gaussians' parameters (mean position with respect to the SS fitted polynomial μ, width w, and separation between them ΔG); for the observed CMD (left column) and averaged over the 30 CMD realisations of each simulated OC (right column). Only for the latter we have a measure of the dispersion of the fitted GMM binary fraction over the 30 different CMD realisations of each simulated OC, which we also average over all the 84 simulated OCs ($\overline{\sigma_{f^{sim}_{b,\mathrm{GMM}}}}$).

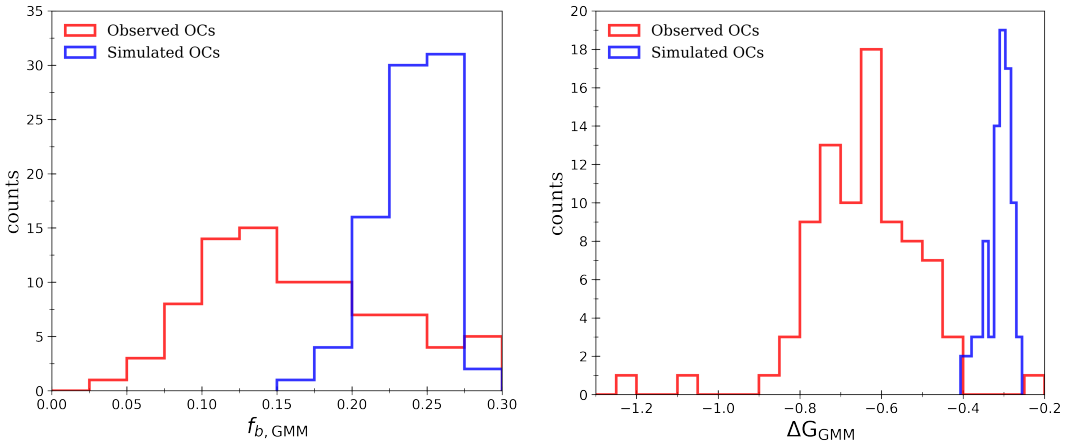


FIGURE 10: *Left:* Histograms of the estimated unresolved binary fraction applying GMM to the observed CMDs ($f_{b,\mathrm{GMM}}$; red) and to 30 CMD realisations of each simulated OC ($\overline{f^{sim}_{b,\mathrm{GMM}}}$; blue) for the 84 OCs studied through GMM. *Right:* Histograms of the separation between the SS and BS fitted Gaussians applying GMM, for the same samples as in the left panel.

takes into account more systems) than for the observed CMDs. Hence, we conclude that $q^{\mathrm{GMM,sim}}_{\mathrm{lim}} < q^{\mathrm{GMM,obs}}_{\mathrm{lim}}$: the BS Gaussian for the simulated OCs takes into account systems down to a mass ratio smaller than for the observed OCs, for which it is centred at higher q values (nearer to the equal-mass binary sequence) so that it is the SS Gaussian the one accounting mainly for the low-q systems. This is critical for our intended purpose of identifying the q above which our estimated f_b for the observed OCs takes into account binaries, because as the GMM method does not behave equivalently for their simulated counterparts, $q^{\mathrm{GMM,sim}}_{\mathrm{lim}} = 0.3^{+0.1}_{-0.1}$ cannot be directly equalled to our unknown $q^{\mathrm{GMM,obs}}_{\mathrm{lim}}$ (it is rather a lower limit of it). Therefore,

the simulations of the OCs are not realistic enough for the purpose of characterizing the GMM method applied to the observed OCs, because this method is highly sensitive to the dispersion in the CMD to determine the mean position of the BS Gaussian and the Gaussians' widths, and this dispersion is significantly lower in the simulated CMDs than in the real observed ones (reflected as smaller uncertainties of the parameters in Table 1 for the simulated OCs). Consequently, the derived results applying GMM to simulated CMDs cannot be directly used to gain insight into the results obtained applying GMM to observed CMDs. This explains why in 42/84 (50%) of the cases $f_{b,\mathrm{GMM}}$ is not compatible within 1σ with $\overline{f^{sim}_{b,\mathrm{GMM}}}$. Within 2σ they are compatible in 70/84 (83%) of the cases.

6.2 Binary fractions estimated through MCMC

The sample of 324 observed OCs studied through MCMC that have a reliability parameter greater than 0.5 comprises values of $f_{b,\mathrm{MCMC}}$ (taken as the median) between 4% and 62%, thus doubling the f_b range covered with respect to GMM method. Unlike GMM, MCMC takes into account that its errors are not symmetric: the upper errorbar $\sigma^U_{f_{b,\mathrm{MCMC}}} = f_{b,\mathrm{MCMC}}(84^{\mathrm{th}}) - f_{b,\mathrm{MCMC}}$ over the sample has a median 4.3% and for 93% of the OCs is smaller than 10%, while the lower one $\sigma^L_{f_{b,\mathrm{MCMC}}} = f_{b,\mathrm{MCMC}} - f_{b,\mathrm{MCMC}}(16^{\mathrm{th}})$ has a median 3.8% and for 97% of the OCs is smaller than 10%. The most populated OC has 1791 MS members, the least populated one 31; and there is an 8% of OCs with more than 500 MS members. The median of the least massive MS member selected is $0.72M_{\odot}$, and of the most massive member is $2.04M_{\odot}$.

Out of this sample of 324 observed OCs studied through MCMC, the 84 OCs that have also been studied through GMM and that are thus simulated too have values of $f_{b,\mathrm{MCMC}}$ between 7% and 38%. Their corresponding determinations of $\overline{f^{sim}_{b,\mathrm{MCMC}}}$ over 10 CMD realisations of each simulated OC to which MCMC is applied are found to be within a similar range, between 7% and 35%. The mean $\overline{\sigma^U_{f_{b,\mathrm{MCMC}}}}$ has a median 4.0% and is between 1.4% and 9.1%; while $\overline{\sigma^L_{f_{b,\mathrm{MCMC}}}}$ has a median 3.3% and is between 1.4% and 5.4%. On the other hand, $\sigma_{f^{sim}_{b,\mathrm{MCMC}}}$ ranges from 1% to 11%. Its relation to the mean over the 10 realisations of the nominal error $\delta[f^{sim}_{b,\mathrm{MCMC}}] = \frac{\sigma^U_{f^{sim}_{b,\mathrm{MCMC}}} + \sigma^L_{f^{sim}_{b,\mathrm{MCMC}}}}{2}$ is represented in Fig. 11. It shows that both error measures are positively correlated between them, and negatively correlated with the number of MS members as expected. Unlike for some OCs studied with GMM, applying MCMC to the 84 simulated OCs $\overline{f^{sim}_{b,\mathrm{MCMC}}}$ is actually found to be smaller than $\overline{f^{sim,theo}_b}(q > 0)$ for all OCs (as expected).

The distribution of $q^{\mathrm{MCMC,sim}}_{\mathrm{lim}}$ values for the 84 simulated OCs for which $\overline{f^{sim,theo}_b}(q \geq q^{\mathrm{MCMC,sim}}_{\mathrm{lim}})$ is the closest possible to $\overline{f^{sim}_{b,\mathrm{MCMC}}}$ for each OC, displays an almost bimodal dependence on ΔG, changing at $\Delta G \simeq -0.65$. As for 311/324 (96%) studied observed OCs $\Delta G < -0.65$, we choose to estimate $q^{\mathrm{MCMC}}_{\mathrm{lim}}$ with the simulated OCs for which MCMC gives $\Delta G < -0.65$, which are 60/84. Their distribution of $q^{\mathrm{MCMC}}_{\mathrm{lim}}$ has values ranging from 0.2 to 0.7, being the median 0.6, and the 10^{th} and 90^{th} percentiles 0.45 and 0.65. Therefore, this derived $q^{\mathrm{MCMC}}_{\mathrm{lim}}$ value is accurate for the vast majority of the observed OCs studied with MCMC, because a very low proportion of OCs in the sample have $\Delta G > -0.65$ (as can be seen in the right panel of Fig. 12). Our found $q^{\mathrm{MCMC}}_{\mathrm{lim}}$ median of 0.6 is sensible considering what is stated in Jadhav et al. (2021): for the $M_G \in [1, 10]$ mag range they study (which is also approximately our

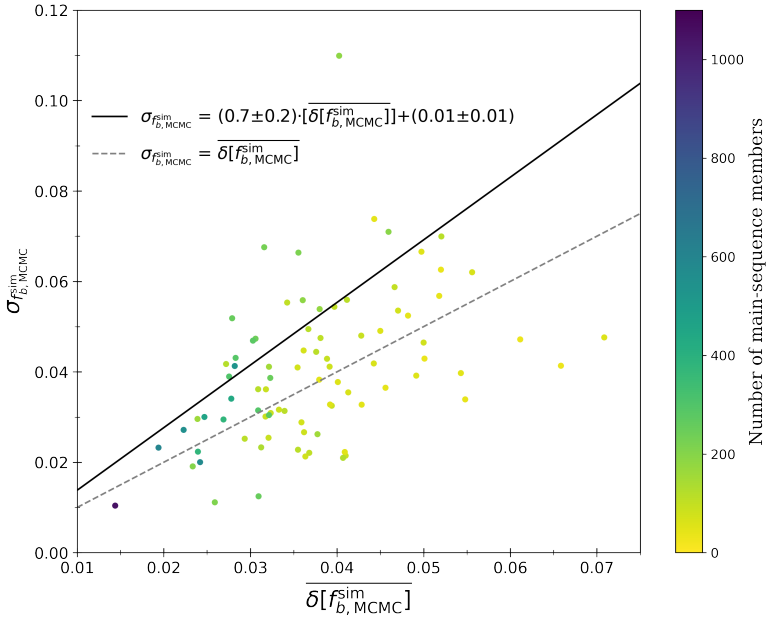


FIGURE 11: Standard deviation of the estimated $f^{sim}_{b,\mathrm{MCMC}}$ over the 10 CMD realisations of each simulated OC to which MCMC is applied ($\sigma_{f^{sim}_{b,\mathrm{MCMC}}}$) against the mean over the same realisations of the $f^{sim}_{b,\mathrm{MCMC}}$ nominal errors; colour-coded by the number of selected MS members. The black solid line is a linear fit, and the dashed grey one the identity line.

range of selected MS members), setting $q = 0.6$ as the cut-off above which binaries are considered as such ensures that the single MS stars are 3σ away from the thus defined BS. It is reasonable too that we find $q^{\mathrm{MCMC,sim}}_{\mathrm{lim}} = 0.6^{+0.05}_{-0.15} > q^{\mathrm{GMM,sim}}_{\mathrm{lim}} = 0.3^{+0.1}_{-0.1}$, because as can be seen in Tables 1 and 2, for SS and BS widths of the same order, applying MCMC the absolute value of ΔG is doubled with respect to the value for GMM.

Table 2 shows the means of the estimated f_b and of the Gaussians' widths and separation, and the means of their 16$^{\mathrm{th}}$ and 84$^{\mathrm{th}}$ uncertainty percentiles, applying MCMC to the observed and simulated CMDs. Comparing the results for the common observed and simulated OCs, for 51/84 OCs (61%) their $f_{b,\mathrm{MCMC}} > \overline{f^{sim}_{b,\mathrm{MCMC}}}$. For 53/84 OCs (63%), $f_{b,\mathrm{MCMC}}$ is compatible within 1σ with $\overline{f^{sim}_{b,\mathrm{MCMC}}}$ (see their similar distributions in Fig. 12, left panel), and for 78/84 OCs (93%) they are compatible within 2σ.

