UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA

Mass loss in Open Cluster Populations

Duarte Miguel de Almeida

Mestrado em Física
Especialização em Astrofísica e Cosmologia

Dissertação orientada por:
André Moitinho de Almeida

2022

Agradecimentos

Primeiro que tudo, tenho de agradecer aos meus pais pela oportunidade de prosseguir os meus estudos. Ambos apoiaram-me durante esta (e muitas outras) jornadas e estou extremamente grato pelo esforço que têm feito por mim.

Em seguida agradeço ao meu orientador, o professor André Moitinho, por me proporcionar um ambiente fantástico onde foi possível crescer não só enquanto estudante mas também enquanto pessoa. A sua paixão pela astronomia e pelo ensino é claramento visível e sinto que isso tornou esta experiência muito mais enriquecedora, tanto a nível profissional como pessoal. Estou extremamente grato por todo o apoio e todo o tempo que dedicou ao meu trabalho.

Quero também agradecer aos meus amigos que partilharam comigo a aventura de estar num mestrado, em particular à Lara Piscarreta, Maria Gonçalves, Constança Freire e Duarte Santos. Foram, sem dúvida, as pessoas que mais me deram ânimo e força durante estes meses. Quero também agradecer às pessoas que deram um contributo essencial para esta tese, André Silva e Sandro Moreira, cujo trabalho serviu de base para o código utilizado numa parte deste trabalho e ao Gonçalo Fernandes pelos bons momentos passados na organização do ENAA e no gabinete.

Finalmente, agradeço também aos meus amigos André Mendes, David, Elias, Duran, Nico, Bernardo e Nota a quem carinhosamente chamo "família" que sempre me apoiaram e se preocuparam comigo. Agradeço também a André Velez, por todo o seu carinho e apoio, que foi muito importante para mim. Agradeço ao Rodrigo pelos comentários criticos ao texto e pela amizade. Por fim, agradeço também a tantos outros que, de alguma forma, me ajudaram durante este trabalho.

Resumo

Crê-se que a maior parte das estrelas é formada em grupos (Lada and Lada, 1991) que, mais tarde, se dissolvem e cujas estrelas passam a integrar a população de campo. Este processo de dissolução determina a distribuição de massa e idade dos grupos de estrelas observadas e, também, as propriedades da população de campo. Neste contexto, determinar a distribuição de massa e idade dos aglomerados na nossa galáxia irá impor constrangimentos sobre os modelos teóricos de dissolução e aumentar o nosso conhecimento sobre os processos de perda de massa dos grupos de estrelas.

Estes grupos de estrelas formam-se devido ao colapso gravitacional nas regiões mais densas das nuvens moleculares gigantes que se encontram no disco galático. Como as estrelas de cada grupo formam-se ao mesmo tempo, essas estrelas partilham a mesma idade, distância e composição química que faz delas objetos ideais para estudar não só a formação e evolução estelar, como também a estrutura espacial da nossa galáxia.

A massa é o elemento que governa a dinâmica interna e, também, a interação com o campo de maré da galáxia, pelo que é de extrema importância saber qual a distribuição de massa dos aglomerados na nossa galáxia. Contudo, um catálogo de massas homogéneo e de larga escala ainda não está disponível, pelo que um dos objetivos principais deste trabalho é a determinação das massas para aglomerados abertos com fotometria de elevada precisão que agora é possível graças à missão *Gaia*.

A nossa amostra vem do catálogo de Dias et al. (2021) que contém a idade e distância de aglomerados abertos, determinadas a partir de dados da missão *Gaia* (*Gaia* Data Release 2, Gaia Collaboration et al. (2018)). Os dados da missão *Gaia* fornecem-nos movimentos próprios, posições e fotometria (no visível) para cada estrela. Esta missão revolucionou o campo da astronomia ao trazer dados de elevada precisão para mais de 1 bilião de estrelas. Isto tem permitido estudar a estrutura e dinâmica da Via Láctea com um nível de detalhe nunca antes visto. No total, a amostra proveniente do catálogo de Dias et al. (2021) é composta por 1743 aglomerados abertos que têm idades entre alguns milhões a alguns biliões de anos e distâncias até alguns milhares de parsecs à volta do Sol.

Considerando que o limite de cada aglomerado pode ser definido como o raio até ao qual as estrelas ainda estão gravitacionalmente ligadas a este, por vezes chamado de "raio de maré", foi necessário começar pela determinação do raio de maré de cada aglomerado. Este é o raio onde a força de maré da galáxia onde o aglomerado se encontra é igual a atração gravitacional do próprio aglomerado. As estrelas fora do raio de maré do aglomerado estão a dissipar-se para a população de campo e já não devem ser consideradas parte deste. Cada aglomerado possui um perfil de densidade cuja parametrização mais usada é o perfil de King (King, 1962) que nos permite obter o raio de maré e o raio do núcleo de cada aglomerado. Os nossos resultados mostram um raio de maré mediano de 10 parsec (excludindo os resultados de má qualidade) e um raio do núcleo mediano de 2 parsec que é compatível com os valores encontrados na literatura (Piskunov et al., 2007; Kharchenko et al., 2013). Contudo, o ajuste do perfil de King apresentou algumas dificuldades que se traduziram em incertezas elevadas para o raio de maré embora as incertezas medianas para o raio do núcleo sejam da ordem de 20%.

Através de modelos de evolução estelar e usando a Função de Massa Inicial, que nos dá o número de estrelas de cada massa que se formam inicialmente num aglomerado, é possível saber o número de estrelas que um aglomerado com uma certa massa tem de cada magnitude. Isto constitui a função de luminosidade que varia de acordo com a idade de cada aglomerado. Através da comparação entre a distribuição de luminosidade observada e a função de luminosidade teórica, obtida através dos modelos de evolução estelar de Padova, podemos determinar a massa luminosa. Os nossos resultados mostram uma distribuição logarítmica de massa que apresenta uma forma gaussiana com valor médio 2.7 dex e desvio-padrão de 0.4 dex. A distribuição linear tem uma massa média de 797 $M_{\odot}$. Os erros associados ao nosso método de determinação da massa foram obtidos através de *bootstrapping* e são cerca de 4%. Para avaliar o efeito das incertezas elevadas no raio de maré, determinámos a massa usando o valor máximo e mínimo do raio de maré e comprovámos que o efeito na determinação da massa não é significativo.

Uma vez que nem todos os resultados são de boa qualidade, classificámos cada aglomerado com base na qualidade da determinação do raio de maré, massa e ajuste da isócrona no diagrama cor-magnitude e selecionámos uma amostra com os resultados de qualidade elevada e intermédia.

Para podermos caraterizar não só a forma como as estrelas passam de um ambiente em que estão aglomeradas para uma população de campo bastante uniforme, como também estudar o campo de maré da nossa galáxia, é necessário entender o processo de disrupção que atua sobre os aglomerados. Para determinar os parâmetros relacionados com a disrupção dos aglomerados, simulámos a formação e evolução de uma população de aglomerados abertos que perde massa ao longo do tempo de acordo com as equações obtidas por Lamers et al. (2005b) e comparámos a distribuição simulada de idades e massas com as observações. No nosso modelo, assumimos uma taxa de formação constante e uma função de massa inicial exponencial, comumente utlizada na literatura, obtida usando os aglomerados imersos nas nuvens moleculares onde são formados (Gieles, 2009).

Percorrendo várias combinações para os parâmetros do nosso modelo de disrupção e, comparando as simulações com as observações, foi possível obter os mesmos valores que existem na literatura que indicam uma escala de tempo de disrupção de cerca de 5 milhões de anos (Lamers et al., 2004; Boutloukos and Lamers, 2003). Obtivémos também uma zona adicional, que parece estar degenerada, onde são favorecidas escalas de tempo de disrupção mais pequenas. Estas zonas verificam-se também quando comparamos as simulações com as observações vindas de uma amostra com os aglomerados que apresentam apenas resultados de qualidade elevada e também para uma amostra de menor dimensão (contendo apenas aglomerados num raio de 1500 parsec à volta do Sol). Apesar de existir compatibilidade entre as observações e as simulações para a distribuição de idades usando os valores da literatura, o mesmo não acontece para a distribuição de massas cujo pico das massas simuladas se verifica para valores de massa mais pequenos que o pico das massas observadas. Esta incompatibilidade na distribuição das massas era inesperada e verifica-se para todas as combinações de parâmetros consideradas.

Uma vez que o nosso modelo apenas considera a evolução dos aglomerados após sairem das nuvens moleculares, a distribuição de massa inicial usada deverá refletir a distribuição de massa dos aglomerados jovens. Para que a distribuição de massa simulada se aproximasse das observações, testámos uma função de massa inicial modificada obtida através do ajuste de uma gaussiana à distribuição logarítmica de massa dos aglomerados jovens (com idades inferiores a 50 milhões de anos). Embora este ajuste tenha sido feito apenas como um teste heurístico, os resultados mostram que o pico da distribuição das massas simuladas se aproxima do pico das observações, mantendo o bom ajuste na distribuição das idades. Isto indica que a função de massa inicial utilizada na literatura, obtida a partir dos aglomerados imersos, poderá não ser adequada para os aglomerados expostos.

Este trabalho constitui o primeiro estudo onde é possível comparar, para uma amostra de grande dimensão, as distribuições de idade e, pela primeira vez, massa observadas na vizinhança solar com o resultado de simulações da formação e evolução de aglomerados de forma a impor constrangimentos nos modelos de disrupção de aglomerados abertos.

Palavras chave: Galáxia: evolução, Galáxia: conteúdo estelar, Aglomerados abertos e associações: geral, *Surveys*: Gaia

Abstract

Mass is the main quantity driving the formation, structure, and evolution of stars but it also governs the dynamics of stellar systems such as open clusters (OCs) that provide crucial information about the dynamical evolution of the Galactic disc where they are formed.

Recently, several large-scale OC studies have been published (Cantat-Gaudin et al. (2018), Monteiro et al. (2020) and others). However, high quality and systematic mass determinations for those OCs are not available so, in our work, we performed homogeneous determinations of luminous mass for 1724 OCs from the catalogue of Milky Way OCs of Dias et al. (2021). We also determined the tidal and core radii by fitting the density profile of each cluster with the King profile (King, 1962) which was required to determine the mass of each cluster.

Using the resulting mass distribution, we attempted to constrain the disruption experienced by clusters in the solar neighbourhood by simulating the build up and mass evolution of a population of OCs using the equations derived by Lamers et al. (2005a). Comparing the simulated age and, for the first time, mass distributions to the observations, we recover the same parameter values related to disruption obtained in the literature with an additional optimal area for smaller disruption timescales. However, despite the reasonable agreement for the age distribution, the simulations do not generate a mass distribution similar to the observations. To improve the simulated mass distribution, we tested a different Cluster Initial Mass Function (CIMF) obtained from the mass distribution of clusters with age under 50 million years (Myr), which led to an improvement of the agreement with the observed mass distribution. This suggests that the CIMF of non-embedded clusters might be different from the one universally used, which is determined using embedded clusters.

Keywords: Galaxy: evolution, Galaxy: stellar content, open clusters and associations: general, Surveys: Gaia

Index

List of Figures

List of Tables

Glossary

2MASS Two Micron All Sky Survey.

ASCC All-Sky Compiled Catalogue.

CIMF Cluster Initial Mass Function.

CMD Color-magnitude diagram.

CPU Central Processing Unit.

DR Data Release.

GMC Giant Molecular Cloud.

Gyr billion years.

HR Hertzsprung–Russell.

JWST James Webb Space Telescope.

LF Luminosity Function.

MCMC Markov chain Monte Carlo.

MWSC Milky Way Star Cluster catalog.

Myr million years.

OC Open Cluster.

RMS root mean square.

VVV VISTA Variables in the Via Lactea.

Chapter 1

Introduction

1.1 Star clusters

Most stars are believed to be born in groups (Lada and Lada, 1991) which later dissolve into the field population of their host galaxy. Thus, the dissolution process drives the mass and age distribution of the observed star cluster population as well as the properties of the field population. In this context, determining the mass and age distribution of star clusters in the Milky Way is an essential observational foundation for constraining models of star cluster evolution and shedding light on the dissolution processes they experience.

Stellar groups have been noticed since ancient times and have been object of interest for many theoreticians over time including early astronomers. One example is Galileo Galilei that systematically recorded and catalogued stellar aggregates in 1610 (D'Onofrio and Burigana, 2009). In Galileo's words:

"Est enim GALAXIA nihil aliud, quam innumerarum Stellarum coacervatim consitarum congeries: in quamcumque enim regionem illius Perspicillum dirigas, statim Stellarum ingens frequentia sese in conspectum profert, quarum complures satis magnæ ac valde conspicuæ videntur; sed exiguarum multitudo prorsus inexplorabilis est."

"The Galaxy is nothing else but a mass of innumerable stars planted together in clusters. Upon whatever part of it you direct the telescope straightway a vast crowd of stars presents itself to view; many of them are tolerably large and extremely bright, but the number of small ones is quite beyond determination" - English translation of the Sidereus Nuncius by Edward Stafford Carlos in Project Gutenberg (Carlos, 2014).

These groups of stars are formed due to gravitational collapse in the densest regions of Giant Molecular Clouds (GMCs), found in the galactic disc. In the first 2-3 million years, these groups are still embedded in the cloud (Lada and Lada, 2003) but, after that, the gas is dispersed due to stellar winds, radiation pressure or shock waves caused by the death of the most massive stars within the group, so they leave the GMC and become exposed. According to Lada and Lada (2003), there are very few groups of stars (less than 10%) that can survive the process of emerging from the parent cloud, yielding a very high early disruption rate (sometimes referred to as "infant mortality" rate).

There are three main types of stellar groups. There are the stellar associations that are loose groups of stars, not gravitationally bounded that drift apart from each other over time. There are also open clusters (OCs) where the stars are gravitationally bounded to each other. OCs are groups of a few hundred to a few thousand young stars, that may have irregular shapes, but, in general, have a more concentrated core that is surrounded by a less dense halo. This type of cluster is usually found in the Galactic disc and

has been extensively used in characterizing the structure of our galaxy (Moitinho et al., 2006; Vázquez et al., 2008; Bobylev and Bajkova, 2014; Cantat-Gaudin et al., 2020). As they tend to dissolve before most of their stars reach older ages, stars in OCs are usually young and cover a high span in brightness and masses.

Finally, there are the globular clusters which contain a much larger number of stars, in the order of $10^4 - 10^6$ stars. Their structure is characterized by a high-density core and a less dense halo. This type of clusters lasts longer than OCs and, in the Milky Way, they have ages over 10 Gyrs (e.g. Harris (1996)). They belong mostly to the galactic halo and some to the bulge, and can be found at all galactic latitudes.