The values of the mean over all 84 OCs of the standard deviations (widths) of the SS and BS Gaussians are quite similar applying GMM and MCMC, for both observed and simulated OCs (see Tables 1 and 2). The values of ΔG, however, differ more for both the observed and simulated OCs applying GMM or MCMC. With MCMC, for the observed CMDs the mean ΔG is -0.732, very close to the imposed prior, and for the simulated CMDs the mean ΔG is reduced to -0.66. Hence, both ΔG for simulated and observed OCs are closer for MCMC than for GMM. Given this similarity between the SS and BS widths and ΔG for the observed and simulated OCs obtained through MCMC, we conclude that the GOG-simulated CMDs are realistic enough so that applying MCMC to them we are able to derive information of the MCMC application to the observed CMDs, in particular: $q^{\mathrm{MCMC,sim}}_{\mathrm{lim}} = 0.6^{+0.05}_{-0.15} \approx q^{\mathrm{MCMC,obs}}_{\mathrm{lim}}$.

MCMC mean parameters	Observed OCs	Observed OCs (84/324)	Simulated OCs
$\overline{f_{b,\mathrm{MCMC}}}$	$0.20^{+0.06}_{-0.04}$	$0.16^{+0.05}_{-0.04}$	$0.15^{+0.04}_{-0.03}$
$\overline{\sigma_{f^{sim}_{b,\mathrm{MCMC}}}}$	-	-	0.04
$\overline{\sigma_{SS}}$	$0.160^{+0.012}_{-0.013}$	$0.121^{+0.009}_{-0.008}$	$0.083^{+0.007}_{-0.007}$
$\overline{\sigma_{BS}}$	$0.209^{+0.018}_{-0.018}$	$0.184^{+0.016}_{-0.014}$	$0.190^{+0.019}_{-0.017}$
$\overline{\Delta G}$	$-0.75^{+0.02}_{-0.02}$	$-0.732^{+0.012}_{-0.011}$	$-0.66^{+0.02}_{-0.02}$

TABLE 2: Means over the OCs studied through MCMC of the median unresolved binary fraction and its 16th and 84th uncertainty percentiles, and of the widths of the fitted single-sequence and binary sequence Gaussians in the G-magnitude dimension (σ_{SS} and σ_{BS}) and the separation between their means (ΔG). The left column shows these parameters for all the 324 observed OCs studied through MCMC, the central one for the subsample of 84 observed OCs that coincide with the simulated ones, and the right column for the 84 simulated OCs (for which we have a measure of the dispersion of the fitted MCMC binary fraction over the 10 different CMD realisations, which we average over all the 84 OCs).

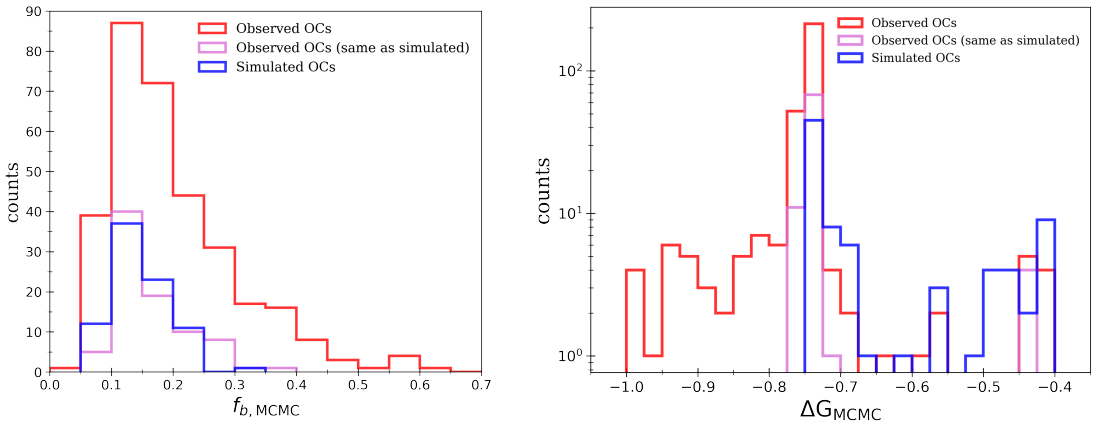


FIGURE 12: *Left:* Histograms of the estimated unresolved binary fraction applying MCMC to all the 324 observed CMDs ($f_{b,\mathrm{MCMC}}$; red), to the observed CMDs of the 84 OCs that have been simulated (pink), and to 10 CMD realisations of each of the 84 simulated OCs ($\overline{f^{sim}_{MCMC}}$; blue). *Right:* Histograms of the separation between the SS and BS fitted Gaussians applying MCMC, for the same samples as in the left panel and with the counts in logarithmic scale.

6.3 Comparison between GMM and MCMC methods

In order to compare the performance of both methods for the f_b estimation applied to observed data, we evaluate the results for the 84 common observed OCs. For 73/84 (87%) of them, $f_{b,\mathrm{GMM}}$ and $f_{b,\mathrm{MCMC}}$ are compatible within 1σ; and for 81/84 (96%) they are compatible within 2σ (considering the Wald formula errors for GMM and the nominal ones for MCMC). For 68/84 OCs (81%), $\left|f_{b,\mathrm{GMM}} - f_{b,\mathrm{MCMC}}\right| < 5\%$. The distribution of their difference has a mean approximately zero: for nearly half of these OCs (43/84) $f_{b,\mathrm{GMM}} > f_{b,\mathrm{MCMC}}$; and for the rest it is the other way around. It does not follow a Gaussian distribution, nor it appears to depend on the OC's parameters. This compatibility between both results (Fig. 13, left panel) exists despite the difference in the ΔG values of both methods: MCMC mainly adopts values near

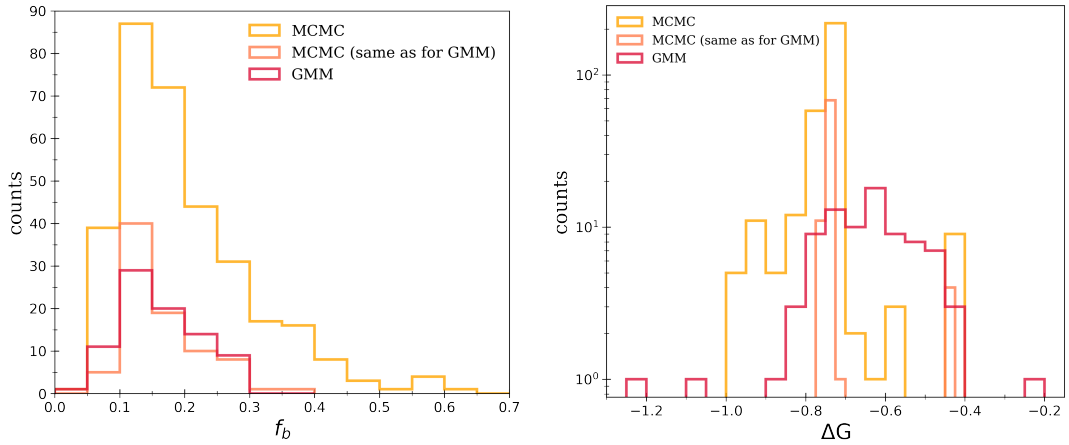


FIGURE 13: *Left:* Histograms of the estimated unresolved binary fraction of the observed OCs, applying MCMC to all the 324 observed CMDs (yellow) and to the observed CMDs of the 84 OCs that have also been studied with GMM (orange), and applying GMM to these 84 observed OCs (crimson). *Right:* Histograms of the separation between the SS and BS fitted Gaussians applying MCMC and GMM to observed OCs, for the same samples as in the left panel, with the counts in logarithmic scale.

the prior at -0.75, while GMM displays a wider range (Fig. 13, right panel).

The f_b uncertainties derived with both methods are also overall compatible. $\sigma^L_{f_{b,\mathrm{MCMC}}}$ are more similar to the errors from Wald formula than $\sigma^U_{f_{b,\mathrm{MCMC}}}$ but, as can be seen in Fig. 14, MCMC nominal error $\left(\delta[f_{b,\mathrm{MCMC}}] = \frac{\sigma^U_{f_{b,\mathrm{MCMC}}} + \sigma^L_{f_{b,\mathrm{MCMC}}}}{2}\right)$ is clearly correlated with the error estimated with the Wald formula from GMM ($\delta[f_{b,\mathrm{GMM}}]$). Both if the clear outliers are taken into account or not, the resulting linear regression is compatible with the identity line. The slight tendency of MCMC uncertainties to be larger can be attributed to the fact that they are more realistic because they include the uncertainties in the polynomial fitting, while GMM does not.

It must be recalled that both estimated f_b do not necessarily have the same definition, because for MCMC it takes into account systems of $q > q^{\mathrm{MCMC,obs}}_{\mathrm{lim}} = 0.6^{+0.05}_{-0.15}$ while for GMM it has not been possible to characterise its $q^{\mathrm{GMM,obs}}_{\mathrm{lim}}$. But, still, their values are compatible and with similar errors for both methods. So, the possible effect introduced by a different q_{lim} is found to be within the errors for most of the sample, which is reasonable given that it is expected that $q^{\mathrm{GMM,sim}}_{\mathrm{lim}} = 0.3^{+0.1}_{-0.1} < q^{\mathrm{GMM,obs}}_{\mathrm{lim}}$, and that observing the GMM fitting, $q^{\mathrm{GMM,obs}}_{\mathrm{lim}}$ cannot be very close to $q = 1$, so that $q^{\mathrm{GMM,obs}}_{\mathrm{lim}}$ and $q^{\mathrm{MCMC,obs}}_{\mathrm{lim}}$ do not have extremely different values.

Regarding the comparison of the performance of both methods for the f_b estimation applied to simulated OCs, Fig. 15 shows what has already been discussed separately before. For the observed OCs, $f_{b,\mathrm{GMM}}$ and $f_{b,\mathrm{MCMC}}$ are overall compatible, and for MCMC the f_b for the observed and simulated OCs ($f_{b,\mathrm{MCMC}}$ and $\overline{f^{sim}_{b,\mathrm{MCMC}}}$) are also generally compatible; but for GMM not. It is this discrepancy in GMM's performance applied to observed and simulated data in contrast to MCMC's similar performance for observed and simulated CMDs that causes that for only 39/84 (46%) of the simulated OCs $\overline{f^{sim}_{b,\mathrm{GMM}}}$ and $\overline{f^{sim}_{b,\mathrm{MCMC}}}$ are compatible within 1σ. The value found

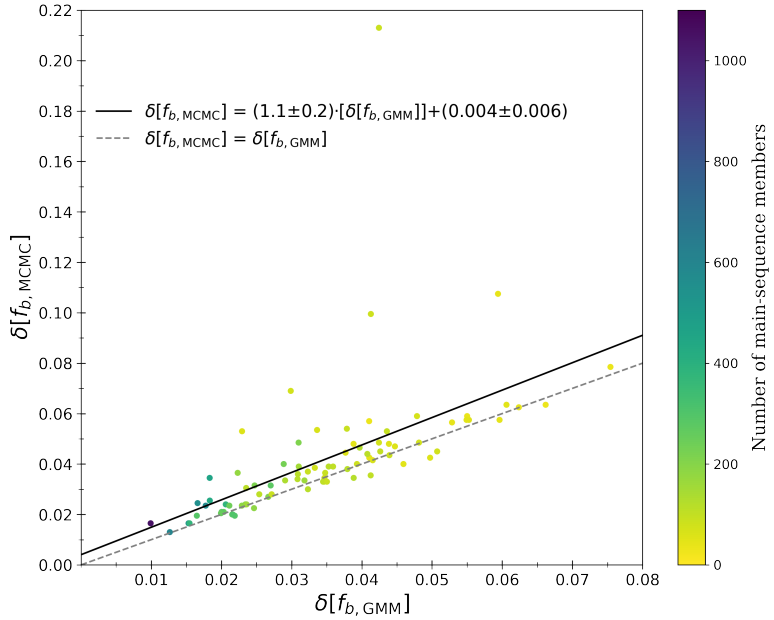


FIGURE 14: Nominal error of the unresolved binary fraction estimation through MCMC against its error estimated with the Wald formula when GMM is applied, for the 84 common observed OCs. The colour scale indicates the number of selected MS members. The black solid line is a linear fit, and the dashed grey one the identity line.

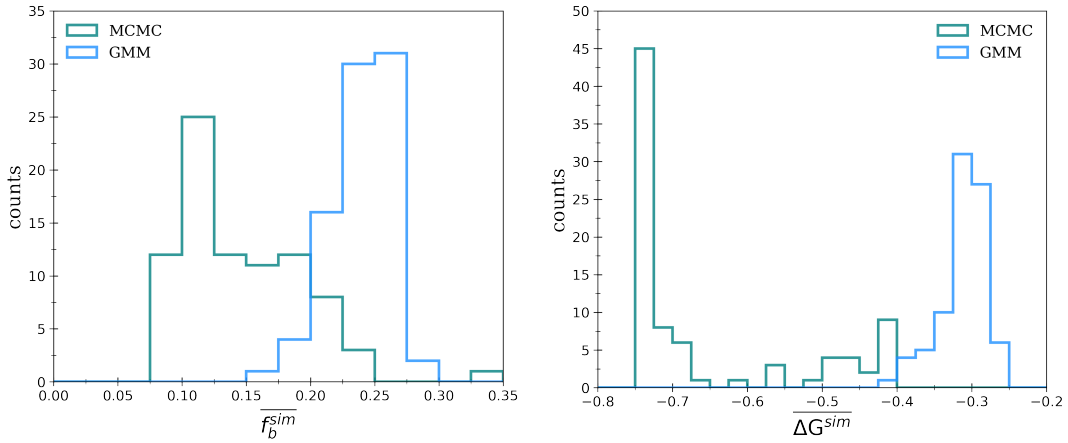


FIGURE 15: *Left:* Histograms of the estimated mean unresolved binary fraction of the 84 simulated OCs, applying MCMC to 10 CMD realisations (green) and GMM to 30 CMD realisations (blue). *Right:* Histograms of the mean separation between the SS and BS fitted Gaussians applying MCMC and GMM to simulated OCs, for the same samples as in the left panel.

for $q_{\mathrm{lim}}^{\mathrm{GMM,sim}} = 0.3^{+0.1}_{-0.1}$, although not being translatable into a minimum q threshold detected for the estimated f_b of the observed OCs ($f_{b,\mathrm{GMM}}$), is reasonable because it is found that for all 84 OCs $\overline{f_{b,\mathrm{GMM}}^{sim}} > \overline{f_{b,\mathrm{MCMC}}^{sim}}$, and this makes sense as it verifies $q_{\mathrm{lim}}^{\mathrm{GMM,sim}} < q_{\mathrm{lim}}^{\mathrm{MCMC,sim}}$.

In conclusion, we have applied two different methods for the unresolved f_b estimation in OCs, GMM and MCMC, and for their 84 common studied OCs the results are overall compatible. MCMC has the advantages that it is successfully applied to study OCs of higher f_b than GMM, and also with a larger number of MS members. While GMM works essentially for clearly delineated SS and BS, being restricted to values $f_{b,\mathrm{GMM}} < 30\%$, MCMC can deal with more disperse CMDs, which tend to be more populated too. For instance, for the Pleiades OC, with 1158 MS selected members, its f_b estimation has been possible through MCMC but not with GMM. This is why the sample of 324 OCs studied with MCMC is nearly four times larger than the one studied with GMM. However, due to our modelling and the limitations on the photometric determinations of binaries in the CMD, our estimated f_b cannot take into account systems down to the smallest q. Therefore, our obtained values are not the total unresolved f_b and we need to characterise how many of the unresolved binaries we are actually taking into account. This is what simulations allow us to estimate for the simulated OCs (see Sect. 5.2.2), but there must be a sufficient correspondence between the f_b estimated for the observed and simulated OC in order to consider that their q_{lim} values are compatible and thus deduce $q_{\mathrm{lim}}^{\mathrm{obs}} \approx q_{\mathrm{lim}}^{\mathrm{sim}}$. For GMM this compatibility between observed and simulated results is very poor because the method is very sensitive to the smaller dispersion the simulated OCs have. Therefore, our determined f_b comprises an unknown q range. But for MCMC the compatibility is remarkable and we have been able to infer $q_{\mathrm{lim}}^{\mathrm{MCMC,obs}} = 0.6^{+0.05}_{-0.15}$, so that our estimated f_b corresponds to systems of $q_{\mathrm{lim}}^{\mathrm{MCMC,obs}} < q \leq 1$, and also to estimate the resolved binary fraction and the total binary fraction of systems of $q > q_{\mathrm{lim}}^{\mathrm{MCMC,obs}}$ for each OC (as explained in Sect. 5.3).The rest $f_b^{tot}(q > q_{\mathrm{lim}}^{\mathrm{MCMC,obs}}) - f_{b,\mathrm{MCMC}}(q > q_{\mathrm{lim}}^{\mathrm{MCMC,obs}})$ has a mean of 0.03, and for 282/324 (87%) OCs is smaller than 0.05. Its mean uncertainties are $\sigma^U_{f_b^{tot}} = 0.07$ and $\sigma^L_{f_b^{tot}} = 0.05$, and have been calculated carrying out the same estimation of f_b^{tot} but of q values above the two percentiles of $q_{\mathrm{lim}}^{\mathrm{MCMC,obs}}$. Again, they do not include the systematic errors derived from the accuracy of the simulations' modelling. It is reasonable the small correction that the resolved f_b introduces, due to the distances of the studied OCs and the typical separations between the components of a binary system.

Consequently, MCMC is a better method because it enables the study of a larger sample of OCs and having a wider range of f_b, thus providing more statistically significant results; and also because $q_{\mathrm{lim}}^{\mathrm{MCMC,obs}}$ can be inferred. This is why in the following Chapter 7 we perform the analysis of the dependencies on several physical parameters of the total f_b of systems with $q > 0.6$, $f_b^{tot}(q > 0.6^{+0.05}_{-0.15})$, calculated using the unresolved f_b estimated with MCMC ($f_{b,\mathrm{MCMC}}(q > 0.6^{+0.05}_{-0.15})$).

6.4 Catalogue of open cluster main-sequence multiplicity fractions

Our results are summarized in a catalogue, which will soon be published in an online repository along with a paper on this study (work in progress). It corresponds to our sample of 324 OCs studied with MCMC which have a reliability parameter greater than 0.5, 228 from Tarricq et al. (2022) and 96 from Cantat-Gaudin et al. (2020). We thus present the largest homogeneous catalogue of multiplicity fractions in OCs to date.

For each OC the catalogue provides the values and uncertainties of the unresolved binary fraction deduced with MCMC and the estimated total binary fraction (of both resolved and unresolved systems), both of MS systems with $q > 0.6^{+0.05}_{-0.15}$. There is also the intrinsic colour index range that the considered MS systems comprise and its translation into masses; as well as the reliability parameter of $f_{b,\mathrm{MCMC}}(q > 0.6^{+0.05}_{-0.15})$ and the values of the rest of parameters which MCMC fits.

Chapter 7

Discussion

We compare, when possible, our estimated unresolved $f_{b,\mathrm{MCMC}}(q > 0.6^{+0.05}_{-0.15})$ of the OCs' MSs, integrated over q dependencies, with f_b estimations of some recent studies for which we have common studied OCs. Then we discuss the dependence of $f^{tot}_{b,\mathrm{MCMC}}(q > 0.6^{+0.05}_{-0.15})$ on OC distance and members' mass, and on position and age.

7.1 Comparison to other open cluster studies

The multiplicity fractions in many OCs have been studied before; however, large homogeneous studies are still rare. The results of different studies, moreover, are very seldom directly comparable. Apart from the data and OC's 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. 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 a well-accepted approach to model binaries in CMD-fitting works. While it generally provides f_b values very similar to the ones found for more detailed prescriptions for the binaries, its problem is that obviously does not properly account for the fraction of resolved binaries (considered inexistent) and the fraction of unresolved binaries. As the fraction of resolved binaries decreases with distance, the closer the studied sample, the less appropriate it is to assimilate the fitted f_b assuming all binaries to be unresolved to the actual unresolved f_b.

In this section 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 $f_{b,\mathrm{MCMC}}$ with their results, instead of our calculated f^{tot}_b that takes into account the resolved binaries. The first three commented studies rely on the CMD of G vs. $(BP - RP)$, as our study, while Li et al. (2022) use V vs. (V - I).

Niu, Wang, and Fu (2020) used *Gaia* DR2 photometry and LAMOST spectroscopy 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 also found q distribution in their OCs to be flat, in accordance with the finding of Torres, Latham, and Quinn (2021) for the Pleiades. They also provide the limiting values m_{min} (minimum mass estimated by the faintest MS star in the OC) and m_{max} (estimated by the brightest MS star in the OC) of the mass range for which f_b has been estimated. Table 3 shows the comparison to our study for the 8 OCs we have in common. None of the studied mass ranges coincides and, as is shown in Fig. 16A, our estimated f_b are smaller than theirs for all cases. This is consistent with the fact that they both study mass ranges up to higher masses (while m_{min} does not differ so much) and are sensitive to a wider q range than us, thus detecting more binaries.