Figure 1.1: Optical images of IC 142 stellar association, part of Triangulum Galaxy (left); open cluster M45 (middle) and Globular Cluster NGC 1466 (right). Credits: ESA/Hubble

As the stars of each cluster are formed in the same region in the Giant Molecular Cloud, they share the same age, distance and chemical composition, but differ by mass. In the Hertzsprung-Russell (HR) diagram, which is a scatter plot of stellar luminosities versus their effective temperatures, stars from an open cluster align along a distinctive sequence. That sequence depends on the mass distribution and age of the cluster. It can be fitted by age-dependent theoretical models of stellar evolution (called isochrones) to determine the age of the cluster. In Figure 1.2, we present the observational photometric version of the HR diagram which is the color-magnitude diagram (CMD). In the CMD, the stars' apparent magnitude is plotted against their color (difference of magnitude between 2 photometric bands). The effect of distance and extinction - absorption of the stellar light by the interstellar medium- which move the isochrone in the vertical and diagonal directions (respectively) needs to be considered when fitting the isochrone. Since determining ages and distances is beyond the scope of this work, we refer to Dias et al. (2021), which is our reference catalogue, for the description of the fitting method used to determine the values used in this work. In Figure 1.2, we present an example of the CMD with fitted isochrone for log(age) = 7.8 for open cluster Alessi 5. For isolated stars, their position in the CMD is ambiguous which makes it difficult to compare to theoretical isochrones.

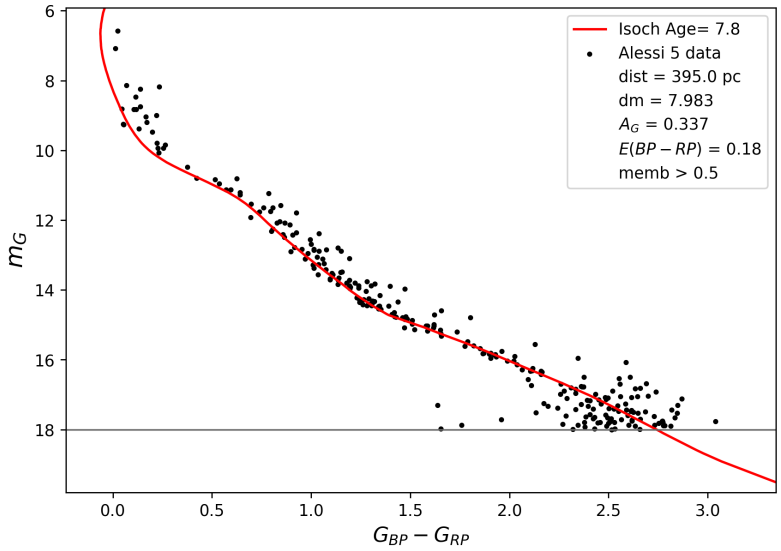

Figure 1.2: Color-Magnitude Diagram for open cluster Alessi 5 (Dias et al., 2021) with isochrone of log(age) = 7.8 overplotted for a distance of 395 pc and $A_G = 0.337$.

Despite the distinction between the 3 stellar groups, observations show that there is no "clear cut" when it comes to their classification as a specific type because groups that appear to be in the transition between these types have been observed. Recent detections have found clusters of young stars with hundreds of thousands of stars which is characteristic of globular clusters. One example is open cluster NGC 2070 in the Tarantula Nebula that contains about 500,000 young stars which is unusual for an open cluster. Figure 1.3 presents the image of NGC 2070 in the Tarantula Nebula produced recently by the James Webb Space Telescope (JWST).

Figure 1.3: Tarantula Nebula seen by James Webb's Near-Infrared Camera (NIRCam). Near the centre of the image open cluster NGC 2070 is shown. Credits: NASA, ESA, CSA, STScI, Webb ERO Production Team.

With the development of large-scale surveys from space (e.g. *Gaia*) as well as ground-based surveys such as 2MASS (Two Micron All Sky Survey) (Skrutskie et al., 2006), VVV (VISTA Variables in the Via Lactea) (Minniti et al., 2010), and others, more observations have been carried out which have brought a comprehensive view of star clusters. So far, about two thousand open clusters have been detected (Dias et al., 2002; Cantat-Gaudin, 2022) which are mostly inside a few kpc around the Sun (often called the solar neighbourhood) but that number will certainly increase in the future.

1.2 Mass of open clusters

Mass is the quantity that governs the internal dynamics of the clusters as well as its interactions with external galactic gravitational forces. Despite the wealth of theoretical studies of cluster evolution (Gieles et al. (2014), Alexander and Gieles (2012), Kruijssen et al. (2011), and others), there is a lack of observational results to make comparisons. A comprehensive determination of the mass distribution for the OC population is the crucial step.

For a given age, mass and chemical composition, stellar evolution models can calculate the luminosity of stars establishing mass-luminosity relations. So, in principle, it is possible to derive stellar masses from luminosity measurements. These mass determinations are often referred to as "photometric" or "luminous masses". Masses derived from kinematic measurements are referred to as "dynamical masses" and "virial masses" (for clusters assumed to be in virial equilibrium). Additionally, one may consider "tidal masses" which are masses determined directly from the radius of a cluster given the galactic tidal forces acting on the cluster, which will be discussed in more detail in Section 4.2.5.

In the literature, the only large scale catalogue of OC masses is that of Piskunov et al. (2007) containing tidal masses for 650 OCs. In addition to that catalogue, there are only a few studies (for example, Bisht et al. (2020) and Gao et al. (2021)) that focus on a small number of clusters for which the authors provide mass determinations. In total, masses have only been determined for about $\sim 35\%$ of the known clusters. However, since most were based on the ASCC 2.5 catalogue (Kharchenko, 2001) which has a very limited photometric and astrometric depth ($K \sim 12$) and high errors, a more complete catalogue of OC masses determined from high quality data is clearly needed. This is one of the key goals of the work here presented.

1.3 Mass loss mechanisms in open clusters

Bright massive and intermediate-mass stars suffer violent deaths that generate stellar winds and supernovae shocks. In a cluster, these phenomena may remove a significant part of the cluster's mass, dispersing it into the surrounding medium. The death of these stars quickly reduces the mass of the cluster which weakens the bounds between the remaining stars. This is referred to as mass loss by stellar evolution and it is the dominant mechanism of mass loss for ages up to about 4 Myr (Lamers et al. (2010), see Figure 1).

On longer timescales, clusters loose mass through evaporation (or dissolution) which is the process in which stars leave the cluster if their velocity is higher than the escape velocity of the cluster. Stars interact with each other leading to the equipartition of energy, in a process called relaxation. On average, the stars will have similar kinetic energy so the low mass stars will have higher velocities than the high mass stars ($E_k = 1/2mv^2$). The lowest mass stars will have velocities higher than the escape velocity of the cluster so they will exit the cluster, producing a preferential loss of low mass stars. Over time, as the total mass of the cluster decreases, the escape velocity at the boundary of the cluster (R) also decreases ($v_{esc} = \sqrt{2GM/R}$) so stars of increasingly higher masses will gradually escape from the cluster until it completely dissolves into the field population.

From the outside, clusters suffer an additional dissolution effect due to the tidal forces from the gravitational field of the host galaxy, pulling the stars away from the cluster. The clusters closer to the centre of the Galaxy experience higher tidal forces as more material is concentrated in the inner regions of the Galaxy, than those on the outer regions therefore disruption is dependent on the location of the clusters within the Galaxy.

Besides the tidal interaction with the Galaxy, encounters with Giant Molecular Clouds also disrupt clusters due to the strong gravitation field of GMCs (Lamers and Gieles (2006)) which introduce extra tidal forces. These GMCs are mainly located within the plane of the Galaxy so, in contrast to globular clusters that spend most of its life away from the plane, open clusters are subject to multiple encounters during their lifetime.

Understanding these processes of disruption is important to explain not only how the transition from clustered formation to the smooth field population occurs but also to characterize the tidal field of the Galaxy itself.

The *Gaia* mission (Gaia Collaboration et al., 2021) of the European Space Agency (ESA) has the goal of pinpointing the positions of nearly two billion objects with extreme accuracy and provide accurate astrometric and photometric data. It is the deepest all-sky survey ever done, reaching magnitude depths around 20.5 in the G band with accuracies at the mili-magnitude level. With 3 data releases so far, it has provided astrometric and photometric data for more than 1 billion objects, allowing to obtain the most accurate 3D view of the Milky Way and study its dynamics with unprecedented detail. For stellar clusters, in particular, *Gaia* data has made it possible to identify more cluster members, find hundreds of new clusters and perform homogeneous determinations of cluster parameters.

In this work, we made use of data from *Gaia* and the cluster parameters determined in a state-of-the-art catalogue of open clusters (Dias et al., 2021) to determine the first Gaia-based homogeneous mass catalogue of OCs. This catalogue was then used to study the disruption experienced by clusters in the solar neighbourhood.

In the next chapter, we will present the methods used for the determination of cluster masses. As will be explained, this requires also determining their radii. Chapter 3 is dedicated to the characterization of the observational data used in this work. Chapter 4 present our results for the masses and radii of 1724 OCs, followed by the investigation of the OC disruption timescale in the solar neighbourhood. Chapter 5 closes this work with conclusions and lays out directions for follow up research.

Chapter 2

Methods

To obtain the mass of a cluster, we must first know which stars to consider. The boundary of each cluster can be defined as the radius at which the stars inside of it are still gravitationally connected to the cluster, often called the tidal radius (R_t). Beyond that radius, the tidal forces from the host galaxy overcome the gravitational attraction of the cluster. So, despite the observation of many stars in the outskirts of the clusters, with similar kinematic and photometric properties, stars beyond R_t are already dissipating into the field population and should not be considered part of the cluster. For this reason, we consider only stars inside R_t when determining the cluster's mass.

Before the *Gaia* mission, the catalogues of OC radii available were those of Piskunov et al. (2007) and Piskunov et al. (2008) which contained tidal radii for a total of 650 OCs. Since *Gaia*, several studies have determined the tidal radii for small numbers of clusters under study (e.g., Yeh et al. (2019), Bisht et al. (2020), and Angelo et al. (2021)). Recently, Tarricq et al. (2021) have produced a catalogue containing radii for 164 OCs using *Gaia* data. However, this new catalogue only covers about $\sim$ 10% of the total number of clusters with available *Gaia* based lists of cluster members. Thus, a first step towards building the comprehensive catalogue of cluster masses needed for this work, is to determine their radii.

2.1 Cluster radii determinations

In a cluster, stars are not homogeneously distributed in space. Typically they are more concentrated near the centre of the cluster and show a radial density profile that follows a smooth decreasing curve. The most widely used parameterization is the King empirical function (King, 1962). Although it was first introduced to describe the surface density of globular clusters, it has been widely used for open clusters. It contains 3 parameters: the tidal radius (R_t) which is the radius beyond which the density of stars is indistinguishable from the density of the background, the core radius (R_c) defined as the radius for which the density is equal to half of the central density and N_0 which is a normalization factor related to the central density of the cluster. By fitting the observed radial density profile with the King function, it is possible to determine the core and tidal radius of each cluster. As this function has been used in other large-scale studies, we opted for this profile to make a direct comparison with related work. In the future, we plan to test other profiles to access the effect of choosing particular profiles.

The original formula derived by King, which provides the number of stars per pc^2, is:

$$n(R) = N_0 \left(\frac{1}{\sqrt{1+(R/R_c)^2}} - \frac{1}{\sqrt{1+(R_t/R_c)^2}} \right)^2 \tag{2.1}$$

An additional constant c was introduced to this equation, following Küpper et al. (2010), to account for the foreground/background population of stars that act as contamination. In every line of sight, we detect stars that are not part of any cluster (called field stars) that influence the determinations of cluster parameters. As they are randomly distributed in space, they lead to an overall increase of the counts in the density profile, so parameter c, which is an addictive constant, allows to account for this effect. Outside the tidal radius, all stars are considered to be from the background so, for R $\geq R_t$, the density remains constant in the vicinity of the cluster.

$$n(R) = \begin{cases} N_0 \left(\frac{1}{\sqrt{1+(R/R_c)^2}} - \frac{1}{\sqrt{1+(R_t/R_c)^2}} \right)^2 + c & if\ R < R_t \\ c & if\ R \geq R_t \end{cases} \tag{2.2}$$

To determine the radial density profile for each cluster, we obtained the distance of each star to the centre of the cluster which is given by α_c in right ascension and δ_c in declination. First, we determine their projected coordinates through the following expressions, as suggested by Ven et al. (2006) and Olivares et al. (2018), where D is the distance to the cluster.

$$\begin{cases} x = D\ sin(\alpha - \alpha_c) cos(\delta) \\ y = D\ cos(\delta_c) sin(\delta) - sin(\delta_c) cos(\delta) cos(\alpha - \alpha_c) \end{cases} \tag{2.3}$$

These coordinates (x, y) correspond to the projection of each star coordinates (α, δ) on the plane of the sky, tangential to the celestial sphere at the coordinates of the cluster's centre. This avoids distortions caused by strong projection effects for clusters in high declinations or those nearby that also suffer projection effects due to the curvature of the celestial sphere. Then, to obtain the radial distance (R) of each star to the centre of the cluster, we use $R = \sqrt{x^2 + y^2}$.

Considering the projected spatial distribution, we counted the number of stars inside concentric rings to obtain the density distribution for each cluster (Figure 2.1). For clusters where the farthest star is less than 12 pc from the centre, we used a ring size of 0.5 pc and opted for 1 pc for the other cases. This choice was somewhat arbitrary, but the results were checked to ensure that there was no under/over sampling of the density profile in clusters with fewer stars.

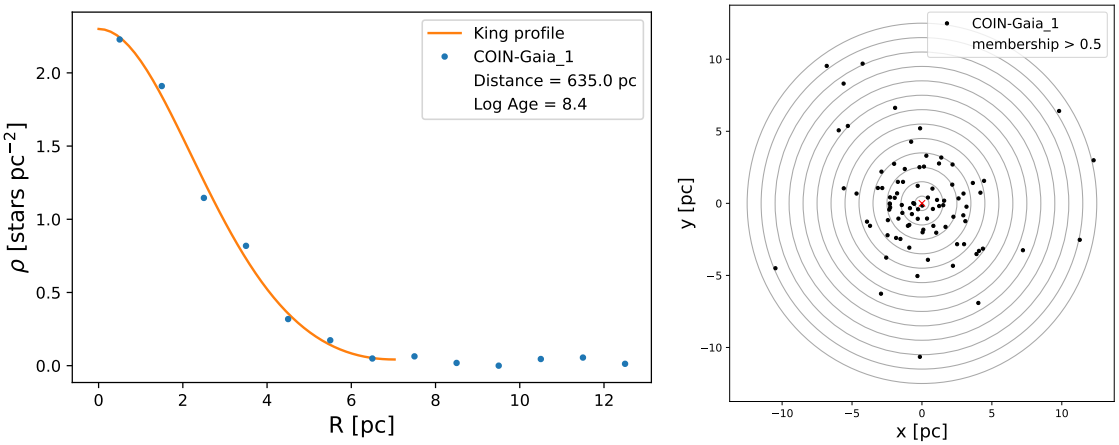

Figure 2.1: (Left) Radial density profile for open cluster COIN-Gaia 1 (blue points) with fitted King profile (orange line). (Right) Projected spatial distribution of the stars in open cluster COIN-Gaia 1 (black dots) with concentric rings (light grey) used to obtain the density profile.