Open cluster	This work		Niu, Wang, and Fu (2020)	
	Mass range ($M_\odot$)	$f_{b,\mathrm{MCMC}}(q > 0.6^{+0.05}_{-0.15})$	Mass range ($M_\odot$)	$f_b(0 < q < 1)$
Melotte 22	[0.18, 2.29]	$0.078^{+0.009}_{-0.009}$	[0.57, 3.75]	0.41 ± 0.04
NGC 1662	[0.56, 1.60]	$0.16^{+0.04}_{-0.04}$	[0.44, 2.30]	0.42 ± 0.06
NGC 1750	[0.73, 1.89]	$0.28^{+0.03}_{-0.03}$	[0.61, 3.28]	0.54 ± 0.06
NGC 2099	[0.96, 1.73]	$0.223^{+0.02}_{-0.019}$	[0.72, 2.54]	0.53 ± 0.04
NGC 2168	[0.73, 2.07]	$0.138^{+0.03}_{-0.019}$	[0.61, 3.43]	0.50 ± 0.08
NGC 2281	[0.58, 1.65]	$0.095^{+0.018}_{-0.015}$	[0.47, 2.59]	0.47 ± 0.04
NGC 2548	[0.71, 1.76]	$0.097^{+0.014}_{-0.013}$	[0.54, 2.48]	0.38 ± 0.06
Roslund 6	[0.40, 2.23]	$0.085^{+0.02}_{-0.017}$	[0.34, 3.60]	0.41 ± 0.08

TABLE 3: Comparison between our estimated unresolved binary fraction of systems with $q > 0.6$ and the one of Niu, Wang, and Fu (2020) (of systems with $0 < q < 1$ and for different open cluster mass ranges), for the 8 open clusters we have in common.

None of the values is compatible within 1σ or 2σ, and for only 2/6 OCs both f_b are compatible within 3σ.

Furthermore, Niu, Wang, and Fu (2020) recalculate the f_b for Melotte 22 and NGC 2099 in mass ranges coincident with the ones of other published studies. We also perform our $f_{b,\mathrm{MCMC}}$ calculation in these two same mass ranges, 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 (see Table 4). For the Pleiades OC, our inferred f_b is still smaller than the one of Niu, Wang, and Fu (2020), being compatible only within 3σ. This could be attributed to the fact their minimum detected q is lower, aside from intrinsic discrepancies. Our inferred f_b is also compatible only within 3σ with the one of Pinfield et al. (2003), which is larger than ours too despite being computed for a q range compatible with ours. For NGC 2099 our inferred f_b is also smaller than the one of Niu, Wang, and Fu (2020), being compatible only within 3σ. Compared to the work of Cordoni et al. (2018), however, our value and theirs are computed for very similar q ranges and are compatible within 1σ.

Open cluster	Mass range	This work	Niu, Wang, and Fu (2020)	Other works
		$f_{b,\mathrm{MCMC}}(q > 0.6^{+0.05}_{-0.15})$	$f_b(0 < q < 1)$	f_b
Melotte 22	[0.6, 1.0]	$0.086^{+0.012}_{-0.011}$	0.20 ± 0.03	$0.23^{+0.06}_{-0.05}$ ($q \in [0.5, 1.0]$) Pinfield et al. (2003)
NGC 2099	[1.06, 1.63]	$0.11^{+0.03}_{-0.03}$	0.23 ± 0.02	0.085 ($q \in [0.7, 1.0]$) Cordoni et al. (2018)

TABLE 4: Comparison between our estimated unresolved binary fraction of $q > 0.6$ and two estimations of it from studies that comprise different q ranges than ours but the same main-sequence mass range for each open cluster.

Jadhav et al. (2021) 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. We take into account, therefore, unresolved binaries in the same q range. They do not provide, however, the mass range for which f_b is estimated, only the total number of MS members for each OC. Table

Open cluster	This work			Jadhav et al. (2021)	
	Mass range ($M_\odot$)	Number of members	$f_{b,\mathrm{MCMC}}(q > 0.6^{+0.05}_{-0.15})$	Number of members	$f_b(q > 0.6)$
IC 4756	[0.66, 1.48]	381	$0.16^{+0.03}_{-0.02}$	543	0.31 ± 0.03
Melotte 22	[0.18, 2.29]	1158	$0.078^{+0.009}_{-0.009}$	1326	0.14 ± 0.02
NGC 1039	[0.49, 2.11]	601	$0.106^{+0.013}_{-0.013}$	764	0.17 ± 0.02
NGC 2168	[0.73, 2.07]	1243	$0.138^{+0.03}_{-0.019}$	1794	0.23 ± 0.01
NGC 2360	[0.79, 1.55]	577	$0.25^{+0.02}_{-0.02}$	1037	0.19 ± 0.02
NGC 2422	[0.43, 2.18]	530	$0.082^{+0.013}_{-0.012}$	907	0.14 ± 0.02
NGC 2423	[0.77, 1.54]	256	$0.119^{+0.02}_{-0.019}$	694	0.17 ± 0.02
NGC 2447	[0.73, 1.66]	616	$0.24^{+0.02}_{-0.02}$	926	0.19 ± 0.02
NGC 2516	[0.34, 1.91]	1791	$0.23^{+0.02}_{-0.12}$	2518	0.16 ± 0.01
NGC 2547	[0.32, 2.63]	206	$0.14^{+0.04}_{-0.03}$	644	0.22 ± 0.04
NGC 2548	[0.71, 1.76]	500	$0.097^{+0.014}_{-0.013}$	509	0.17 ± 0.02
NGC 2682	[0.71, 1.17]	658	$0.37^{+0.02}_{-0.02}$	1520	0.22 ± 0.02
NGC 3532	[0.52, 1.76]	1727	$0.231^{+0.013}_{-0.013}$	1879	0.13 ± 0.01
NGC 6025	[0.86, 2.18]	253	$0.111^{+0.02}_{-0.019}$	452	0.21 ± 0.02
NGC 6281	[0.64, 1.67]	456	$0.32^{+0.04}_{-0.03}$	573	0.27 ± 0.02
NGC 6405	[0.37, 2.79]	534	$0.124^{+0.017}_{-0.016}$	967	0.17 ± 0.02
NGC 6793	[0.71, 1.80]	205	$0.179^{+0.03}_{-0.03}$	465	0.17 ± 0.03
NGC 752	[0.58, 1.52]	278	$0.20^{+0.03}_{-0.03}$	433	0.19 ± 0.03
Trumpler 10	[0.29, 2.48]	436	$0.159^{+0.019}_{-0.018}$	947	0.12 ± 0.02

TABLE 5: Comparison between our estimation of the unresolved binary fraction of $q > 0.6$ and its estimation from Jadhav et al. (2021) (also for $q > 0.6$ but for unspecified open cluster mass ranges), for the 19 open clusters we have in common.

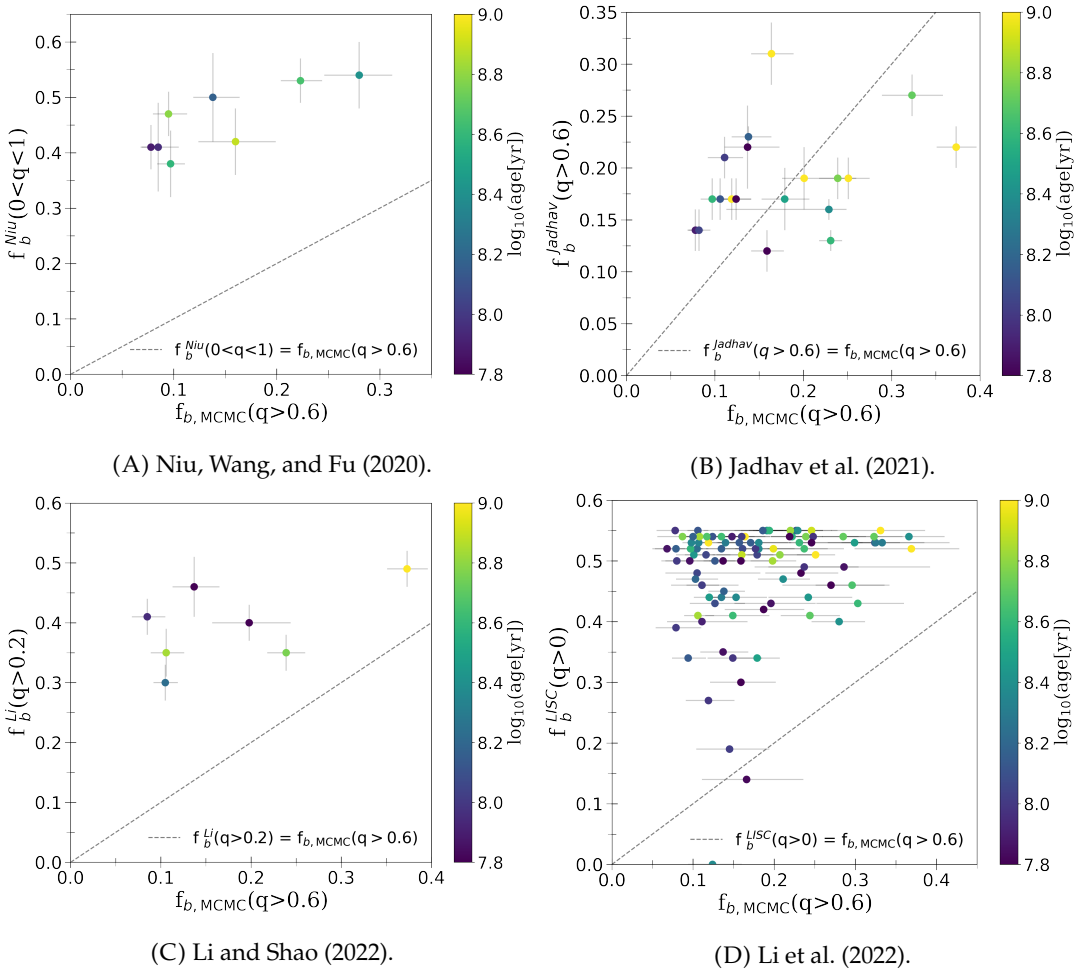
5 shows the comparison to our study for the 19 OCs we have in common. Their considered number of members is larger than ours for all OCs, although for some OCs they are comparable and for others our sample represents only a 30% of theirs. As seen in Fig. 16B, for 9 OCs our estimated f_b is larger than theirs, while for 10 OCs is smaller. For 4 OCs (NGC 2516, NGC 6281, NGC 6793 and NGC 752), our estimated f_b is compatible within 1σ with theirs. In total there are 12 OCs with compatible f_b estimations within 2σ, and 17 within 3σ. These discrepancies can be real if the studied mass range is common for each OC, or can otherwise arise (at least partially) from the fact that they refer to different OC members.

Li and 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 they 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, but do not provide neither the studied mass range nor the number of selected MS members. Table 6 and Fig. 16C show the comparison to our study for the 7 OCs we have in common. None is compatible within 1σ or 2σ, and 3 are compatible within 3σ. Both results, therefore, are not overall compatible. Their f_b is larger than ours for all OCs, what could be attributed to the fact that they reach a lower q threshold, but the studied mass range must also be taken into account and, as they do not provide it, no conclusions can be drawn.

Finally, Li et al. (2022) determine the primordial f_b fitting the CMD morphologies using the *Powerful CMD* code. They provide a catalogue for 309 OCs for which they consider the fit to be good, and for 288 OCs for which it is not as good, without

Open cluster	This work		Li and Shao (2022)
	Mass range ($M_\odot$)	$f_{b,\mathrm{MCMC}}(q > 0.6^{+0.05}_{-0.15})$	$f_b(q > 0.2)$
ASCC 21	[0.18, 4.47]	$0.20^{+0.05}_{-0.04}$	0.40 ± 0.03
NGC 2287	[0.68, 2.01]	$0.105^{+0.014}_{-0.013}$	0.30 ± 0.03
NGC 2447	[0.73, 1.66]	$0.24^{+0.02}_{-0.02}$	0.35 ± 0.03
NGC 2527	[0.71, 1.62]	$0.106^{+0.02}_{-0.017}$	0.35 ± 0.04
NGC 2682	[0.71, 1.17]	$0.37^{+0.02}_{-0.02}$	0.49 ± 0.03
Roslund 6	[0.40, 2.23]	$0.085^{+0.02}_{-0.017}$	0.41 ± 0.03
Trumpler 3	[0.71, 2.29]	$0.14^{+0.03}_{-0.02}$	0.46 ± 0.05

TABLE 6: Comparison between our estimation of the unresolved binary fraction of $q > 0.6$ and its estimation from Li and Shao (2022) (for systems with $q > 0.2$ but for unspecified open cluster mass ranges), for the 7 open clusters we have in common.



(A) Niu, Wang, and Fu (2020).

(B) Jadhav et al. (2021).

(C) Li and Shao (2022).

(D) Li et al. (2022).

FIGURE 16: Unresolved binary fraction estimation from four different studies (specified in the caption of each panel) as a function of our estimated unresolved binary fraction of $q > 0.6$, for the common open clusters between the two studies in each case. As a guidance for comparison, the equality line is drawn (dashed grey line).

specifying either the mass or q ranges considered or the f_b uncertainty. Our study and their good-fit one have 98 OCs in common. For 96 OCs their estimated f_b is larger than ours (see Fig. 16D), but for UBC 260 they have determined a $f_b^{\mathrm{LISC}}(q > 0) = 0.0$ (unrealistic), while ours is $f_{b,\mathrm{MCMC}}(q > 0.6^{+0.05}_{-0.15}) = 0.12^{+0.04}_{-0.03}$. For one OC (UBC 183) both measures are compatible within 1σ, for 2 OCs they are compatible within 2σ, and for 6 OCs they are compatible within 3σ. This study, compared to ours and the three previously mentioned, has the particularity of having many f_b values between 0.5 and 0.55. Out of the 98 OCs we have in common, 64 have values in this range. This f_b distribution they obtain might be the result of an imposed prior and does not seem realistic enough. A similar situation happens applying `ASteCA` to 84 observed OCs (see Appendix): its three f_b estimators are larger than our f_b estimation for all OCs, although their f_b distributions are broader than the one of Li et al. (2022).

7.2 Dependence of the binary fraction on the distance and mass

Figure 17 shows the dependency with distance of the unresolved f_b of $q > 0.6$ derived with MCMC (top) and of the estimated total f_b of $q > 0.6$ (bottom). Firstly, it shows that the number of studied OCs increases with distance, with only 11 OCs out of the 324 studied being inside a sphere of radius of 200 pc centred on the Sun. This is simply because the volume studied increases as the third power of the distance. Beyond this distance, there is a wide dispersion of f_b values for a same value of distance. Still, for both the unresolved and total f_b the running median displays a general tendency to increase with distance. For $f_b^{tot}(q > 0.6^{+0.05}_{-0.15})$, however, it is slightly less steep. This could be attributed to the fact that, to calculate it, the considered resolved f_b term does take into account the dependence with distance; although its reliability depends on how realistic the simulations used are. In any case, the dependence on distance coming from the estimated unresolved f_b plays a major role. The main reason for its increase with distance is the fact that for further 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. This is clearly seen in the bottom panel of Fig. 17, where the colourbar corresponds to the mean MS mass of each OC, computed as follows 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 $m \geq 0.5\ \mathrm{M}_{\odot}$):

$$\langle m \rangle = \frac{\int_{m_{min}}^{m_{max}} m\xi(m)\,dm}{\int_{m_{min}}^{m_{max}} \xi(m)\,dm} = \frac{1-\alpha}{2-\alpha} \cdot \frac{{m_{max}}^{2-\alpha} - {m_{min}}^{2-\alpha}}{{m_{max}}^{1-\alpha} - {m_{min}}^{1-\alpha}} \tag{17}$$

We find that for increasing distance, the m_{max} observed does not vary significantly, while m_{min} increases. This results in the MS being shortened from the reddest end so that the further the OC is, a smaller portion of its upper MS end is seen, which comprises stars that are more massive (yielding thus a larger $\langle m \rangle$) and have higher f_b. Hence, the tendency that Fig. 17 shows of the unresolved and total f_b increasing with distance reflects not an intrinsic dependence, but 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 depends on the mass of the stars. We therefore find the increase of f_b with the mass of the primary star to be in agreement with observational evidence.

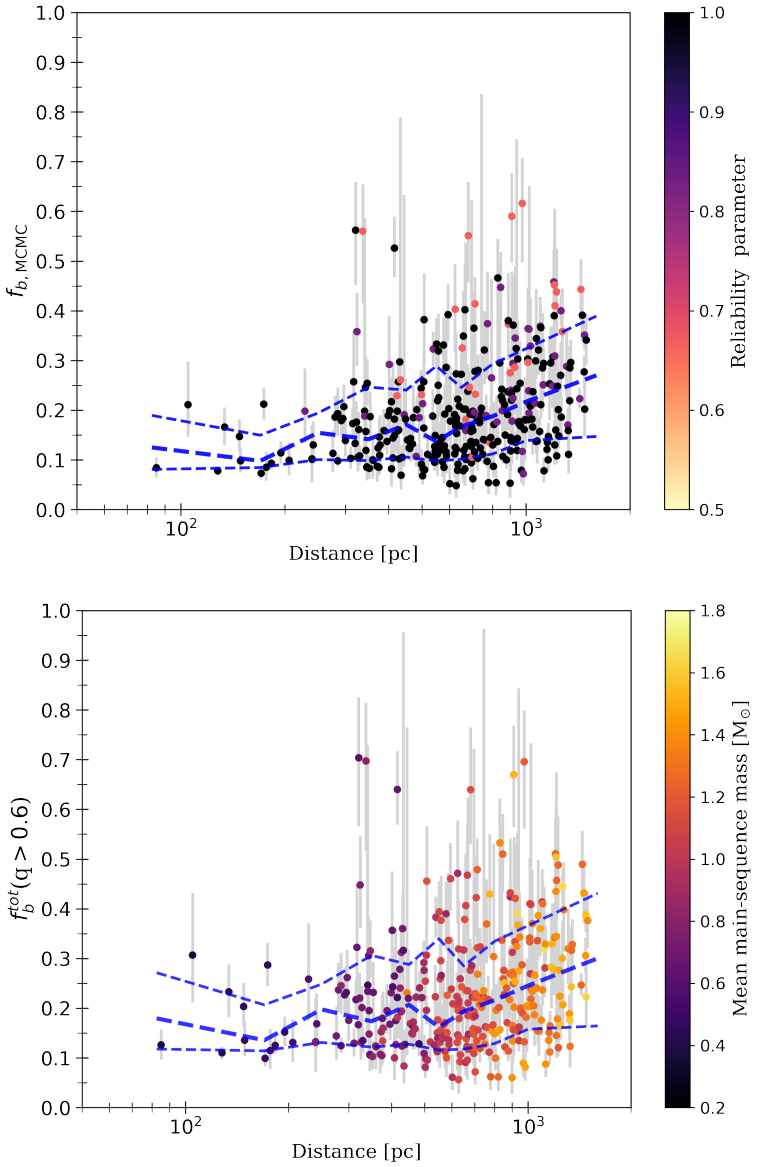


FIGURE 17: *Top*: Unresolved binary fraction of systems with $q > 0.6$ for the 324 open clusters in our sample, estimated through MCMC, as a function of the distance in logarithmic scale, colour-coded by the reliability parameter of the MCMC fit. *Bottom*: Total binary fraction of systems with $q > 0.6$ for the 324 open clusters in the sample as a function of the distance in logarithmic scale, colour-coded by the mean main-sequence mass of each open cluster. For both figures, the central thick dashed blue line is the running median, and the thin blue dashed lines are the 16^{th} and 84^{th} running percentiles.

7.3 Dependence of the binary fraction on position

Figures 18 and 19 account for the three-dimensional spatial distribution of the sample of 324 OCs studied and its relation to the total f_b. Figure 18 shows the projection onto the Galactic plane of their positions, colour-coded by the total f_b. They are approximately evenly distributed around the Sun, some of them reaching the centre of the Sagittarius–Carina arm, while the centre of Perseus arm is out of the reach of the distance limit of 1.5 kpc of our sample. No apparent correlation between the total f_b and the $(X_\odot, Y_\odot)$ position is seen.

Figure 19 shows the OC's height above the Galactic plane as a function of the distance from the Galactic centre, colour-coded by the total f_b. All the OCs fall inside the solar neighbourhood scale height for the thin disc, $z_d \approx 300$ pc, except one: NGC 2682, at $Z = 470$ pc. As before, no clear correlation is seen between the total f_b and the Z coordinate.

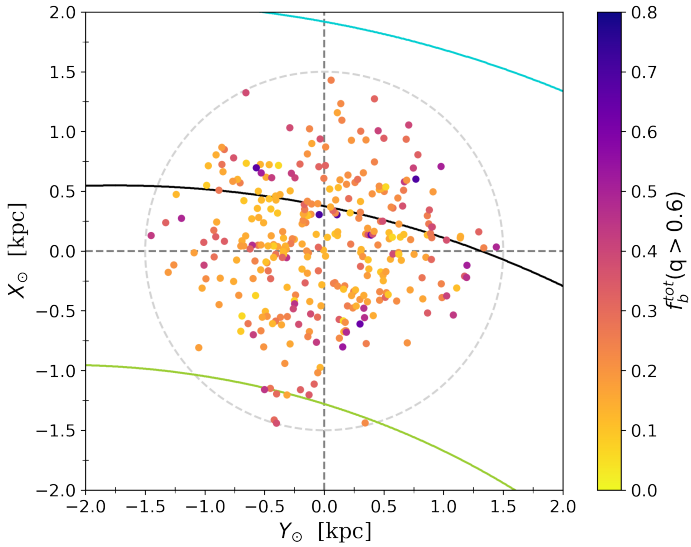


FIGURE 18: Projection on the Galactic plane of the locations of the 324 open clusters in our catalogue (within a distance of 1.5 kpc, indicated by the dashed circumference), colour-coded by the estimated total binary fraction of main-sequence members with $q > 0.6$ (the Galactic Centre is towards negative $X_\odot$). The coloured solid curves are the centres of the Local arm (black), Sagittarius–Carina arm (green) and Perseus arm (blue), as defined in Reid et al. (2019).