Then, we fitted the King profile to the observed density distribution using a Non-Linear Least-

Squares Fitting python package, LMFIT (Newville et al., 2014). This package finds the optimal set of parameters by exploring the parameter space to determine the maximum likelihood values. We used the Maximum Likelihood Estimation from the *emcee* method within LMFIT to obtain the King parameters and adopted +1σ and -1σ error estimates as the upper and lower uncertainty of each King parameter, respectively.

Emcee (Foreman-Mackey et al., 2013) is a python implementation of the Markov chain Monte Carlo (MCMC) Ensembler sampler that samples the parameter space to find the distribution of the King parameters. Very briefly, we can describe MCMC as a random sampling technique in which each step (or each new parameter combination) depends only on the previous step. It "walks" over the parameter space to find the combination that is most likely to be the optimal (i.e. find the combination that maximizes the likelihood). Instead of going through every single combination, its algorithm samples the distribution of parameters to find the maximum which makes this method much faster than doing a grid search for multidimensional parameter spaces. Codes such as *emcee* include improvements to the original MCMC method to increase the speed and accuracy of the methods (Goodman and Weare, 2010) and have been widely used in astronomy.

To avoid having results dependent on the starting point, it requires that we provide initial "burn-in" steps to allow the method to converge. For that reason, we assigned 100,000 steps and discarded the first 5000. As recommended in Goodman and Weare (2010), the convergence of the chains was checked based on the value of the integrated autocorrelation time (τ), that works as a measure of the robustness of the MCMC analysis, to ensure that the Monte Carlo error (σ) is within a few percent ($\sigma^2 = (\tau \times$ variance) / number of steps). According to Goodman and Weare (2010), τ must be above 50 for every parameter so that the estimates can be trusted, i.e., the Monte Carlo error (or sampling error) is around a few percent. Considering the 100,000 steps, for clusters where τ was under 50, LMFIT raised a warning and we increased the number of steps. More details on this are discussed in Section 4.1.1.

As it would be too computationally expensive to run in a personal computer (about 60 days considering 100,000 steps for each OC) which would prevent the possibility of testing and repeating runs, we made use of the remote server of CENTRA-SIM to parallelize the code by running it simultaneously in 6 CPU, each with 4 cores. Each CPU was commanded to run on a specific portion of the sample (about 300 open clusters per CPU) therefore reducing the computation time by a factor of 6, leading to a total time of 10 days.

While testing the code, convergence problems were found for some clusters where the tidal radius given by the code was at the maximum limit allowed by us. Problems in fitting the King profile are often attributed to the sparse nature of open clusters that combined with contamination by field stars, make the fitting more difficult. We noticed that the results were very sensitive to the value of parameter c so, to overcome such problem, we tested several small modifications to the model. These modifications include considering parameter c as the median of the points outside the tidal radius in each iteration or removing the constant c entirely. None of these modifications provided significant improvements in the convergence of the method.

In addition, we noticed some cases where the density in the outer regions of the cluster is of order $\sim 10^{-2}$ $stars/pc^2$. If the density is already as low as that in the outer regions, we consider it to be the background so the tidal radius should not be larger than that. For that reason we implemented a threshold at 0.02 $stars/pc^2$. For each combination of parameters, the radius at which the King profile is smaller than 0.02 is checked and the run is discarded if the tidal radius considered in that run is above that value. The implementation of this threshold allowed to reduce the number of clusters that were not converging.

2.2 Mass determination

In this work, we address the determination of open clusters' luminous mass through the comparison of the observed luminosity distribution to the theoretical luminosity distribution for each cluster. The theoretical luminosity distributions, commonly designated as luminosity functions (LFs), give us the number of stars in each magnitude bin for a cluster with a known mass. The LFs were taken from the Padova database of stellar evolutionary tracks and isochrones (Bressan et al. (2012), Marigo et al. (2013), Marigo et al. (2008), Chen et al. (2015), PARSEC version 1.2S). We retrieved a grid of LFs calculated for the *Gaia* filter passbands of Maíz Apellániz and Weiler (2018) and Kroupa (2001) initial mass function (distribution of stellar masses in the cluster at birth) corrected for unresolved binaries of stars, scaled to the solar metallicity of $Z = 0.0152$.

The luminosity functions obtained from Padova models are normalized to 1 solar mass, for the mass of the stellar population initially born. Since Padova models are referred to the initial mass but include the effect of stellar evolution (as death of stars), the LF for the cluster's age will no longer be normalized to 1 solar mass as some stars have already died. For this reason, the scaling of the LF to the observed luminosity distribution does not provide the current mass of the cluster. So, a scaling correction had to be implemented to account for that effect. To do so, we generated, from Padova models, a population of stars with known mass and evaluated the mass of the cluster (by summing the individual masses of each star) at different ages from log(age) = 6.6 to 10, with steps of 0.1 dex. We observed that the mass decreases with age in a linear trend so we performed a linear fit (Figure 2.2) and obtained $corr = -0.135 \times log(t) + 1.781$, where "*corr*" is the mass of the LF at the selected age instead of 1 solar mass.

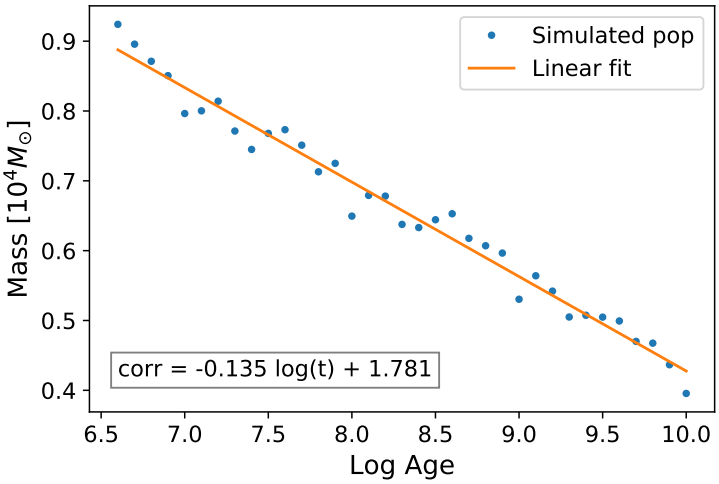

Figure 2.2: Mass of a simulated cluster (from PADOVA models) at different ages from log(age) = 6.6 to 10, with steps of 0.1 dex with linear fit overplotted.

To compare the theoretical and observed luminosity distributions, we plotted the histogram of the absolute magnitudes and scaled it to match the theoretical LF. The absolute magnitudes were obtained through equation $(M = m - 5log(d) + 5 - A)$ where M is the absolute magnitude, m is apparent magnitude, d is distance (in parsec) and A is the interstellar absorption.

We adopted two mass determination methods. One method consists in the least-squares fit of the model luminosity distribution to the observed one. The LF is multiplied by a number, called "scaling factor", to match the observed luminosity distribution. Using equation 2.4, where LF_{count} stands for the number of stars in each bin predicted by the luminosity function and obs_{count} is the observed number of stars (in each bin), we obtain the deviation of the observed star counts in relation to the theoretical

counts, as is shown by the black arrows in Figure 2.3. The scaling factor that minimizes that equation multiplied by the scaling correction (*corr*) is the current total mass of the cluster, in units of solar masses.

$$\chi^2 = \frac{\sum_N (LF_{count} \times factor - obs_{count})^2}{N_{bins}} \tag{2.4}$$

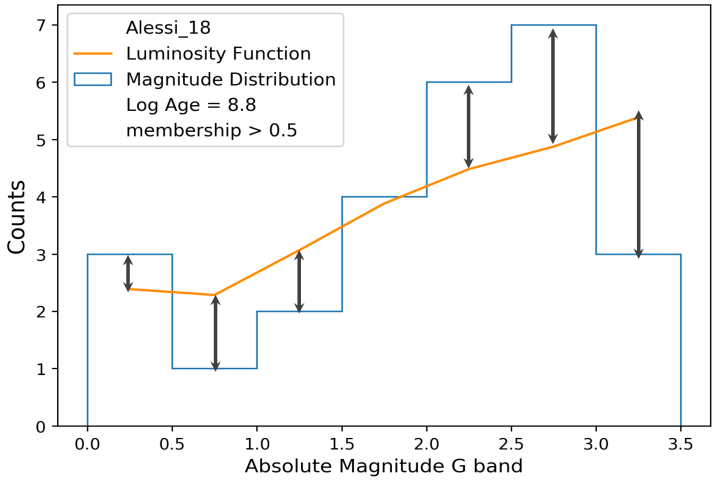

Figure 2.3: Histogram of the stellar absolute magnitude distribution, in the *G* band, for open cluster Alessi 18. Orange line represents the luminosity function for the cluster's age multiplied by 269 (scaling factor) which corresponds to a mass of 122 $M_\odot$.

The second method is the direct division of the areas under the magnitude distributions. Since both the theoretical and observed magnitude distributions are evaluated using the same bin width, dividing the areas is simply dividing the sum of the observed number of stars (or counts) by the sum of the predicted counts.

The data used in this work comes from *Gaia* DR2 (Gaia Collaboration et al., 2018) which contains photometry in 3 bands (G, G_{BP}, and G_{RP}) that fall in the visible region of the spectrum. Details on the data used are given in Section 3.1. Both mass determination methods were applied to the G band to make comparisons on their performance and to the G_{BP} and G_{RP} bands to check the robustness of the results, as will be discussed in Section 4.2. As we wish to compare the results between the different bands, only stars detected in the 3 *Gaia* bands were considered.

The two methods share the same principle however, they produce slightly different results. The second method does not consider the shape of the magnitude distribution as it only considers the sum of the counts. On the contrary, the least-squares method is sensitive to the shape because each bin has a different weight so, in case the cluster parameters (such as age or distance) are not well determined and/or there is incompleteness, the results will be more affected. This will be seen later as a larger difference in the mass determined using different bands considering the least-squares method. Since the luminosity function is dependent on age and distance, if those are not correctly determined, the shape of the LF will not be adequate to the cluster.

2.2.1 Completeness Cut

Although *Gaia* photometry is complete down to an apparent magnitude of ~ 20 in the G band, since the membership probabilities from Dias et al. (2021) were only assigned for stars with apparent magnitude G ≤ 18, we must only consider stars with $G_{apparent} \leq 18$. We must also note that when the apparent magnitudes are converted to absolute magnitudes and discretized in bins, the absolute magnitude bins will not be complete as the last bin may include magnitude values that correspond to $G_{app} > 18$. In fact, in some cases, there is indication of this effect as the last bin has very few stars when compared to the predicted luminosity function (see Figure 2.4).

Using luminosity functions in apparent magnitude would cancel this effect however, retrieving LFs from Padova in apparent magnitude is not possible in the website interface. So, to account for this artificial incompleteness effect, we imposed a completeness cut in the G band by removing the absolute magnitude bins that correspond to $G_{apparent} \geq 18$. In most cases, this results in the removal of just the last bin, as exemplified in Figure 2.4, with few exceptions where more than one bin is removed. Since the true completeness limit in G_{BP} and G_{RP} bands is not known, we chose to remove the last bin in those bands.

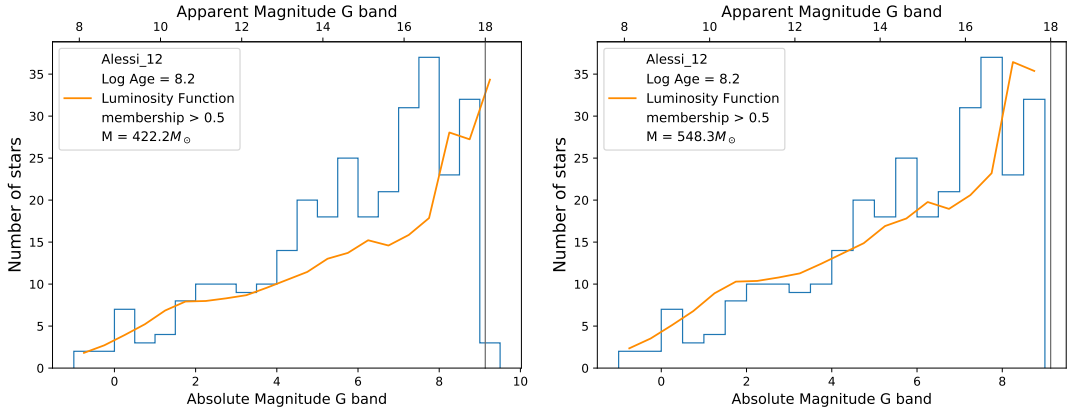

Figure 2.4: Absolute magnitude distribution for open cluster Alessi 12 with luminosity function scaled to match the observed distribution before (left) and after (right) implementing the magnitude cut in the G band, using the least squares' method.

Additionally, as the Padova interface forces us to manually retrieve LFs for each bin width, our freedom of choice for the bin width was limited. Choosing a different bin width for every cluster would become unpractical so, for that reason, we tried to find a couple of bin widths that could work for the entire sample. We opted for a bin width of 0.5 magnitudes, however, for clusters with fewer stars this choice was not appropriate as it was oversampling the data, so we increased the bin width to 1, in those cases.

In future work, we plan to build our own luminosity functions using the Padova stellar evolution models and mass functions. This will give us more freedom in the choice of bin width and more easily convert between apparent and absolute magnitudes.

2.2.2 Bootstrap error analysis

To get the error associated with our mass determination methods, a bootstrap analysis was performed. The bootstrap method is a resampling technique used to estimate statistics on a population by sampling a data set with replacement. It generates new hypothetical samples that help to test the estimated value and efficiently determine the standard error. This method has the advantage that it is easy to implement

and it does not require any pre-assumptions to work.

In our work, it translates into taking samples (with the same size as the original) with replacement of the magnitude distribution for each cluster and then calculate the mass of the new samples. To select the number of samples necessary to have a robust error determination, we checked the error using 10, 50, 100 and 200 samples. Above 50 samples, the bootstrap error showed fluctuations outside the significant digits, so we decided to use 100 samples. This process was only applied to the least-squares method because, in the other method, the total number of stars is kept fixed so the sums are always the same, what changes is how the stars are distributed through the bins therefore bootstrapping cannot be applied to the second method. Considering the least-squares method, for each repetition, a mass is determined, so we adopt the standard deviation from the 100 mass values as the error of our method, for each cluster.

Chapter 3

Sources

3.1 Catalogue of Dias et al., 2021

The *Gaia* mission of the European Space Agency (ESA) was launched in 2013 with the goal of pinpointing the positions of nearly two billion objects with extreme accuracy at the microarcsecond level. This is allowing astronomers to study the Milky Way structure and dynamics with unprecedented detail. This mission has had 3 data releases and more are planned for the future. The first *Gaia* Data Release (DR1: Gaia Collaboration et al. (2016)), contained positions and magnitudes for more than one billion sources. The second and third data releases (DR2: Gaia Collaboration et al. (2018), DR3: Gaia Collaboration et al. (2022) and Gaia Collaboration et al. (2021)) presented astronomers with astrometric data of unprecedented accuracy (up to a hundred times better than previous proper motion catalogues) for more than 1 billion sources.