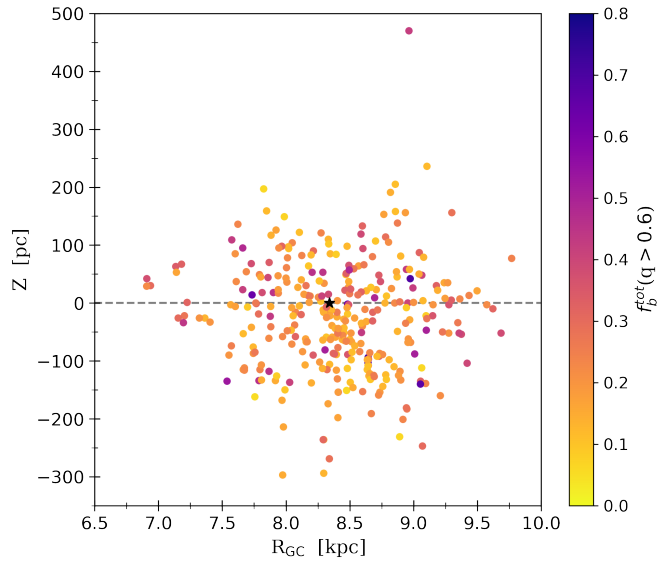


FIGURE 19: Height above the Galactic plane versus Galactocentric radius for the 324 open clusters in our catalogue, colour-coded by the estimated total binary fraction of main-sequence members with $q > 0.6$. The Sun's position at $R_{GC} = 8.34$ kpc is indicated with a black asterisk symbol.

7.4 Dependence of the binary fraction on age

Figure 20 shows the dependence of $f^{tot}_{b,\mathrm{MCMC}}(q > 0.6^{+0.05}_{-0.15})$ on the OC's age. It displays a large dispersion, although the running median decreases with age for the younger OCs up to approximately $\log_{10}(\mathrm{age[yr]}) \in [8.0, 8.5]$, and for OCs older than this value it increases with age. The initial decreasing trend can be explained by binary disruption through gravitational interactions, while creating binaries from two single stars is much less common. The age range for which we find an inversion of the trend is compatible with the value found for 8 OCs in Thompson et al. (2021), $\log_{10}(200 \cdot 10^6 yr) = 8.30$, for which the decreasing f_b trend with age stops; although in their case after these first 200 Myr of the OC's lifetime f_b becomes fairly constant instead of increasing.

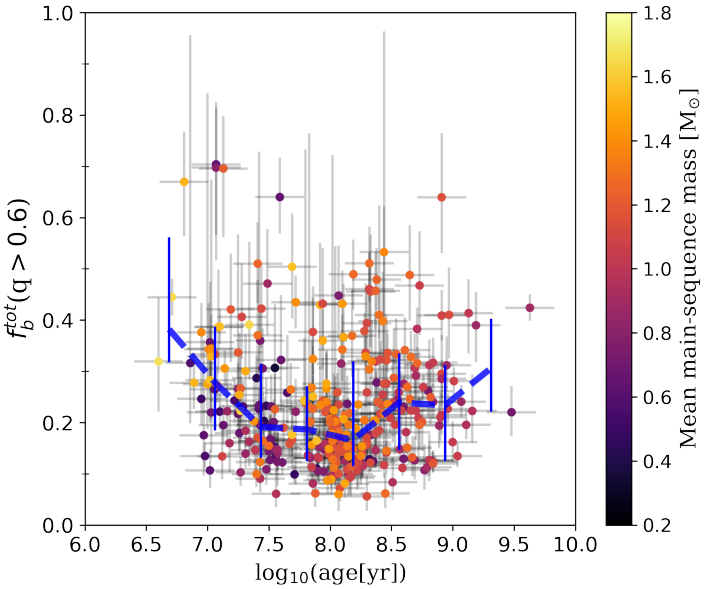


FIGURE 20: Total binary fraction of systems with $q > 0.6$ for the 324 open clusters in our sample as a function of the decimal logarithm of their age in years, colour-coded by the mean main-sequence mass of each open cluster. The dashed blue line is the running median, and the errorbars are calculated as the 16^{th} and 84^{th} percentiles.

In the left panel of Fig. 21 we confirm what was already seen before: 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 redder values of the faint end and, so, the less massive and more abundant stars are the ones 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 it is the reason why the top right corner of the figure is empty. The right image of this figure is the same but now coloured according to the total f_b instead of the distance. In this case, a clear correlation cannot be appreciated.

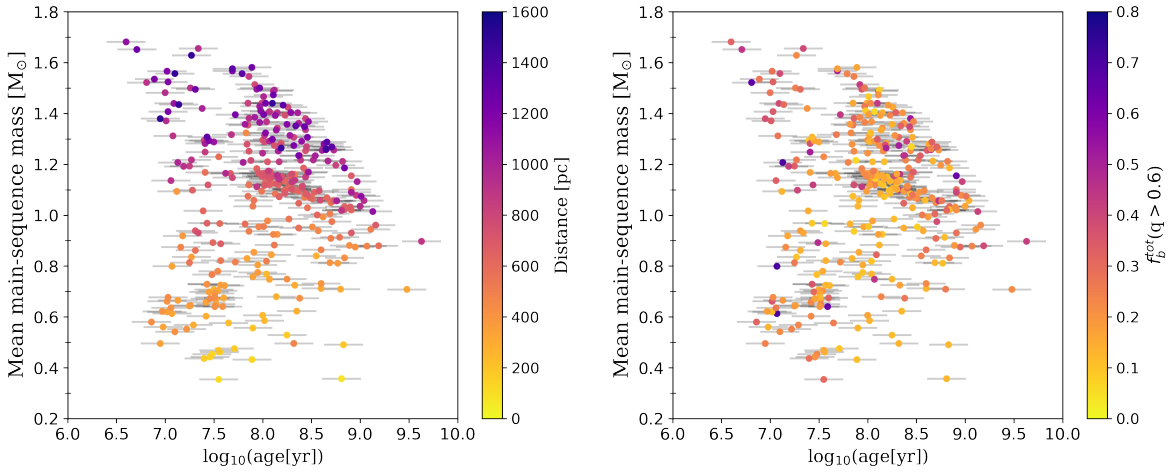


FIGURE 21: Mean main-sequence mass of the open cluster as a function of the decimal logarithm of its age in years for the 324 open clusters in our catalogue, colour-coded by their distance (left) and by their total binary fraction of systems with $q > 0.6$ (right).

7.5 Comparison to field stars

In OCs, the region between the single-star main sequence (SS) and the secondary sequence (BS) is not very densely populated. This could be due to the fact that the mass-ratio distribution is biased towards equal masses, although this is not necessarily the case because the BS is not only composed of equal-mass binaries but has also contribution from a range of high q. For field binaries, however, a clear peak in the q distribution near unity has been reported (Fisher, Schröder, and Smith, 2005), which seems to be present at small periods and it may even decrease with mass (Söderhjelm, 2007). Nevertheless, field binaries are not necessarily representative of OC binaries here studied. Cluster binaries are in the process of diffusing from the OC at a significant rate over timescales $\sim 10^8$ yr (Fisher, Schröder, and Smith, 2005), and without a better understanding of their ejection processes, the overall f_b of OCs and field stars cannot be easily compared (Thompson et al., 2021); and such comparison is out of the scope of this work.

A recent review on this topic is Offner et al. (2022), which states that for close and intermediate separations the f_b of FGK MS stars in young OCs are consistent with their field MS counterparts; and the close f_b of old OCs also match the field values, though the cluster cores have an excess of binaries due to mass segregation. Regarding wide binaries in OCs, *Gaia* observations reveal a deficit of them especially in the OCs with higher stellar densities, likely due to dynamical disruptions.

Chapter 8

Conclusions

The study of the binary fraction in open clusters 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. We have studied 324 OCs, 228 from Tarricq et al. (2022) and 96 from Cantat-Gaudin et al. (2020) catalogues, retaining only their members having a membership probability larger than 0.7.

The study of the f_b in OCs is usually restricted to the unresolved binaries. Their identification can be done individually mainly through two methods. The first is the detection of photometric binaries, either because they are eclipsing binaries or using more sophisticated analysis like comparing observed magnitudes from multiple photometric filters to synthetic star spectral energy distributions (Thompson et al., 2021).The second method is through radial velocities. But, given the number of stars in each OC and the velocity precision necessary to detect most binary systems, these spectroscopic surveys can take many years to complete. Such individual studies are not feasible, therefore, to be carried out for the majority of OC members, nor for large OC samples.

This is why the CMD, only requiring imaging in two filters, is 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) and 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 Niu, Wang, and Fu (2020) and `ASteCA`. Still, these methods rely on isochrone fitting and treat the rest of parameters as nuisance parameters. Finally, Li and Shao (2022) fitted all cluster parameters simultaneously in a fully probabilistic and selfconsistent way, modelling the OC's observed CMD as a mixture distribution of single stars, unresolved binaries, and field stars.

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. We are not concerned with estimating all the OC's fundamental parameters, only f_b, which is estimated as the proportion of the BS Gaussian area (counts) to the SS one, without need for estimating the q of each unresolved binary system. But we do need to characterize the q range our BS Gaussian (and, thus, f_b) takes into account, which by construction involves the higher q values down to $q_{\rm lim}$. This value is estimated applying the method to a reduced representative sample of simulated CMDs.

This can be successfully done with our MCMC implementation, which is applicable to 371 OCs out of the 374 that verify the conditions on the number of MS selected members and MS width. Through visual inspection, it is found to work properly for the final studied sample of 324 OCs, implying a high percentage (87%) of successful applications. It yields values of $f_b(q > 0.6^{+0.05}_{-0.15})$ in the range $[0.04, 0.62]$ with a median nominal uncertainty of 0.04 (not taking into account systematics). This method is largely preferred over the GMM implementation, which yields compatible f_b estimations for the common observed OCs, because the latter is restricted to a sample four times smaller of OCs of lower f_b and for which $q_{\rm lim}$ cannot be estimated due to the different performance of GMM in the observed and simulated CMDs.

The advantages of our modelling are that it does not rely on theoretical isochrone fitting or any prescription on the binaries, it just needs a proper fitting of the SS. The versatility of this polynomial functionality fitted through MCMC enables to study OCs with varied characteristics and degrees of differential extinction, thus carrying out a homogenous study of a large sample. The use of simulations, besides estimating $q_{\rm lim}$, allows us to estimate the total binary fraction $f_b^{tot}(q > q_{\rm lim})$, accounting for the usually neglected small portion of resolved binaries.

The main drawbacks are that all the OCs' selected MS members are regarded as either simple or binary systems, instead of acknowledging the possible presence of field star contamination or higher-order multiple systems; and also that the uncertainties intrinsic to the modelling are difficult to quantify. The uncertainty percentiles of the estimated $q_{\rm lim}^{\rm MCMC,obs} = 0.6^{+0.05}_{-0.15}$ could also be considered too large for some purposes.

Our unresolved f_b estimations are coherent with the values of Niu, Wang, and Fu (2020). Our results are harder to be compared to other studies, which do not specify the considered mass and/or q range. Estimating the f_b for some of our studied observed OCs with an alternative automated analysis, using `ASteCA`, we find degenerated solutions with larger values and uncertainties than ours. However, using the new version of this code we might be able to perform a more accurate study, improving the isochrone fitting and binaries' prescriptions. Overall, the f_b estimation is by no means trivial, and it depends inevitably on the modelling.

We find an increase of the total f_b with the mass of the primary star, in agreement with observational evidence. No correlation appears between its value and the position in or perpendicular to the Galactic plane. There is a great dispersion in the total f_b dependence on the OC age, although the trend of the running median is to decrease with age until approximately $\log_{10}(\text{age[yr]}) \in [8.0, 8.5]$, what could be explained by binary disruption through gravitational interactions, but for older ages this trend increases. As future work, we could assign a probability to each observed object to belong to one Gaussian distribution or the other, and establish a probability threshold to assign them to one. This would facilitate the study of f_b dependence on the spectral type and the distance to the OC's centre (to study whether there is mass segregation). Our assumption of the distributions being Gaussian is not physically motivated, but a first-order approach. Nonetheless, within its limitations, our modelling and its MCMC implementation can be overall considered a valid approach to homogeneously estimate the unresolved binary fraction of high mass ratio for a sample of open clusters. The catalogue we provide is the largest of this kind, and could help in providing insight into stellar and open cluster evolution.

Appendix: ASteCA analysis

Here we summarize the results of the f_b estimation for some of the OCs in our sample through one of the methods of automatic processing of data developed recently to estimate the fundamental parameters of OCs, to evaluate how our developed methods compare to it. In particular, we apply ASteCA (Automated Stellar Cluster Analysis; Perren, G. I., Vázquez, R. A., and Piatti, A. E., 2015), an open source code that determines the fundamental parameters for each OC via CMD analysis, using ptemcee parallel tempering Bayesian inference algorithm (Vousden, Farr, and Mandel, 2016) to look for the synthetic cluster generated from theoretical PARSEC isochrones which fits it better (as defined by a Poisson likelihood ratio). As these isochrones only comprise simple systems, ASteCA implements the prescriptions for the generation of binary systems in the synthetic clusters. In the version we use (0.4.2), the distribution of binary systems as a function of the primary mass is uniform; and the assignation of the secondary mass follows also a uniform distribution in $q \in [q_{min}, 1]$ (q_{min} is an input parameter). Such distributions to generate binary systems differ from the ones implemented by GOG (Sect. 5.1), and also from observations.

We applied ASteCA to the 84 observed OCs in common for both the GMM and MCMC methods, and their respective simulations, and obtained the mean, median and mode of the f_b, as well as its 16^{th} and 84^{th} percentiles.

For the observed OCs, each of these three estimations is compatible within 1σ computing it with only our MS selected members or with all the OC members given by the catalogues. Figure 22 shows the large differences between these three estimations. Moreover, for 78/84 OCs (93%) the rest of the 84^{th} f_b percentile minus the 16^{th} is larger than 0.3; and for 73 OCs the nominal error is greater than 0.9 times the f_b median. This reflects that the f_b marginalized posterior probability distribution does not present a peak, it has not properly converged, so that its large uncertainty range implies an important degeneration in the solution. In the bottom panel of Fig. 22 the distribution of ASteCA's f_b estimations for the 84 OCs are compared with our inferred f_b. All of them (ASteCA's f_b mean, median and mode) are larger than our value for all OCs, and, for instance, ASteCA's f_b mean is compatible with our GMM-inferred one within 1σ for only 14/84 OCs (17%). Given the large uncertainties of ASteCA's f_b, however, their values are not very informative in any case.

Applying ASteCA to the 84 simulated OCs, we find again the same discrepancy between the three f_b estimators (equivalent behaviour as in Fig. 22) and very large uncertainties. For these simulated CMDs having the same number of MS members as the observed CMDs and comprising the same MS range, given the large ASteCA uncertainties we find our GMM-inferred f_b and the f_b mean yield by ASteCA to be compatible within 1σ for 62% of OCs. However, the poorly constrained values of ASteCA are still not very informative.

ASteCA's isochrone fitting process becomes considerably more difficult if the red clump (RC) and/or MSTO are missing, since they are both indicative of the evolutionary stage and play an important role in breaking the degeneracy of the solutions. This could explain the degeneracy of the solutions when applying ASteCA to the selected members of the observed OCs and their respective simulated CMDs. But these results for the observed OCs are found to be compatible with the ones obtained

considering all their members, without performing our MS selection. The problem could be, therefore, that both the MSTO and RC happen to be the least populated parts of the CMDs.

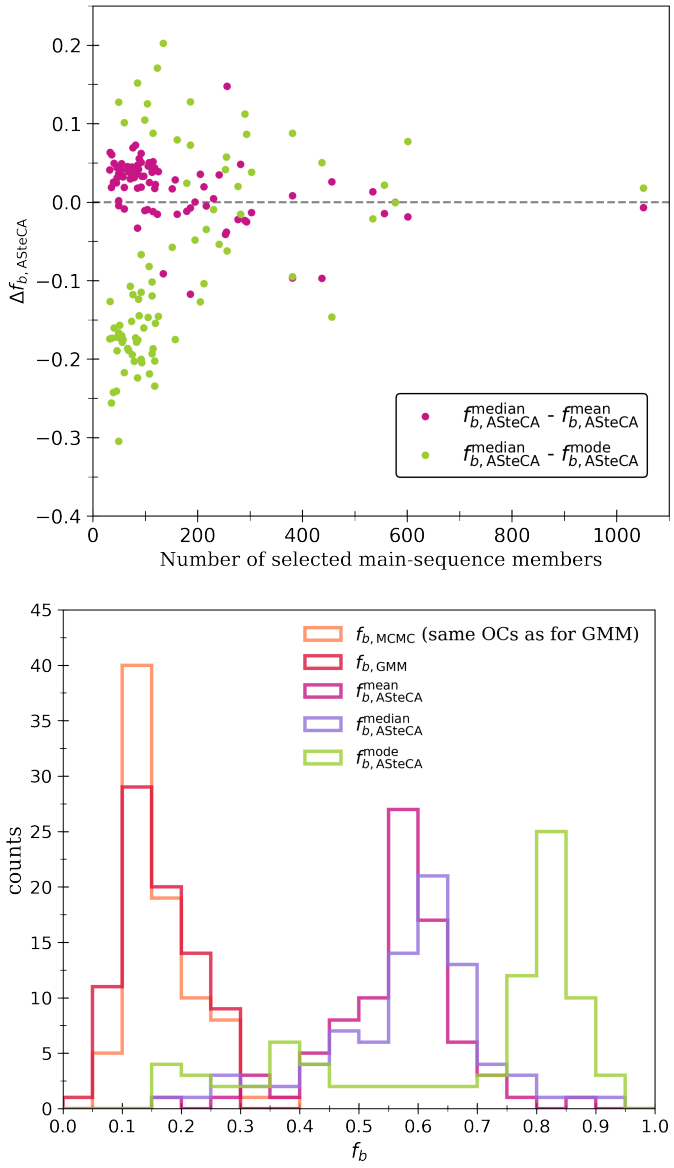


FIGURE 22: *Top*: Differences between ASteCA's median and mean of f_b (magenta) and between ASteCA's median and mode of f_b (green), as a function of the number of selected main-sequence members of each of the 84 observed open clusters. *Bottom*: Histograms of the f_b for the 84 observed open clusters estimated through MCMC (orange), GMM (red), ASteCA's mean (magenta), ASteCA's median (purple) and ASteCA's mode (green).

At the time of this writing, a new ASteCA version (0.4.3) has been released, which includes a large update on the synthetic cluster generation process, involving major improvements on the prescriptions for the generation of binary systems. The application of this new ASteCA version to our sample of OCs, both observed and simulated, as well as the study of how to improve the isochrone fitting in order to obtain better constrained f_b, are deferred to future work.

Bibliography

Agresti, Alan and Brent A. Coull (1998). "Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions". In: *The American Statistician* 52.2, pp. 119–126. DOI: 10.1080/00031305.1998.10480550. eprint: https://doi.org/10.1080/00031305.1998.10480550. URL: https://doi.org/10.1080/00031305.1998.10480550.

Arenou, F. (July 2011). "Simulating multiple stars in preparation for Gaia". In: *International Workshop on Double and Multiple Stars: Dynamics, Physics, and Instrumentation*. Ed. by J. A. Docobo, V. S. Tamazian, and Y. Y. Balega. Vol. 1346. American Institute of Physics Conference Series, pp. 107–121. DOI: 10.1063/1.3597593.

Aros, Francisco I. et al. (Dec. 2021). "Using binaries in globular clusters to catch sight of intermediate-mass black holes". In: *MNRAS* 508.3, pp. 4385–4398. DOI: 10.1093/mnras/stab2872. arXiv: 2110.00590 [astro-ph.GA].

Banerjee, Sambaran (Jan. 2022). "Merger rate density of stellar-mass binary black holes from young massive clusters, open clusters, and isolated binaries: Comparisons with LIGO-Virgo-KAGRA results". In: *Phys. Rev. D* 105.2, 023004, p. 023004. DOI: 10.1103/PhysRevD.105.023004. arXiv: 2108.04250 [astro-ph.HE].

Belloni, Diogo et al. (Nov. 2017). "On the initial binary population for star cluster simulations". In: *MNRAS* 471.3, pp. 2812–2828. DOI: 10.1093/mnras/stx1763. arXiv: 1707.04271 [astro-ph.GA].

Borodina, Olga I. et al. (Apr. 2019). "Unresolved Binaries and Galactic Clusters' Mass Estimates". In: *ApJ* 874.2, 127, p. 127. DOI: 10.3847/1538-4357/ab08e5. arXiv: 1902.10443 [astro-ph.SR].

Bressan, Alessandro et al. (Nov. 2012). "PARSEC: stellar tracks and isochrones with the PAdova and TRieste Stellar Evolution Code". In: *MNRAS* 427.1, pp. 127–145. DOI: 10.1111/j.1365-2966.2012.21948.x. arXiv: 1208.4498 [astro-ph.SR].

Cantat-Gaudin, T. and F. Anders (Jan. 2020). "Clusters and mirages: cataloguing stellar aggregates in the Milky Way". In: *A&A* 633, A99, A99. DOI: 10.1051/0004-6361/201936691. arXiv: 1911.07075 [astro-ph.SR].

Cantat-Gaudin, T. et al. (Aug. 2020). "Painting a portrait of the Galactic disc with its stellar clusters". In: *A&A* 640, A1, A1. DOI: 10.1051/0004-6361/202038192. arXiv: 2004.07274 [astro-ph.GA].

Castro-Ginard, A. et al. (Mar. 2020). "Hunting for open clusters in Gaia DR2: 582 new open clusters in the Galactic disc". In: *A&A* 635, A45, A45. DOI: 10.1051/0004-6361/201937386. arXiv: 2001.07122 [astro-ph.GA].

Cordoni, G. et al. (Dec. 2018). "Extended Main-sequence Turnoff as a Common Feature of Milky Way Open Clusters". In: *ApJ* 869.2, 139, p. 139. DOI: 10.3847/1538-4357/aaedc1. arXiv: 1811.01192 [astro-ph.SR].

Dal Tio, Piero et al. (Oct. 2021). "Dissecting the Gaia HR diagram within 200 pc". In: *MNRAS* 506.4, pp. 5681–5697. DOI: 10.1093/mnras/stab1964. arXiv: 2107.01844 [astro-ph.SR].

de Grijs, Richard et al. (Mar. 2013). "Gravitational Conundrum? Dynamical Mass Segregation versus Disruption of Binary Stars in Dense Stellar Systems". In: *ApJ* 765.1, 4, p. 4. DOI: 10.1088/0004-637X/765/1/4. arXiv: 1301.1926 [astro-ph.SR].

Donada, J. (July 2021). "Binary fraction estimation in open clusters with *Gaia*". In: *http://diposit.ub.edu/dspace/handle/2445/180727*.