It is the deepest all-sky survey ever done, reaching magnitude depths of G $\sim$ 20.5 with accuracies at the mili-magnitude level. It provides 5 astrometric parameters: positions (in right ascension (α) and declination (δ)); proper motions (in α and δ) and parallaxes. It also provides magnitudes in three photometric filters (G, G_{BP} and G_{RP}) for more than one billion sources. These photometric bands fall in the visible region of the electromagnetic spectrum (400 to 900 nm) which is the sensitivity window of the CCDs used in the *Gaia* satellite (Figure 3.1). The nominal uncertainties of *Gaia* reach accuracies of 0.02 mas in parallax and 0.05 mas/yr in proper motions, for G > 14, and 2 mas and 5 mas/yr, respectively, for G $\sim$ 21.

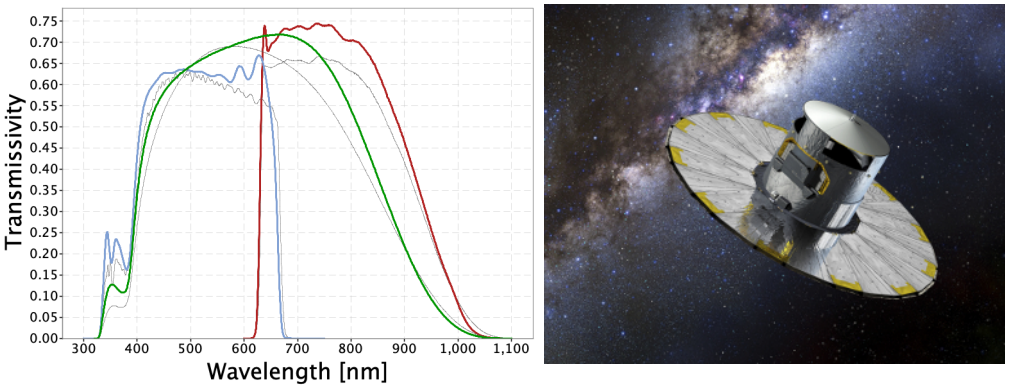

Figure 3.1: (Left) *Gaia* passbands as produced by Coordination Unit 5 of the Gaia Data Processing and Analysis Consortium. The coloured lines in the figure show the G, G_{BP} and G_{RP} passbands (green: G; blue: G_{BP}; red: G_{RP}). Credits: ESA/Gaia/DPAC. (Right) Artist's impression of the *Gaia* spacecraft. Credits: ESA–D. Ducros, 2013.

In what regards open cluster studies, *Gaia* astrometry and photometry has allowed to identify cluster members and find hundreds of new clusters and cluster candidates (Liu and Pang, 2019; Sim et al., 2019; Castro-Ginard et al., 2020; Ferreira et al., 2021). With its deep photometric data, the determination of distances, ages and extinctions with higher accuracy has been made possible which allowed homogeneous determinations of cluster parameters (Cantat-Gaudin et al., 2018; Bossini et al., 2019; Cantat-Gaudin et al., 2020; Monteiro et al., 2020; Dias et al., 2021).

While there are several catalogues that use data from *Gaia*, the majority is focused on a small number of clusters. The two main *Gaia*-based large scale catalogues available in the literature are the Cantat-Gaudin et al. (2020) and Dias et al. (2021) catalogues. The catalogue of Cantat-Gaudin et al. (2020) contains cluster parameters (distance, age, and interstellar reddening) for 1867 OCs determined using an artificial neural network. In this study, we make use of the slightly more recent catalogue of Dias et al. (2021) which contains cluster parameters such as age, distance, extinction, proper motion, metallicity, amongst others determined from *Gaia* Data Release 2 data. The clusters' age, distance and extinction were obtained using a cross-entropy (CE) method to fit theoretical isochrones to the CMDs of cluster member stars, as described in Monteiro et al. (2017) and Monteiro et al. (2021). This catalogue is homogeneous since the parameters were determined with the same isochrone fitting method for all clusters.

In addition to the main catalogue, tables of individual stellar membership probabilities from published literature and individual stellar data, from *Gaia* DR2, are also provided for each cluster. The memberships come from several other catalogues (e.g. Castro-Ginard et al. (2020), Sim et al. (2019), and Liu and Pang (2019)), and were determined using different methods based on the astrometric, photometric and proper motion data of the stars in each cluster. It should be noted that Dias et al. (2021) recalculated the memberships for the clusters in Liu and Pang (2019) and Castro-Ginard et al. (2020) as the stars in those clusters were only assigned, by the authors, 1 or 0 for members and non-members, respectively.

3.2 Sample characterization

3.2.1 Age, distance and spatial distribution

Figure 3.2 shows the distribution of the 1743 open clusters from Dias et al. (2021) in the X-Y plane of the Galaxy, with positive Y in the direction of the Galactic anticentre. The Galactic centre is at (0,0), the Sun is situated at (0,8.3) kpc and the X axis points to the Galactic rotation direction. The spatial distribution does not present a homogeneous distribution in the X-Y plane because there is a higher number of clusters detected in the outer region of the Galaxy (Y > 8.3 kpc) when compared to the inner region (Y < 8.3 kpc). Since our sample from Dias et al. (2021) only contains stars brighter than G $\sim$ 18, it is magnitude limited. Factors that decrease the apparent brightness of clusters members, such as distance, extinction and age lower the distance up to which we can detect stars of G $\sim$ 18. In the direction of the Galactic centre, the level of extinction is much higher due to the higher concentration of dust that block optical wavelengths (see Figure 3.2 on the right). So, we can only detect sources up to a distance that is shorter than in the anticentre direction. Additionally, as some studies may be focused on a specific region, it artificially increases the number of clusters in those regions, leading to a non-uniform distribution of clusters in the X-Y plane.

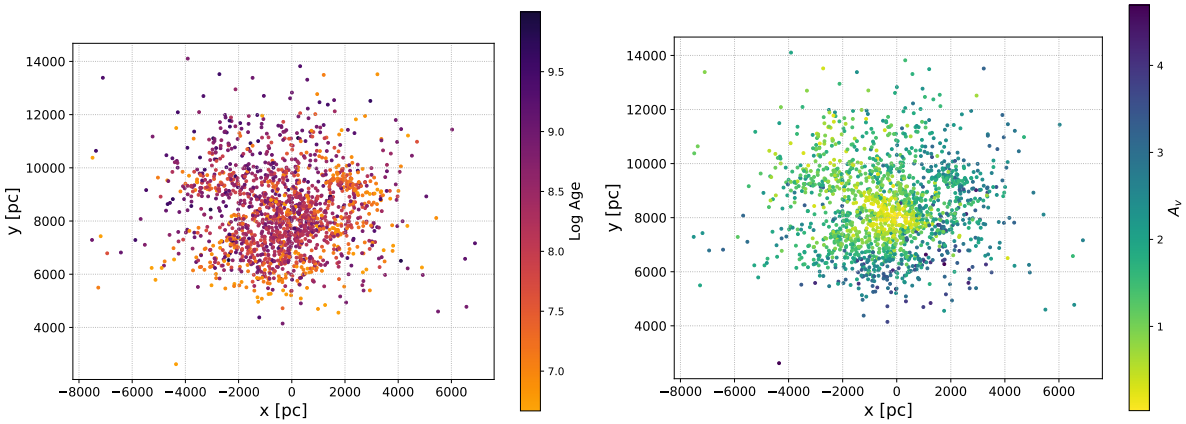

Figure 3.2: Distribution of OCs in the X-Y plane of the Galaxy, with positive Y in the direction of the Galactic anticentre and the X axis pointing to the Galactic rotation direction. Galactic centre is at (0,0) and the Sun at (0, 8.3). On the left, the open clusters are color-coded by age and, on the right, by extinction (A_v).

Figure 3.3 shows the distance from the Galactic plane against Galactocentric distance, color-coded by age. In this plot, the increase of the scale height of the disc is evident in the outer regions. The scale height is defined as the vertical distance over which the number density decreases by $1/e$ (Binney and Tremaine, 2011). In other words, it gives a sense of the scale where the clusters are contained in relation to the distance from the Galactic plane. In the outer regions of the Galaxy, the clusters show a larger spread in the vertical axis corresponding to a larger scale height. We can also see that younger clusters are found near the Galactic plane, while older clusters can be found at all Galactic altitudes as already pointed out by Cantat-Gaudin et al. (2020).

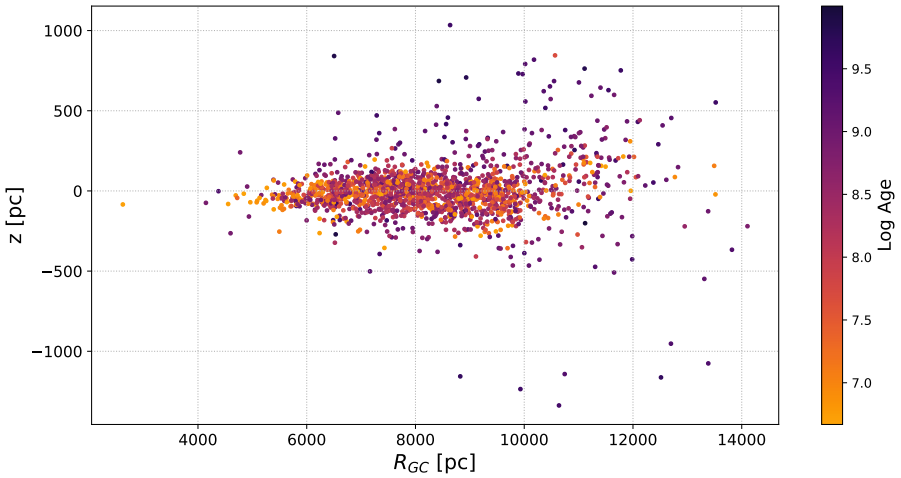

Figure 3.3: Distribution of open clusters in distance from the Galactic plane (z) against Galactocentric distance (R_{GC}), color-coded by age.

In Figure 3.4, we plot the observed distribution of age and distance of our sample. For both distributions, we fitted a Gaussian Kernel Density Estimation that was implemented using the function *gaussian_kde* of Python Scipy package considering a bandwidth of 0.27 dex for the age distribution and 454 pc for the distance distribution.

In general, the sample contains OCs within a few kiloparsec around the Sun with ages of a few million years to a few billion years.

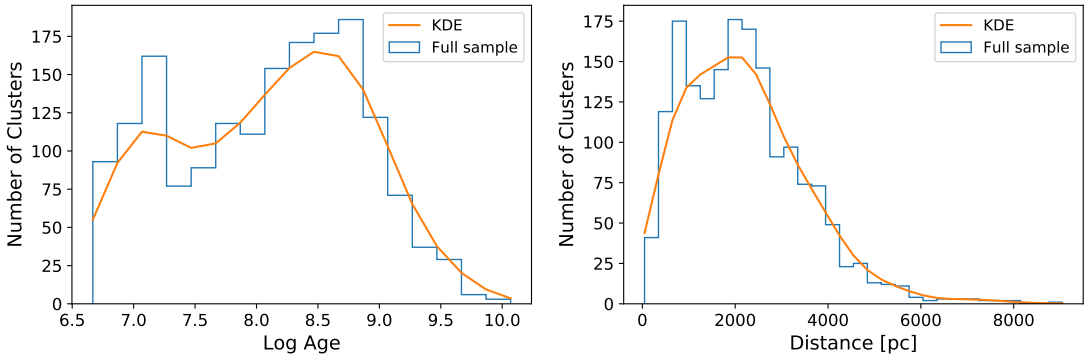

Figure 3.4: Distribution of age (left) and distance (right) of open clusters in the catalogue of Dias et al. (2021), with fitted Gaussian KDE with bandwidth of 0.27 dex for the age and 454 pc for the distance distribution.

3.2.2 Membership Probability Cut-off

To distinguish between member stars and stars from the field population, it is important that we assign a membership probably (P_{memb}) to each star in the field of view around the cluster, translating its probability of belonging to the cluster. In general, the distribution of membership probabilities for stars detected in the region of the sky around a cluster shows a U shape distribution with more stars at the highest and lowest membership probabilities, as seen, for example, for open cluster FSR 0441 in Figure 3.5. This means that it is possible to separate cluster stars from field stars, reducing the level of contamination. In this work, we take a conservative approach by considering only stars with a membership probability above 50%, which are stars that are statistically more likely to belong to the cluster than to be from the field population.

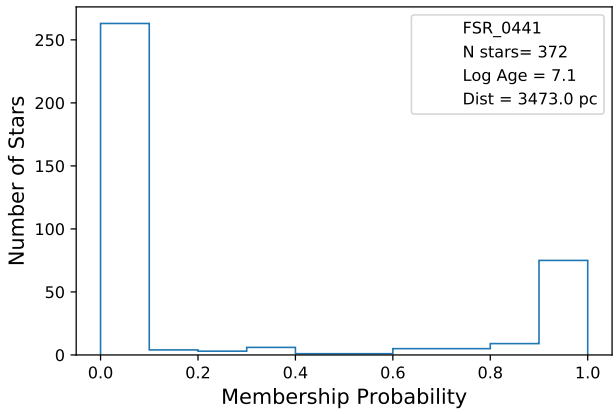

Figure 3.5: Membership probability distribution for open cluster FSR 0441 which is 12.5 million years old and at a distance of 3473 parsec.

The sample from Dias et al. (2021) that we use in our work, as mentioned before, contains stellar memberships determined in other smaller catalogues. Dias et al. (2021) visually inspected the clusters' CMD and did not include in the published catalogue those that seemed dubious or likely not physical objects. These objects come mostly from the catalogue of Liu and Pang (2019) and some from Castro-Ginard et al. (2020). In fact, in the Liu and Pang (2019) catalogue, 72% of the sample was classified, by the authors, as class 3 (worst quality) regarding the width of the main sequence (MS), age and quality of the isochrone fitting. In Castro-Ginard et al. (2020) catalogue, 36% was classified (by the authors) as class C (worst quality) and 15% as class B. According to Dias et al. (2021), these are cluster candidates that require further confirmation.

Dias et al. (2021) visually inspected the CMD with the fitted isochrone over-plotted of each cluster to exclude poor quality determinations. So, in principle, the 1743 OCs reported pass the bona fide criteria however, we detected some cases where there is considerable scatter and/or poor isochrone fit. For that reason, the CMD for the 1743 OCs with the fitted isochrone over-plotted were visually inspected. For those that showed signs of contamination, i.e., stars very far from the isochrone, we increased the membership probability cut-off trying to reduce the contamination. At a first stage, we determined how many showed significant contamination (only $\sim 11\%$) and, later, plotted the CMD for every membership cut-off above 0.5 (from 0.6 to 0.9). As we are dealing with a trade-off between the level of completeness and contamination, we chose the cut-off that removed the contamination by as much as possible but also kept a significant number of stars to avoid losing too much information. The membership cut-off distribution for our sample is the following:

Table 3.1: Number of clusters per membership cut-off used in this work.

Membership Cut-off	P >0.5	P >0.6	P >0.7	P >0.8	P >0.9
Number of clusters	1555	34	66	53	35

As an example, we show in Figure 3.6 the CMD for open cluster NGC 2244 considering, on the left, only stars with a membership probability above 0.5 ($P_{memb} > 0.5$) and, on the right, with the chosen cut-off of $P_{memb} > 0.8$. The comparison shows that increasing the membership cut-off allows to reduce the scatter around the main-sequence, i.e, removing a significant part of the contamination.