Duchêne, Gaspard and Adam Kraus (Aug. 2013). "Stellar Multiplicity". In: *ARA&A* 51.1, pp. 269–310. DOI: 10.1146/annurev-astro-081710-102602. arXiv: 1303.3028 [astro-ph.SR].

Duquennoy, A. and M. Mayor (Aug. 1991). "Multiplicity among Solar Type Stars in the Solar Neighbourhood - Part Two - Distribution of the Orbital Elements in an Unbiased Sample". In: *A&A* 248, p. 485.

Evans, D. W. et al. (Aug. 2018). "Gaia Data Release 2. Photometric content and validation". In: *A&A* 616, A4, A4. DOI: 10.1051/0004-6361/201832756. arXiv: 1804.09368 [astro-ph.IM].

Fabricius, C. et al. (May 2021). "Gaia Early Data Release 3. Catalogue validation". In: *A&A* 649, A5, A5. DOI: 10.1051/0004-6361/202039834. arXiv: 2012.06242 [astro-ph.GA].

Fisher, James, Klaus-Peter Schröder, and Robert Connon Smith (Aug. 2005). "What a local sample of spectroscopic binaries can tell us about the field binary population". In: *MNRAS* 361.2, pp. 495–503. DOI: 10.1111/j.1365-2966.2005.09193.x. arXiv: astro-ph/0508651 [astro-ph].

Foreman-Mackey, Daniel et al. (Mar. 2013). "emcee: The MCMC Hammer". In: *PASP* 125.925, p. 306. DOI: 10.1086/670067. arXiv: 1202.3665 [astro-ph.IM].

Fuhrmann, K. et al. (Feb. 2017). "Multiplicity among Solar-type Stars". In: *ApJ* 836.1, 139, p. 139. DOI: 10.3847/1538-4357/836/1/139.

Gieles, M., H. Sana, and S. F. Portegies Zwart (Mar. 2010). "On the velocity dispersion of young star clusters: super-virial or binaries?" In: *MNRAS* 402.3, pp. 1750–1757. DOI: 10.1111/j.1365-2966.2009.15993.x. arXiv: 0911.1557 [astro-ph.SR].

González, Elena et al. (Feb. 2021). "Intermediate-mass Black Holes from High Massive-star Binary Fractions in Young Star Clusters". In: *ApJ* 908.2, L29, p. L29. DOI: 10.3847/2041-8213/abdf5b. arXiv: 2012.10497 [astro-ph.HE].

Hogg, David W., Jo Bovy, and Dustin Lang (Aug. 2010). "Data analysis recipes: Fitting a model to data". In: *arXiv e-prints*, arXiv:1008.4686, arXiv:1008.4686. arXiv: 1008.4686 [astro-ph.IM].

Huber, Peter J. (1964). "Robust Estimation of a Location Parameter". In: *The Annals of Mathematical Statistics* 35.1, pp. 73 –101. DOI: 10.1214/aoms/1177703732. URL: https://doi.org/10.1214/aoms/1177703732.

Hurley, Jarrod and Christopher A. Tout (Nov. 1998). "The binary second sequence in cluster colour-magnitude diagrams". In: *MNRAS* 300.4, pp. 977–980. DOI: 10.1046/j.1365-8711.1998.01981.x. arXiv: astro-ph/9807108 [astro-ph].

Jadhav, Vikrant V. et al. (Dec. 2021). "High Mass-Ratio Binary Population in Open Clusters: Segregation of Early Type Binaries and an Increasing Binary Fraction with Mass". In: *AJ* 162.6, 264, p. 264. DOI: 10.3847/1538-3881/ac2571. arXiv: 2109.03782 [astro-ph.SR].

Kaczmarek, T., C. Olczak, and S. Pfalzner (Apr. 2011). "Evolution of the binary population in young dense star clusters". In: *A&A* 528, A144, A144. DOI: 10.1051/0004-6361/201015233. arXiv: 1102.2055 [astro-ph.GA].

Kroupa, Pavel (Apr. 2001). "On the variation of the initial mass function". In: *MNRAS* 322.2, pp. 231–246. DOI: 10.1046/j.1365-8711.2001.04022.x. arXiv: astro-ph/0009005 [astro-ph].

— (Jan. 2002). "The Initial Mass Function of Stars: Evidence for Uniformity in Variable Systems". In: *Science* 295.5552, pp. 82–91. DOI: 10.1126/science.1067524. eprint: https://www.science.org/doi/pdf/10.1126/science.1067524. URL: https://www.science.org/doi/abs/10.1126/science.1067524.

Lamers, H. J. G. L. M. et al. (Oct. 2005). "An analytical description of the disruption of star clusters in tidal fields with an application to Galactic open clusters". In:

A&A 441.1, pp. 117–129. DOI: 10.1051/0004-6361:20042241. arXiv: astro-ph/0505558 [astro-ph].

Li, Lu and Zhengyi Shao (May 2022). "MiMO: Mixture Model for Open Clusters in Color-Magnitude Diagrams". In: *ApJ* 930.1, 44, p. 44. DOI: 10.3847/1538-4357/ac5f4f. arXiv: 2112.08028 [astro-ph.GA].

Li, Lu et al. (Sept. 2020). "Modeling Unresolved Binaries of Open Clusters in the Color-Magnitude Diagram. I. Method and Application of NGC 3532". In: *ApJ* 901.1, 49, p. 49. DOI: 10.3847/1538-4357/abaef3. arXiv: 2008.04684 [astro-ph.GA].

Li, Zhongmu et al. (Mar. 2022). "LISC Catalog of Star Clusters. I. Galactic Disk Clusters in Gaia EDR3". In: *ApJS* 259.1, 19, p. 19. DOI: 10.3847/1538-4365/ac3c49.

Luri, X. et al. (June 2014). "Overview and stellar statistics of the expected Gaia Catalogue using the Gaia Object Generator". In: *A&A* 566, A119, A119. DOI: 10.1051/0004-6361/201423636. arXiv: 1404.5861 [astro-ph.IM].

Maeder, A. (May 1974). "Stellar evolution near the main sequence: on some systematic differences between cluster sequences and model calculations." In: *A&A* 32, pp. 177–190.

Marigo, Paola et al. (Jan. 2017). "A New Generation of PARSEC-COLIBRI Stellar Isochrones Including the TP-AGB Phase". In: *ApJ* 835.1, 77, p. 77. DOI: 10.3847/1538-4357/835/1/77. arXiv: 1701.08510 [astro-ph.SR].

Moe, Maxwell and Rosanne Di Stefano (2017). "Mind Your Ps and Qs: The Interrelation between Period Distributions of Binary Stars". In: *The Astrophysical Journal Supplement Series* 230.2, p. 15. DOI: 10.3847/1538-4365/aa6fb6. URL: https://doi.org/10.3847/1538-4365/aa6fb6.

Niu, Hubiao, Jiaxin Wang, and Jianning Fu (Nov. 2020). "Binary Fraction Estimation of Main-sequence Stars in 12 Open Clusters: Based on the Homogeneous Data of LAMOST Survey and Gaia DR2". In: *ApJ* 903.2, 93, p. 93. DOI: 10.3847/1538-4357/abb8d6.

Offner, Stella S. R. et al. (Mar. 2022). "The Origin and Evolution of Multiple Star Systems". In: *arXiv e-prints*, arXiv:2203.10066, arXiv:2203.10066. arXiv: 2203.10066 [astro-ph.SR].

Perren, G. I., Vázquez, R. A., and Piatti, A. E. (Apr. 2015). "ASteCA: Automated Stellar Cluster Analysis". In: *A&A* 576, A6. DOI: 10.1051/0004-6361/201424946. URL: http://dx.doi.org/10.1051/0004-6361/201424946.

Pinfield, D. J. et al. (July 2003). "Brown dwarfs and low-mass stars in the Pleiades and Praesepe: membership and binarity". In: *MNRAS* 342.4, pp. 1241–1259. DOI: 10.1046/j.1365-8711.2003.06630.x. arXiv: astro-ph/0303600 [astro-ph].

Rastello, Sara, Giovanni Carraro, and Roberto Capuzzo-Dolcetta (June 2020). "Effect of Binarity in Star Cluster Dynamical Mass Determination". In: *ApJ* 896.2, 152, p. 152. DOI: 10.3847/1538-4357/ab910b. arXiv: 2005.03037 [astro-ph.SR].

Reid, M. J. et al. (Nov. 2019). "Trigonometric Parallaxes of High-mass Star-forming Regions: Our View of the Milky Way". In: *ApJ* 885.2, 131, p. 131. DOI: 10.3847/1538-4357/ab4a11. arXiv: 1910.03357 [astro-ph.GA].

Riello, M. et al. (May 2021). "Gaia Early Data Release 3. Photometric content and validation". In: *A&A* 649, A3, A3. DOI: 10.1051/0004-6361/202039587. arXiv: 2012.01916 [astro-ph.IM].

Robin, A. C. et al. (July 2012). "Gaia Universe model snapshot. A statistical analysis of the expected contents of the Gaia catalogue". In: *A&A* 543, A100, A100. DOI: 10.1051/0004-6361/201118646. arXiv: 1202.0132 [astro-ph.GA].

Söderhjelm, S. (Feb. 2007). "The q = 1 peak in the mass-ratios for Hipparcos visual binaries". In: *A&A* 463.2, pp. 683–691. DOI: 10.1051/0004-6361:20066024.

Sterzik, M. F. and R. H. Durisen (Aug. 2004). "Are Binary Separations related to their System Mass?" In: *Revista Mexicana de Astronomia y Astrofisica Conference Series*. Ed. by Christine Allen and Colin Scarfe. Vol. 21. Revista Mexicana de Astronomia y Astrofisica Conference Series, pp. 58–62.

Tarricq, Y. et al. (Mar. 2022). "Structural parameters of 389 local open clusters". In: *A&A* 659, A59, A59. DOI: 10.1051/0004-6361/202142186. arXiv: 2111.05291 [astro-ph.GA].

Thies, Ingo and Pavel Kroupa (Dec. 2007). "A Discontinuity in the Low-Mass Initial Mass Function". In: *ApJ* 671.1, pp. 767–780. DOI: 10.1086/522512. arXiv: 0708.1764 [astro-ph].

Thompson, Benjamin A. et al. (Apr. 2021). "The Binary INformation from Open Clusters Using SEDs (BINOCS) Project: Reliable Photometric Mass Determinations of Binary Star Systems in Clusters". In: *AJ* 161.4, 160, p. 160. DOI: 10.3847/1538-3881/abde4c. arXiv: 2101.07857 [astro-ph.SR].

Torres, Guillermo, David W. Latham, and Samuel N. Quinn (Nov. 2021). "Long-term Spectroscopic Survey of the Pleiades Cluster: The Binary Population". In: *ApJ* 921.2, 117, p. 117. DOI: 10.3847/1538-4357/ac1585. arXiv: 2107.10259 [astro-ph.SR].

Vousden, W. D., W. M. Farr, and I. Mandel (Jan. 2016). "Dynamic temperature selection for parallel tempering in Markov chain Monte Carlo simulations". In: *MNRAS* 455.2, pp. 1919–1937. DOI: 10.1093/mnras/stv2422. arXiv: 1501.05823 [astro-ph.IM].