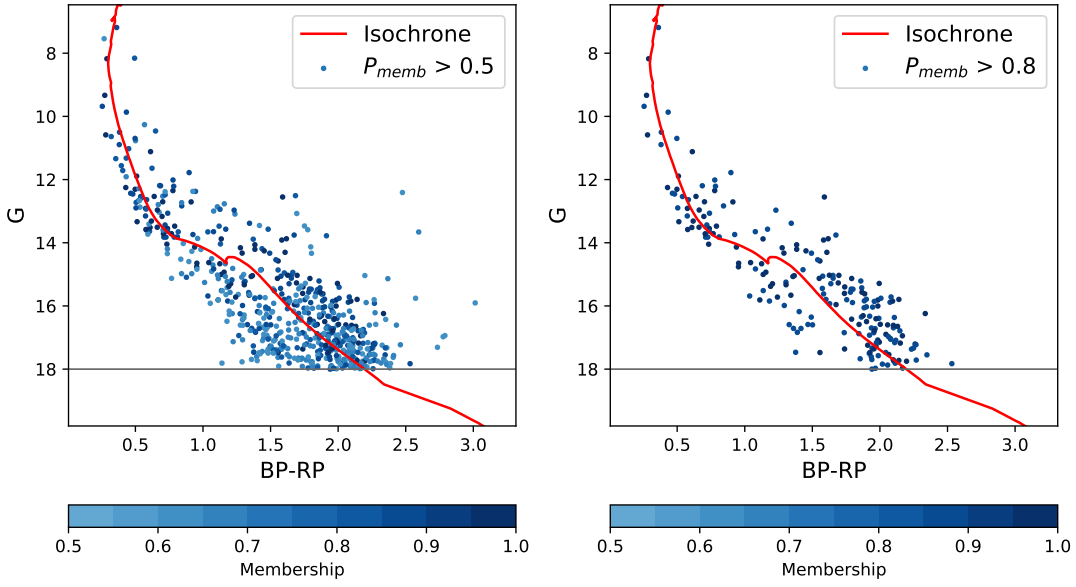

Figure 3.6: Color-magnitude diagram for open cluster NGC 2244 with isochrone for log(age) = 7.1 dex and distance of 1254 pc, considering only stars with membership probability above 0.5 (left) and 0.8 (right).

Of those that contained higher contamination and, therefore, needed a higher membership cut-off, we note that 69 (37%) are from the Castro-Ginard et al. (2020) catalogue and 47 (25%) are from the Liu and Pang (2019) catalogue. In fact, when looking at their color-magnitude diagram, we noticed that some don't show a main sequence nor any type of pattern that can be identified as a cluster. As Dias et al. (2021) suggested, further investigation is needed to conclude on the nature of the objects but that is beyond the scope of this work.

3.2.3 Photometric Classification

Due to the presence of several OCs with poor CMD fits, we classified the clusters using as criterion the width of the main sequence and quality of the isochrone fitting. Considering the CMDs for the selected member stars of each cluster, we visually inspected and classified the clusters into 3 categories: P1 (best quality), P2 (medium) and P3 (worst quality), where P stands for photometry. The first category (P1) contains open clusters that present a good isochrone fit where there is small dispersion, i.e., most of its stars are near the isochrone. The second category is for intermediate quality fits where there is some dispersion around the isochrone, which is evidence of contamination, or the isochrone does not provide the best fit to the data. The latter (P3) contains clusters where the isochrone does not trace the distribution of stars in the CMD, or the points are too scattered to find any pattern that resembles an open cluster. Table 3.2 contains the number of clusters in each category.

Table 3.2: Number of clusters per classification according to the quality of the CMD.

	P1	P2	P3
Number of clusters	939	649	155

Figures 3.7, 3.8 and 3.9 show typical cases for each classification. Despite the low definition of the sequence and/or the presence of scatter in the CMD, the 155 clusters in P3 were kept in the study under a quality flag.

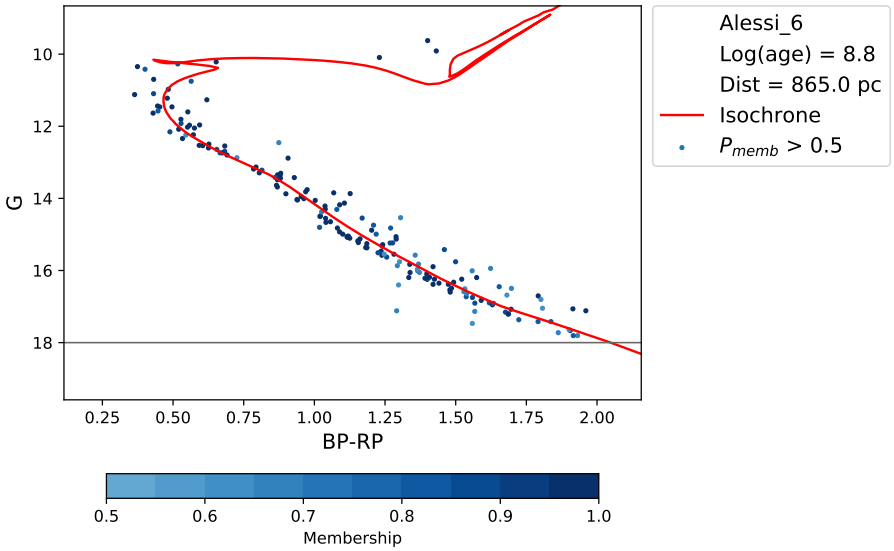

Figure 3.7: CMD for open cluster Alessi 6 which is a representative case of classification P1.

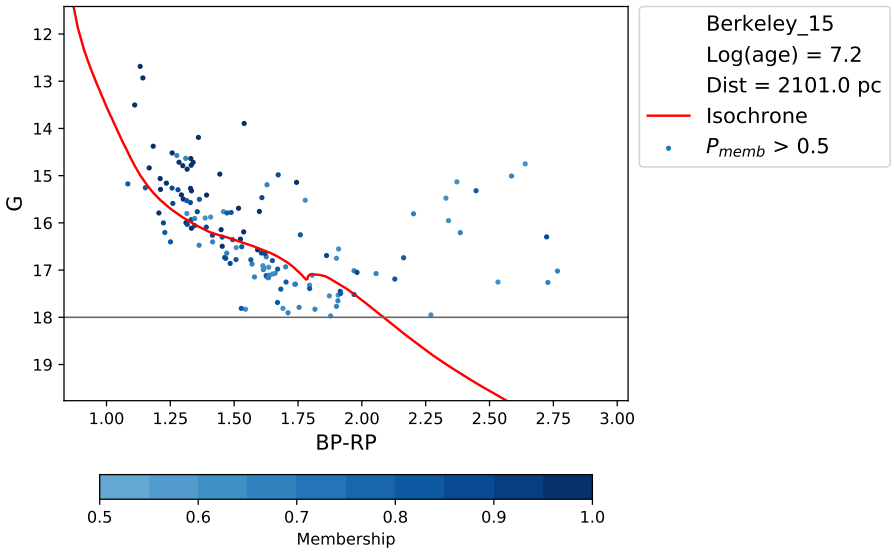

Figure 3.8: CMD for open cluster Berkeley 15 which is representative of P2.

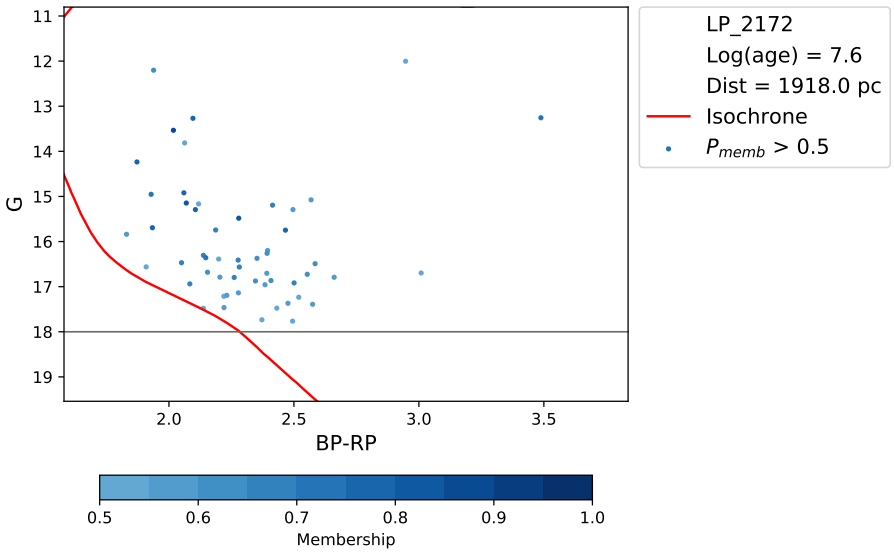

Figure 3.9: CMD for open cluster LP 2172, an example of clusters with classification P3.

Chapter 4

Results

4.1 Determining tidal radii

To determine the tidal radii some constraints were made to the process described in Section 2.1 to ensure the results have physical meaning. As the core radius cannot be larger than the size of the cluster, we restricted the possible values of the core radius to a maximum corresponding to the radial distance of the farthest star, in each cluster. Even though this constraint ignores the fact that additional undetected member stars could exist, the results show that the core radii are never at the maximum value considered so this choice simply defines an appropriate range of values to be considered for R_c. For the tidal radius, we defined a maximum value of 100 pc as most clusters have the farthest star under 30 pc from its centre. This choice has also been used in other studies such as Tarricq et al. (2021), with whom we compare our results in section 4.1.5.1.

An additional constraint was imposed to avoid considering values for the tidal radius that are smaller than the core radius, as it does not have physical meaning. These constraints were implemented such that, if the combination of parameters is not to be considered, the run will be discarded due to a very high χ^2 value, forced by us.

While visually inspecting our early results of the King profile fit, we detected 4 open clusters (Berkeley 58, Berkeley 59, Blanco 1 and NGC 7789) for which the centre is incorrectly determined (the listed centre is off the cluster distribution) so, we decided to remove them from our sample. This leaves us with a sample of 1739 OCs. After testing the best parameter ranges that worked for the entire sample, we applied the described procedure to the 1739 open clusters considering the following parameter ranges:

Table 4.1: Minimum and maximum parameter limits considered to fit the King profile to the observed density distribution. The maximum limit for the core radius is the value of the distance to the centre of the cluster of the farthest member star (max(R)).

Parameter	Minimum Limit	Maximum Limit
R_t	0.5 pc	100 pc
R_c	0.2 pc	max(R) pc
c	0.02 $stars/pc^2$	5 $stars/pc^2$
N_0	0	250

Our initial results showed that 72 OCs did not converge on any value. These are typically clusters that do not show a centrally concentrated distribution in the X-Y plane, so their density profile is almost flat or has an irregular shape that cannot be fitted with the King profile.

To overcome this issue, we changed the number of stars considered, by changing the membership

probability cut-off used for those clusters. For every possible cut-off (from 0.0 to 0.9), we determined the King parameters using the stars with membership above the cut-off considered. Of the 72 clusters for which the code did not converge using the cut-off chosen using the CMD, 57 provided results for at least one cut-off and 15 did not return values for any cut-off. For that reason, we will not consider those 15 clusters in the analysis as we do not have any King parameters estimates for them.

For those that converged on more than one cut-off, we chose the lowest cut-off that resulted in the best fit considering the quality of the King profile fit and how well the core and tidal radii adjust to the spatial distribution. We note, however, that these clusters have poor quality determinations. Our choice to consider the lower membership cut-offs follows the argument of Piskunov et al. (2007) which argued that using a lower membership cut-off provides better results for the King parameters because, despite including more contamination which increases the value of parameter "c", it results in a more robust fit. After implementing the changes in the membership cut-off, we obtained King parameters for 1724 OCs.

4.1.1 Uncertainty

Regarding the uncertainty of the King parameters, we applied the *emcee* method from LMFIT, as described in Section 2.1. Since running our code with a high number of steps for all clusters would be very time consuming (between $\sim$10 to $\sim$25 hours for each one of the 1724 clusters), we decided to start with 1×10^5 for all clusters. However, as mentioned in Section 2.1, for some open clusters, the number of steps given was not enough to fulfill the criterion of $\tau \geq 50$, which is necessary to have low sampling error and robust determinations from the MCMC chain. In those cases, the program recommended the use of a longer chain so, we gradually increased that number until the criterion was satisfied. In Table 4.2, we present the distribution of the number of steps used in our *emcee* analysis.

Table 4.2: Distribution of the number of steps used in the *emcee* analysis. Note that 1437 (83%) of our sample uses 1×10^5 steps and only 287 OCs use a higher number of steps.

Number of steps	Number of OCs
1×10^5	1437
2×10^5	188
5×10^5	79
8×10^5	11
1.2×10^6	5
2×10^6	4

4.1.2 Results

In Figure 4.1, the distribution of tidal and core radii is shown for 1724 OCs. The distribution of tidal radii shows a "bump" (secondary peak) around 50 pc which corresponds to clusters with convergence problems, as will be explained in Section 4.1.4. Excluding the clusters with poor quality fits (classified as R4, see next Section) the "bump" around 50 pc disappears. In the future, we will look in more detail to the convergence of the chains in these clusters. The distribution of tidal radii excluding clusters with poor fits is represented as the green histogram in the left plot of Figure 4.1. It presents a median value of 10 pc, consistent with the values found in the literature (for example, Piskunov et al. (2007) and Kharchenko et al. (2013)). In Figure 4.1 on the right, we show the core radii distribution of our sample which has a median value of 2 pc. The distribution of parameters N_0 and c are shown in the Appendix A. In every histogram, the bin width was chosen using the Knuth rule (Knuth, 2006).

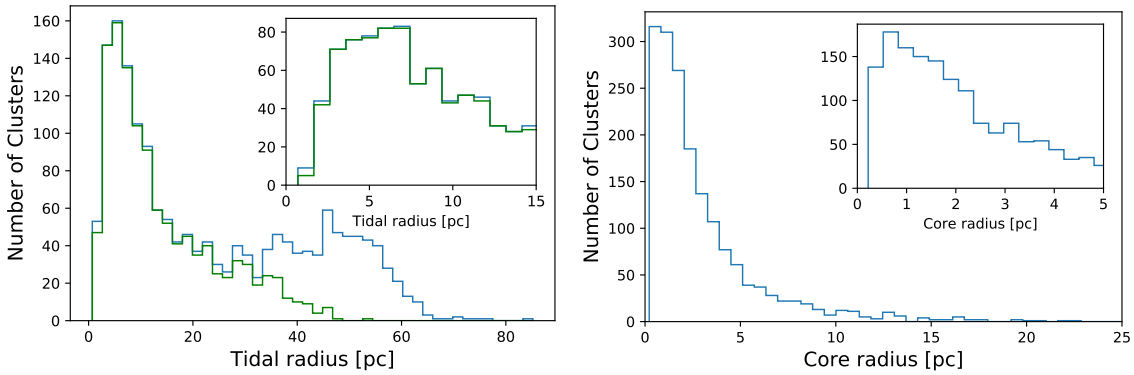

Figure 4.1: Distribution of tidal (left) and core (right) radii for 1724 OCs. In green we represent the distribution of tidal radii excluding poor quality fits. Median tidal radii is 10 pc (from green histogram) and median core radii is 2 pc. In the top right of each figure, we present a zoomed plot within 0 to 15 pc for tidal radii and 0 to 5 pc for the core radii.

The distribution of the upper and lower uncertainties for the core and tidal radii determined in this study are shown in Figures 4.2 and 4.3, for 1724 open clusters. The median of the fractional lower and upper uncertainties for the tidal radius are 40% and 144%, respectively. These large uncertainties reflect the difficulty in obtaining the tidal radii which could be due to, in some cases, the sparse nature of open clusters. For the core radius, the median lower uncertainty is 19% and the median upper uncertainty is 21%. The uncertainties for N_0 and c are also shown in the Appendix A.

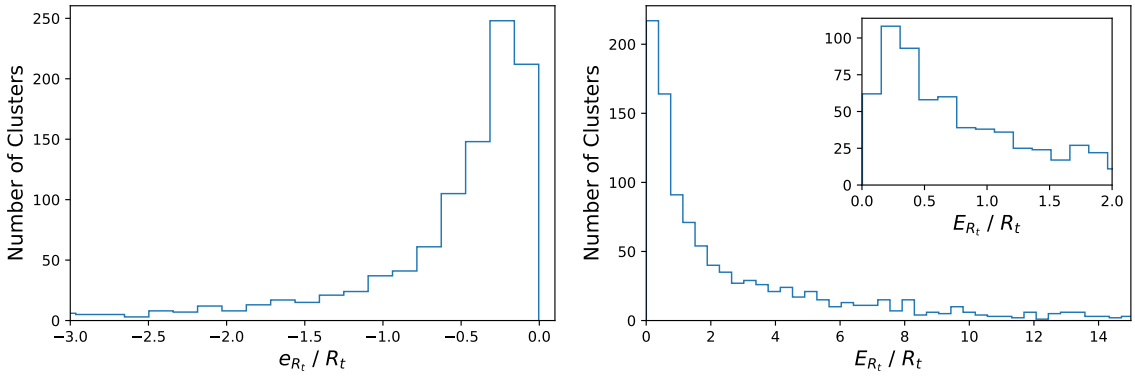

Figure 4.2: Distribution of lower (left) and upper (right) uncertainty for the tidal radii measured for 1724 OCs. On the left plot, there are 71 (4%) clusters with sporadic values below -3 and on the right plot, there are 52 (3%) clusters with values above 15, which we do not include in the figure to allow an easier visualization of the results.

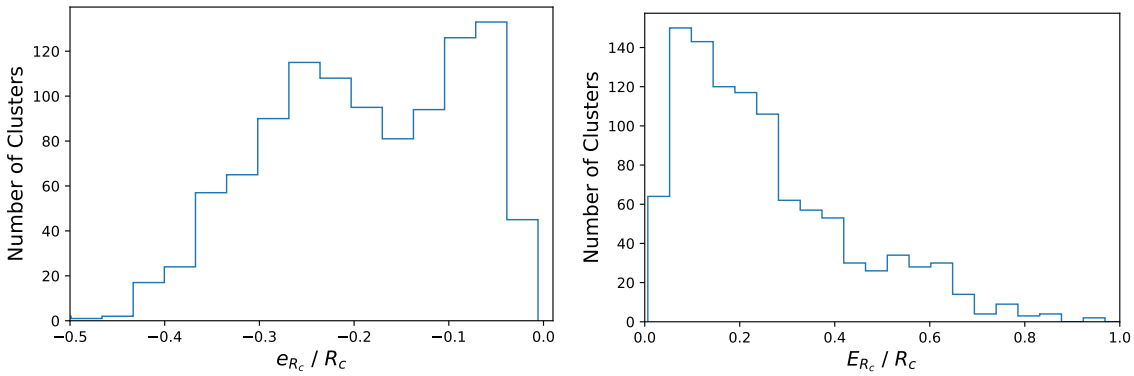

Figure 4.3: Distribution of lower (left) and upper (right) uncertainty for the core radii. On the left plot, there are 9 clusters with values below -0.5 and on the right plot, there are 7 clusters with values above 1, which are not represented.

4.1.3 Tidal Radii Classification

Even though our method provided results for 1724 OCs, we stress that not all present satisfactory results. To separate our results, we classified them according to the quality of the King profile fit to the observed density profile. To do so, we plotted, for each cluster, the observed density profile (in linear and logarithmic scale) with the fitted King profile and the X-Y spatial distribution of the member stars with circles at radius R_c and R_t overplotted. These 3 plots were visually inspected and each cluster was classified into one of 4 classifications: R1 (best quality), R2 (intermediate quality), R3 (worst quality) and R4 (non-reliable).

Classification R1 consists of clusters where there is a very good match between the observed density profile and the King profile, as well as a good fit of R_c and R_t to the X-Y spatial distribution.

Classification R2 consists of clusters that have a tidal radius which is visually well adjusted, but a core radius not well adjusted (too small or too large) to the observed central concentration.

Classification R3 consists of clusters that have similar values of tidal and core radius (which is an indication of possible convergence problems) and/or fewer stars which makes the fit less robust.

Classification R4 consists mostly of clusters where the tidal radius obtained is too large (around 50 pc) and/or is visually not adequate. Cases where the density distribution does not show a profile that can be fitted by the King profile are also included in this category.

Table 4.3: Number of open clusters per classification and respective percentage.

Classification	Number of OCs	Percentage
R1	342	19.8 %
R2	495	28.7 %
R3	338	19.6 %
R4	549	31.9 %

As we will use this classification in the selection of the sample of clusters with good determinations, from which we will base our study of cluster disruption, we analysed the spatial distribution of clusters from each classification separately to assess the spatial bias. As shown in the plots in Appendix A, each classification follows a similar spatial X-Y distribution so there is no spatial trend and no bias is introduced by selecting clusters based on the classification of their tidal radii determination.

In the following figures, we give representative examples of clusters in each classification.

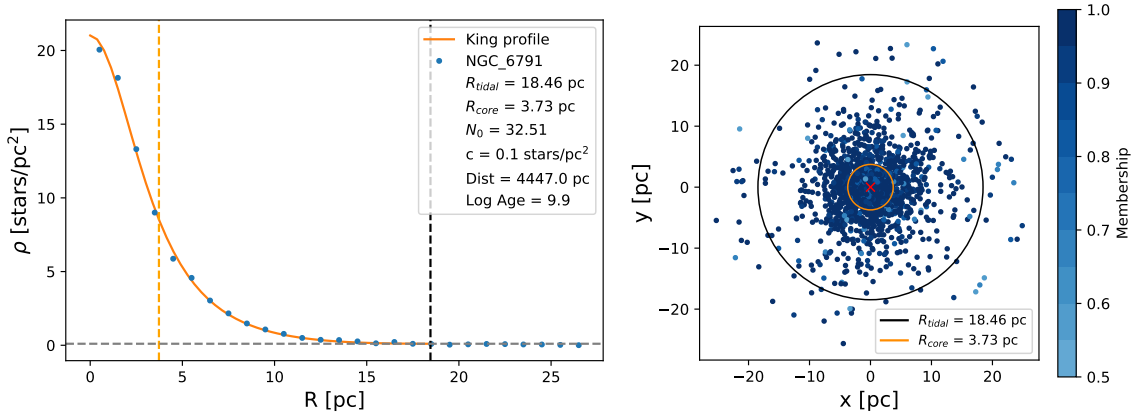

Figure 4.4: Example representative of a cluster with R1 classification. (Left) Spatial density distribution of open cluster NGC 6791, with $10^{9.9}$ years and distance of 4447 pc. (Right) Spatial distribution of the cluster's stars, color-coded by membership probability, with black circle at 18.46 pc (tidal radius) and orange circle at 3.73 pc (core radius).

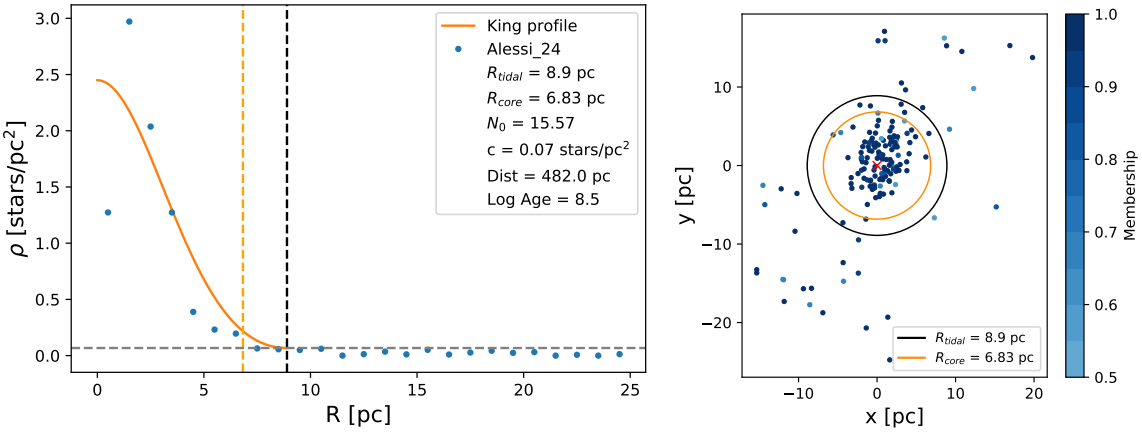

Figure 4.5: Example representative of a cluster with R2 classification. (Left) Spatial density distribution of open cluster Alessi 24, with $10^{8.5}$ years and distance of 482 pc. (Right) Spatial distribution of the cluster's stars, color-coded by membership probability, with black circle at 8.9 pc (tidal radius) and orange circle at 6.83 pc (core radius).

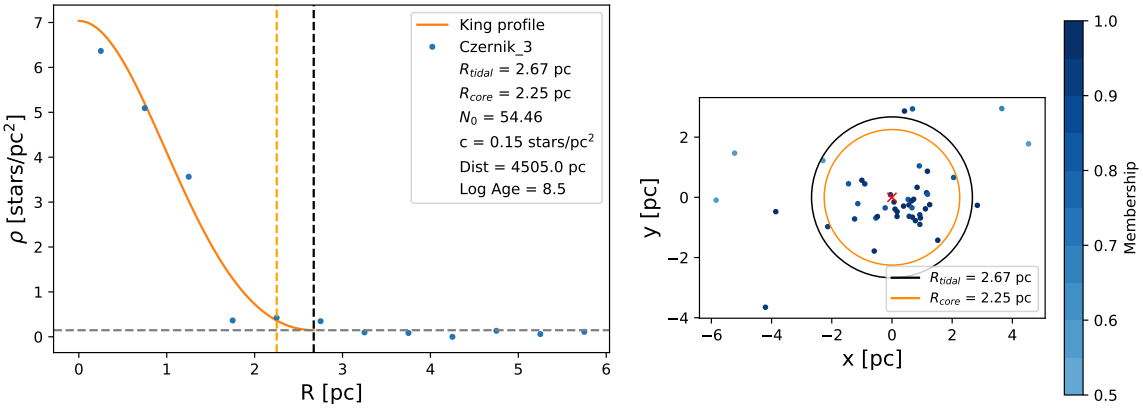

Figure 4.6: Example representative of a cluster with R3 classification. (Left) Spatial density distribution of open cluster Czernik 3, with $10^{8.5}$ years and distance of 4505 pc. (Right) Spatial distribution of the cluster's stars, color-coded by membership probability, with black circle at 2.67 pc (tidal radius) and orange circle at 2.25 pc (core radius).

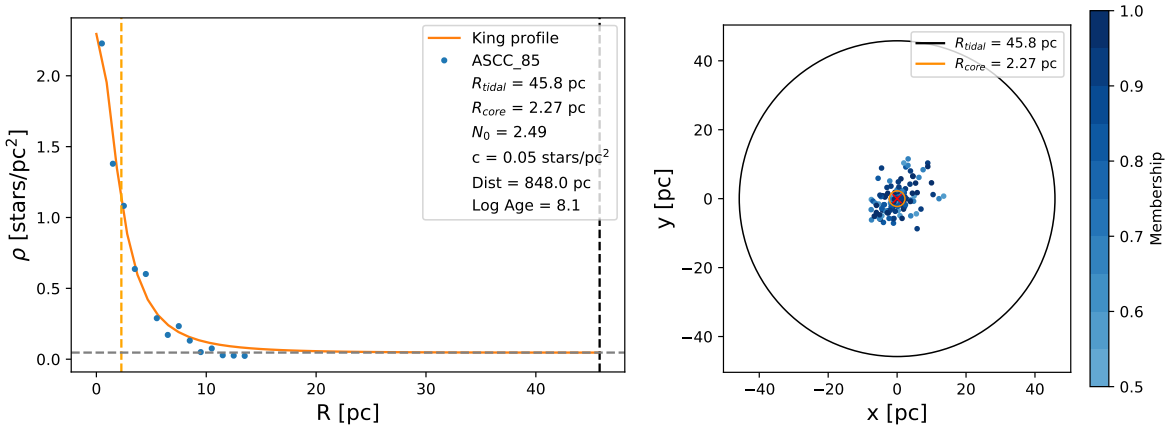

Figure 4.7: Example representative of a cluster with R4 classification. (Left) Spatial density distribution of open cluster ASCC 85, with $10^{8.1}$ years and distance of 848 pc. (Right) Spatial distribution of the cluster's stars, color-coded by membership probability, with black circle at 45.8 pc (tidal radius) and orange circle at 2.27 pc (core radius).

4.1.4 Sanity check

As a sanity check of the LMFIT procedure, we implemented a discrete loop over the tidal radius and obtained directly the χ^2 between the King profile and the observed density distribution. This was done by looping over R_t with steps of 0.5 pc and obtaining the χ^2 for each run. We allowed N_0 and c to vary continuously between the limits given in Table 4.1 as these parameters do not have restrictions. As for R_c, it varied within the limits given previously but, for the runs where R_t is smaller than the radial distance to the furthest star (max(R)), R_c was only allowed to go up to R_t. Using this "loop" method, we can access the values of R_t and R_c directly which is difficult to check inside LMFIT as it varies the parameters almost continuously inside the given ranges. In LMFIT, the combinations considered are not easily accessible so this "loop" method allows to perform a sanity check. We compared the results from this method to the results from LMFIT and verified that, for the high quality determinations (classification R1), the tidal radii have a median difference of 2%, which is an indication that these determinations are robust. For clusters with classification R2 and R3, the results show a higher difference (15% and 37%, respectively). Additionally, we also confirmed that most of the clusters that make up the "bump" in the tidal radii distribution around 50 pc have convergence problems in the loop as these always returned the maximum value chosen for the tidal radius in our sanity check loop.

4.1.5 Comparison with other studies

4.1.5.1 Comparison with Tarricq et al., 2021

The most recent and relevant catalogue of tidal radii is that of Tarricq et al. (2021) where stellar memberships were revisited for a sample of 389 open clusters in the solar vicinity using *Gaia* Early Data Release 3 (eDR3: Gaia Collaboration et al. (2021)). Several structural parameters were determined, including core and tidal radius, elongation and size of the halo using a new list of members extended to the outskirts of the clusters. Of the total sample, Tarricq et al. (2021) reports the determinations of tidal and core radii for 164 and 145 OCs, respectively. In our sample, we have 109 OCs under the same name and centre coordinates.

In Tarricq et al. (2021), the tidal radii were determined by fitting the King density profile with a Maximum Likelihood estimator using the MCMC sampler emcee (Foreman-Mackey et al., 2013). The method differs from ours in the sense that Tarricq et al. (2021) used 10,000 iterations per "walker" (in a total of eight "walkers") and discarded the results where a flag was risen by the integrated auto-correlation time whereas we used 100,000 iterations (with a single walker) for most of the clusters and increased that number when needed, as described in Section 4.1.1.

The authors implemented two criteria to discard the poorly constrained results by excluding clusters where the uncertainty in R_t was above 15 pc and above 2.5 pc for R_c. In our study, we filtered our results based on the quality of the density profile fit and not on the value of the uncertainty. If we apply the criteria used by Tarricq et al. (2021) to our sample, we report the determination of R_t and R_c for 467 OCs, an increase of almost 3 times.

Despite the similarities in our methods, Tarricq et al. (2021) expanded the number of stars considered in each cluster so, the number of stars considered is significantly different between our studies as seen in Figure 4.8 on the left. The ages used by Tarricq et al. (2021) come from the catalogue of Cantat-Gaudin et al. (2020) and were determined using a different method than in our reference catalogue, which explains the difference seen in Figure 4.8 on the right. As for the distance, the values are similar (only differing by about 5%).

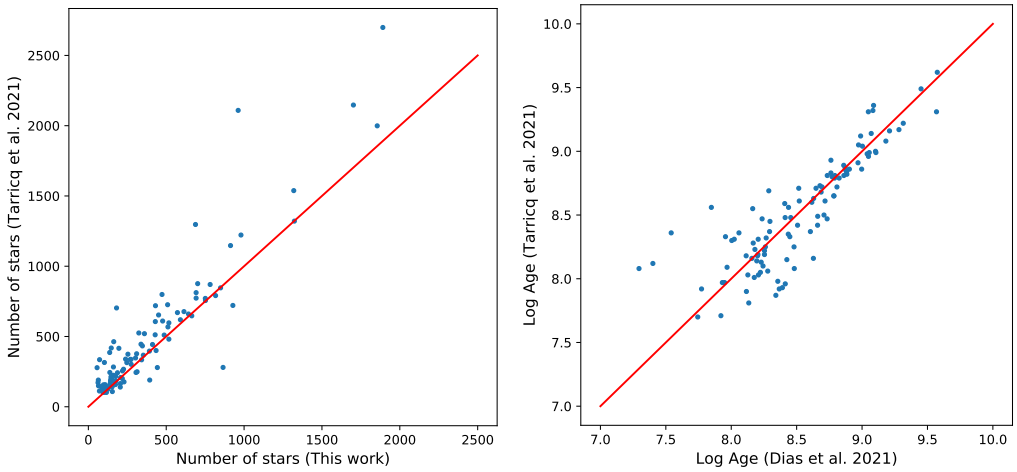

Figure 4.8: (Left) Number of stars used for each OC from this work compared to the number of stars used in Tarricq et al. (2021). (Right) Comparison of the age of each OC used in Tarricq et al. (2021) from the catalogue of Cantat-Gaudin et al. (2020) to the age from Dias et al. (2021).

The comparison between the core and tidal radii obtained by Tarricq et al. (2021) and in this study is presented in Figure 4.9, for the 109 OCs that are common in both studies. Our method, in general, leads to smaller values of R_t and similar values of R_c. This difference in the tidal radii distribution is also verified by Tarricq et al. (2021) when comparing their results to the distribution of tidal radii from Piskunov et al. (2007), Kharchenko et al. (2013), and Angelo et al. (2021).

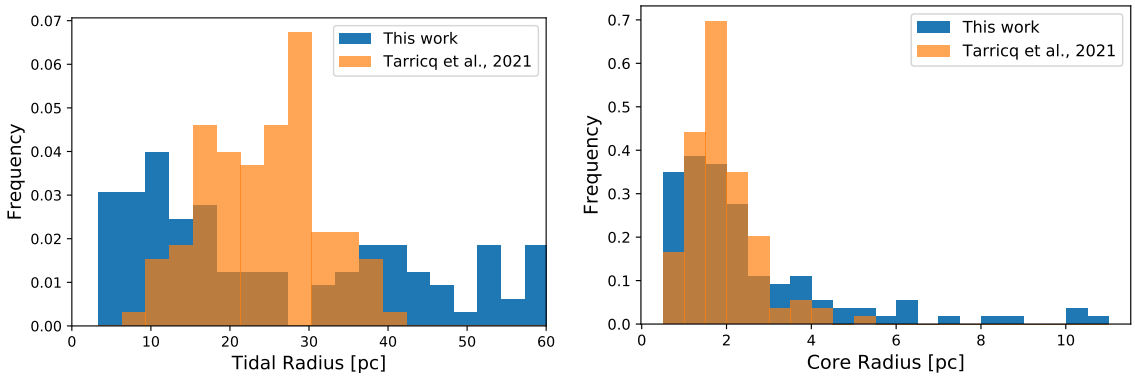

Figure 4.9: Comparison of the distribution of tidal (left) and core (right) radii from this work and from Tarricq et al. (2021), for the 109 OCs in common.

Regarding the uncertainties, since Tarricq et al. (2021) excluded clusters based on the value of the uncertainty, a direct comparison is not possible because we include in our study cluster with large uncertainties (with a quality flag). However, we notice that our results have, in general, similar uncertainties for the core radii and higher uncertainties for the tidal radii, when compared to the ones from Tarricq et al. (2021).

4.1.5.2 Comparison with Piskunov et al., 2007

Before the *Gaia* mission, the main catalogue of tidal radii was from Piskunov et al. (2007) (updated in Piskunov et al. (2008)). It contains tidal radii for 650 OCs from the ASCC 2.5 catalogue, determined by fitting the King profiles to the observed cumulative density distribution. In Figure 4.10, we compare the distance and age of each cluster used in this work to the values used by Piskunov et al. (2007).

Since they used pre-*Gaia* data, as expected, many clusters have different ages and distances in the two catalogues. However, for the determination of the tidal radii, this difference is not expected to impact the results. As seen in Figure 4.11, our results for the tidal radii are comparable with the ones from Piskunov et al. (2007).

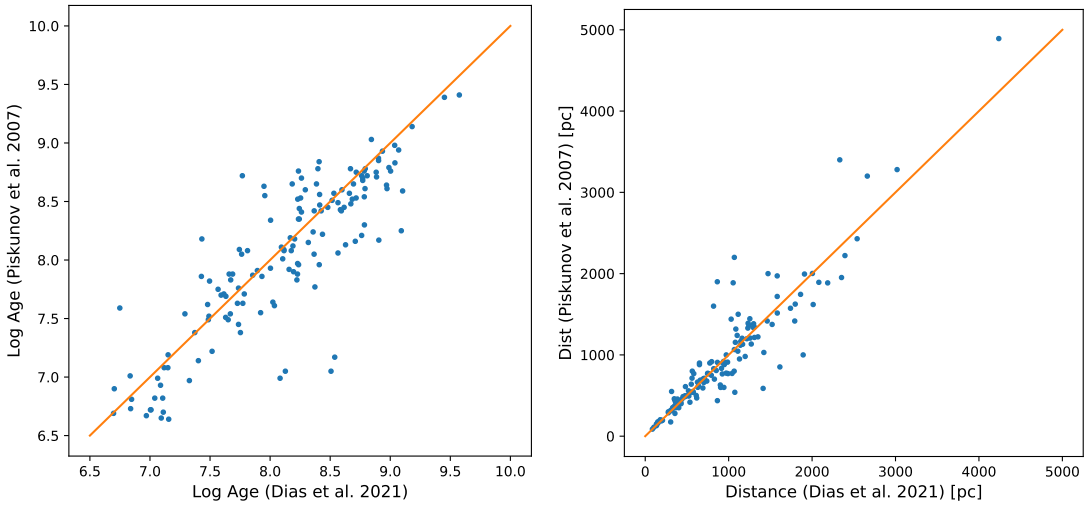

Figure 4.10: Comparison of the age (left) and distance (right) of each OC used in Piskunov et al. (2007) to the ones from Dias et al. (2021).

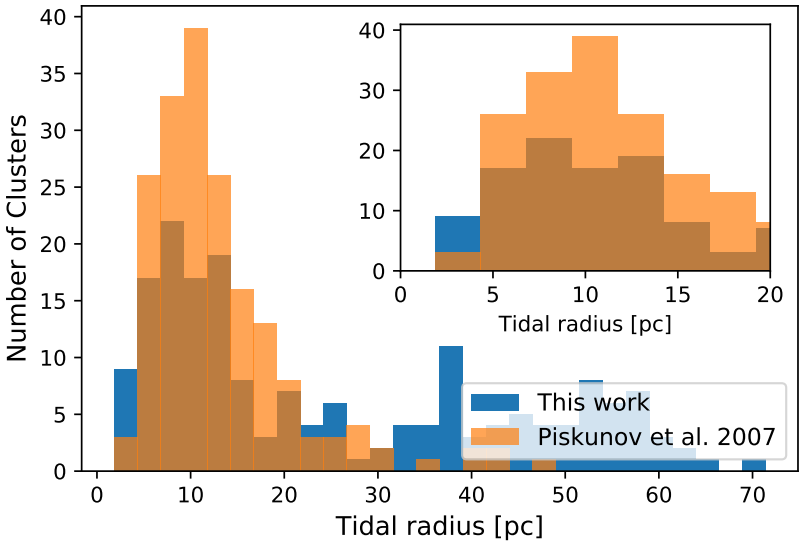

Figure 4.11: Distribution of tidal radii from this work and from Piskunov et al. (2007).

4.2 Determining luminous masses of clusters

With the tidal radii determined, we selected the stars within R_t in each cluster and obtained the luminous masses of the clusters, applying the methods described in Section 2.2 to 1724 open clusters. As mentioned, the bin width has been adjusted to avoid over or under sampling the luminosity distribution. For clusters with fewer stars, which we consider to be those that had less than 10 stars in the highest bin, we doubled the bin width to 1 magnitude. Then, we visually inspected them to confirm the magnitude distribution did not appear to be over or under sampled. In total, 1168 clusters remained with a bin width of 0.5 and 556 changed to 1.

For the membership probability cut-off, we took a conservative approach by considering only stars that are more likely to be from the cluster than from the field population. For this reason, we adopted the cut-offs determined through visual inspection of the CMD as described in Section 3.2.2. Note that we did not use the cut-offs from Section 4.1 which included some cut-offs lower than 0.5, which we were forced to adopt so that the method would converge. Despite this choice leads to clusters with tidal radius and mass determined using different numbers of stars, it should not influence our results because the tidal radius should not depend on the chosen membership cut-off, i.e., it should not depend on the number of stars considered.

4.2.1 Mass distribution

Considering the small adjustments described above, the distribution of mass (in linear and logarithmic scale) obtained for the G band, using the method of the least squares is shown in Figure 4.12, with a Kernel Density Estimation with bandwidth of 0.18 dex. The logarithmic distribution shows a peak at log(M) = 2.7 dex, with standard deviation of 0.4 dex. The linear mean mass is 797 $M_\odot$ and the median is 426 $M_\odot$. Details on why the least squares method was the adopted method are given in Section 4.2.3.2.

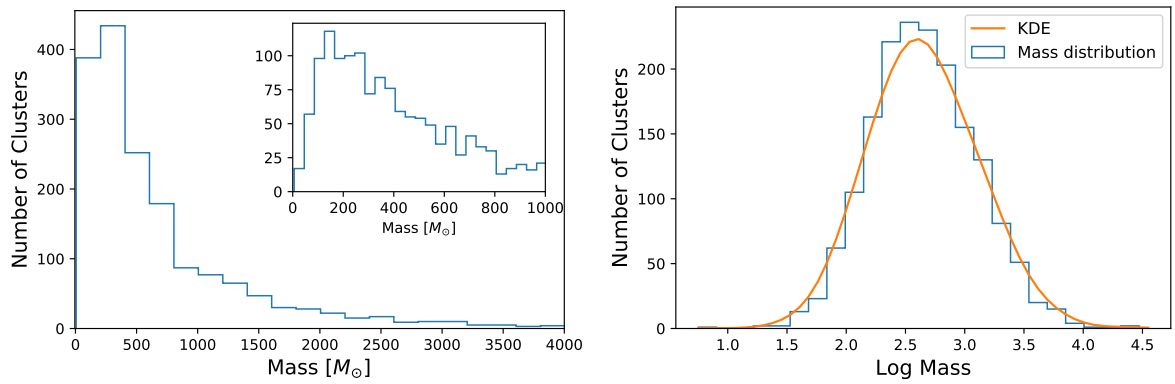

Figure 4.12: Distribution of the luminous mass obtained in this work, in linear (left) and logarithmic (right) scale. There are 34 clusters with mass above 4000 $M_\odot$ that we exclude from the left plot to allow an easier visualization of the results. On the right, the orange line is a Gaussian KDE with bandwidth of 0.18 dex.

4.2.1.1 Spatial distribution

The distribution of open clusters in the X-Z and Y-Z plane color-coded by mass is shown in Figures 4.13 and 4.14. The distribution in the galactic plane shows that clusters further away from the plane are older and more massive. It is known that clusters are usually born at lower latitudes where there is more gas and dust. Fewer clusters are formed at high galactic latitudes. Of those that can form at high

latitudes, since they suffer less disruption (due to the lack of GMCs and other clusters), they survive longer. In this older population of clusters outside the galactic plane, we expect to detect only the more massive ones because clusters naturally dissolve into the field so, the less massive ones will not persist until older ages.

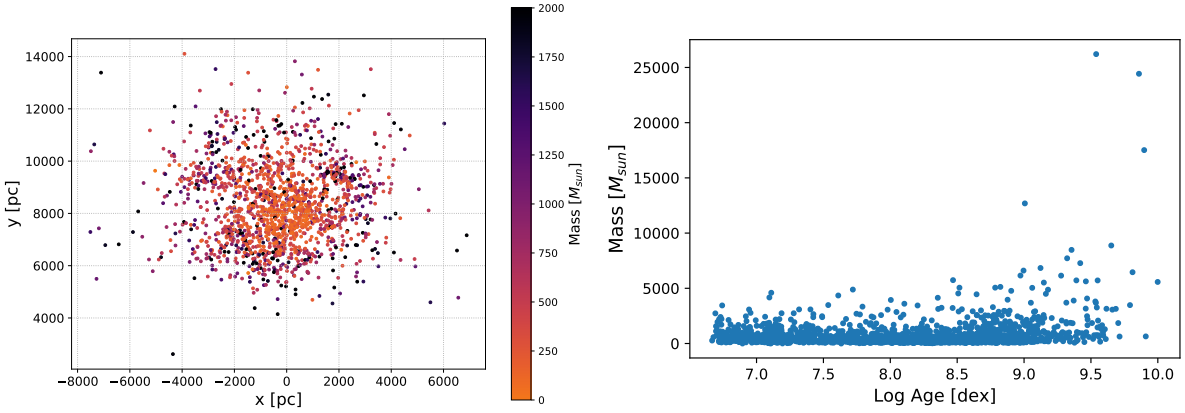

Figure 4.13: Distribution of OCs in the X-Y plane (left) color-coded by luminous mass and distribution of mass by age (right).

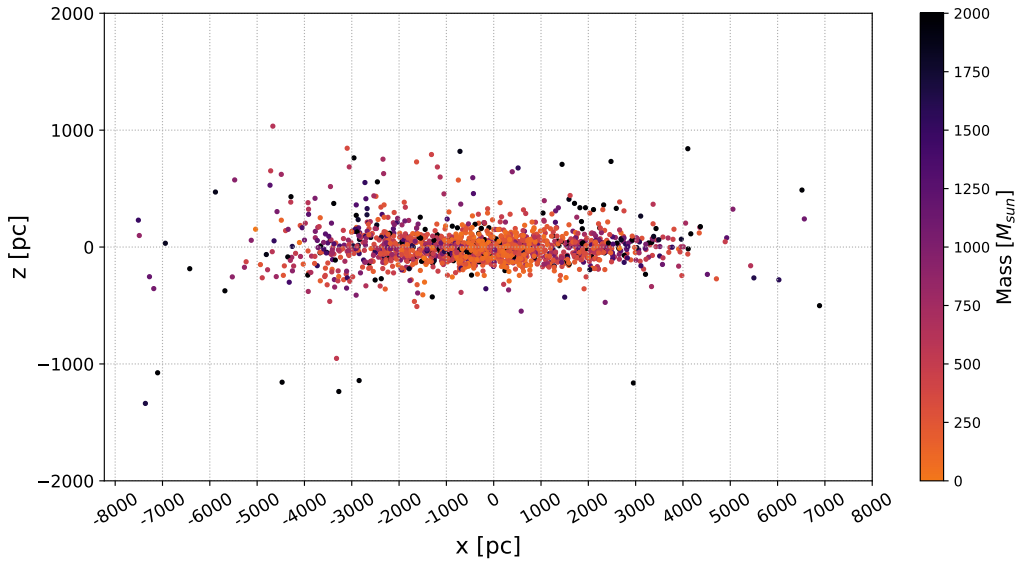

Figure 4.14: Distribution of open clusters in the X-Z plane color-coded by luminous mass.

4.2.2 Error analysis

4.2.2.1 Bootstrap error analysis

The results of the bootstrap error analysis, performed for 100 samples, are shown in Figure 4.15, for the G band. Most of the clusters have a mass error under 6% and the median value for the error is 4% with standard deviation of 3%. We note that these are the statistical uncertainties related to our mass determination method, not the true uncertainty of the mass of each cluster because the systematic uncertainty in the determinations of the age or distance of each cluster are not accounted for.

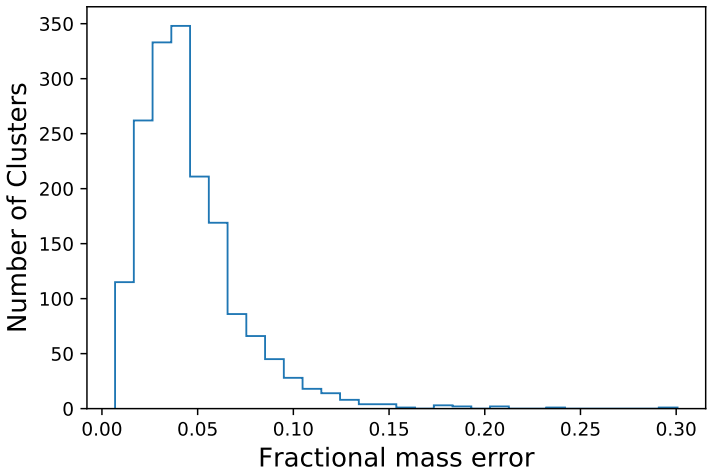

Figure 4.15: Distribution of the fractional mass error obtained through bootstrapping using 100 samples.

4.2.2.2 Effect of tidal radius uncertainty

To account for the uncertainty of the tidal radius and check the influence of that uncertainty in the value of the mass, we calculated the mass of each cluster considering the minimum and maximum value of R_t that corresponds to $R_t - \sigma$ and $R_t + \sigma$, respectively. As seen in Figure 4.16, the mass distribution considering only stars inside the minimum or the maximum value of R_t is very similar to the mass obtained considering R_t. In fact, the mass obtained using the minimum value of R_t is around 8% less than the mass inside R_t (represented as $M(R_t)$) and the mass inside the maximum R_t is around 6% more than $M(R_t)$. This proves that despite the large uncertainties in the tidal radius, it does not have a significant effect on the mass.

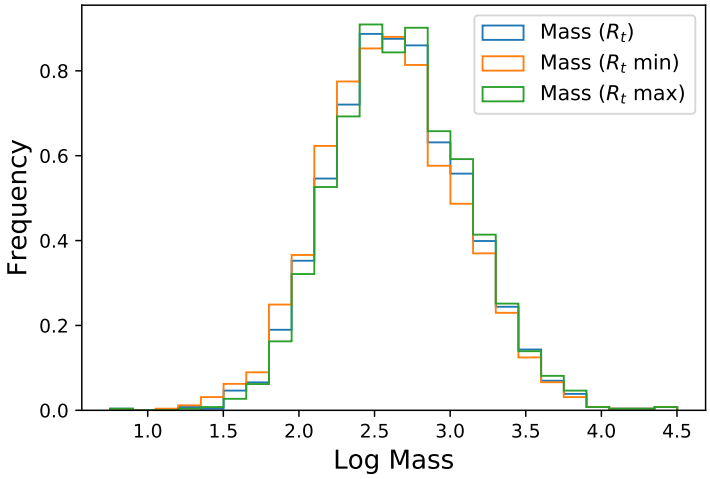

Figure 4.16: Distribution of luminous mass using only stars inside minimum R_t, R_t or maximum R_t.

The distribution of errors for the mass considering the lower and upper limits of R_t is similar to that of $M(R_t)$. The mass errors considering $M(R_t(min))$ and $M(R_t(max))$ show similar values to those of $M(R_t)$, with median difference of 4% and 3%, respectively.

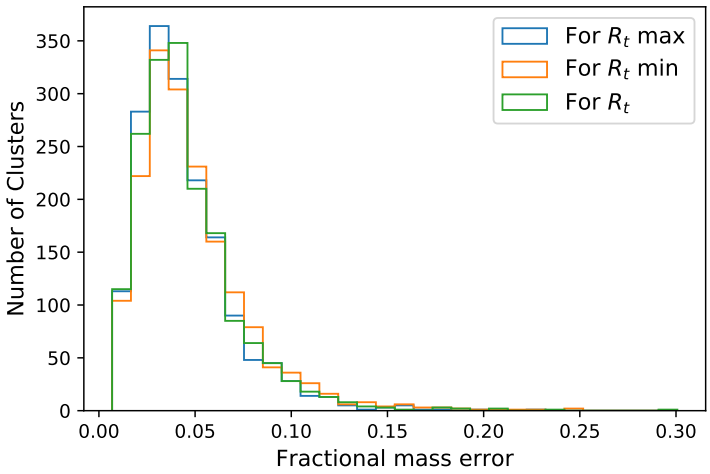

Figure 4.17: Fractional mass error for masses determined using only stars inside minimum R_t, R_t and maximum R_t.

4.2.3 Comparison between Gaia bands and mass determination methods

4.2.3.1 Comparison between G, BP and RP bands

To further assess the robustness of our mass determination method, we obtained the mass using photometric data from G_{BP} and G_{RP} *Gaia* bands, repeating the same process as described in the previous sections. Comparing those to our mass results using photometry from G band, using the least-squares method, we noted a normalized difference of 3% between the mass obtained using G and G_{BP} bands and 5% difference between G and G_{RP} bands, with standard deviation (1σ) of 19% and 20%, respectively. This indicates that our method is robust as there is, in general, a small difference between the mass determined using different bands. The values represent the normalized differences, in the sense of $\frac{M(G_{BP})-M(G)}{M(G)}$, as seen in Figure 4.18 on the left. The mean and standard deviation of the comparison for the mass calculated for different bands is presented in Table 4.4. It shows similar values for both methods however, for the second method (division of sums), there is a lower standard deviation for the difference of the masses obtained with different bands. This might be explained by the argument already mentioned: as this method only considers the sum of the counts, without weighting each bin, it is less sensitive to incompleteness or incorrect age or distance determinations, so it is less likely that the results will differ between bands.

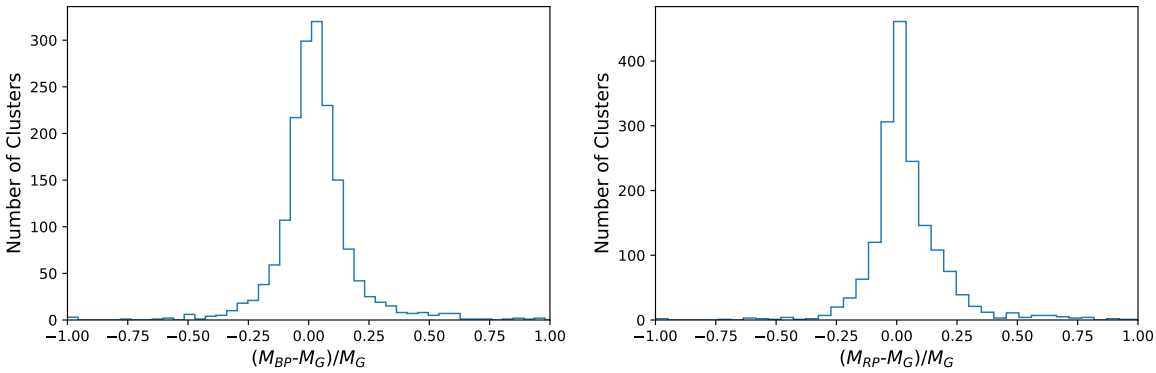

Figure 4.18: Normalized difference for masses determined using G and G_{BP} bands (left) and G and G_{RP} bands (right). We excluded from the figure 4 clusters on both plots with values above 1.

Table 4.4: Mean and standard deviation of the normalized difference between the mass determined using G and G_{BP}, G and G_{RP} bands, considering the least-squares method (ls) and the division of sums (div).

	Mean	Standard dev.
BP-G [ls]	3%	19%
RP-G [ls]	5%	20%
BP-G [div]	3%	13%
RP-G [div]	5%	14%

4.2.3.2 Comparison between mass determination methods

To further explore our results, we also compared the mass obtained using the two described methods: the least squares and the division of sums, for the same band. The comparison of the results from each method for the G band is shown in Figure 4.19. It has a mean mass difference of 6% (with standard deviation of 13%) between the two methods, with the least squares method giving generally lower estimates for the mass of the clusters.

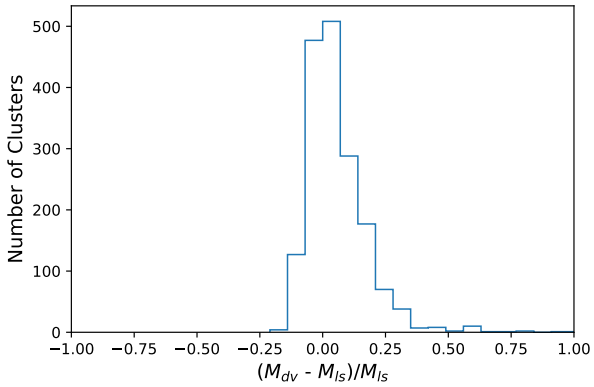

Figure 4.19: Normalized mass difference for masses determined using the method of division of sums and least-squares.

As the differences between the two methods and the three bands are consistent with the typical mass error of < 10%, we conclude that our mass determination method is robust. Given that the results from G_{BP} and G_{RP} bands were only obtained as a sanity check, only the mass determined using the G band should be considered. For the methods, both can be used but we opted for the least-squares as it allows a simpler determination of the error through bootstrap analysis and it is sensitive to the shape of the luminosity function.

4.2.4 Mass Classification

To assess the quality of the mass determinations, we classified the 1724 open clusters based on the quality of the agreement between observed and predicted luminosity distributions (called "LF fits" for simplicity). Several attempts were made to find a criterion that would allow a systematic classification of the LF fits. No criterion was found that allowed a strict classification, however, by using the χ^2 value of the fit, it was possible to loosely sort the clusters and then thoroughly classify, by eye, in three categories: M1 (best), M2 and M3 (worst), where M stands for mass.

For category M1 we selected the clusters which show a good agreement between the observed and predicted counts, on the 3 *Gaia* bands. In category M2 are the clusters for which the shape does not match well the predicted one or where there are not enough bins to give a reliable fit. The remaining clusters, for which the shape of the observed magnitude distribution is very different from the theoretical

one, were classified as M3 as those do not represent reliable results. An additional category (MX) was created for the clusters where the CMD was very poorly populated or the sequence undefined. In this category, we include all the OCs that had classification of P3.

Table 4.5: Number of clusters per classification according to the quality of the mass determination based on the quality of the agreement between observed and predicted luminosity functions.

Classification	Number of OCs
M1	812
M2	695
M3	65
MX	152

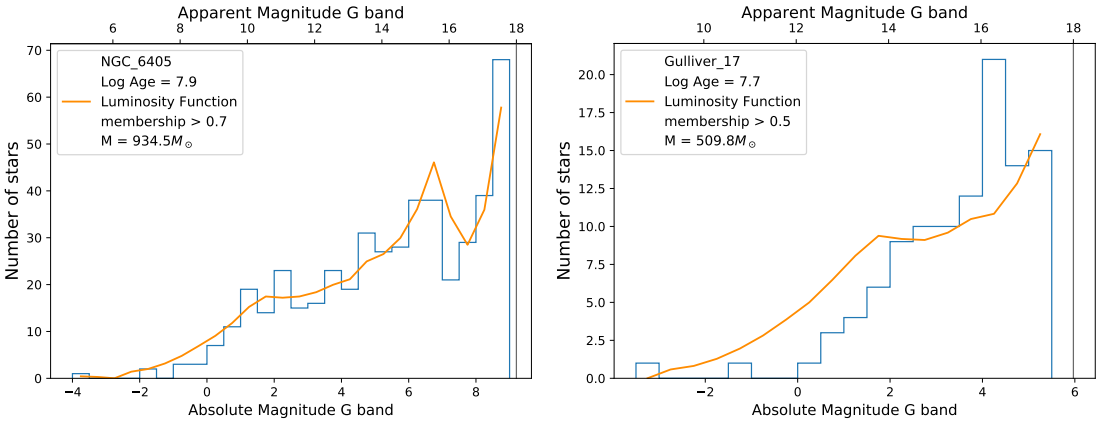

Figure 4.20: Absolute magnitude distribution for OCs NGC 6405 (left) and Gulliver 17 (right) with luminosity functions scaled to match the observed density distributions. NGC 6405 is representative of classification M1 and Gulliver 17 is classified as M2.

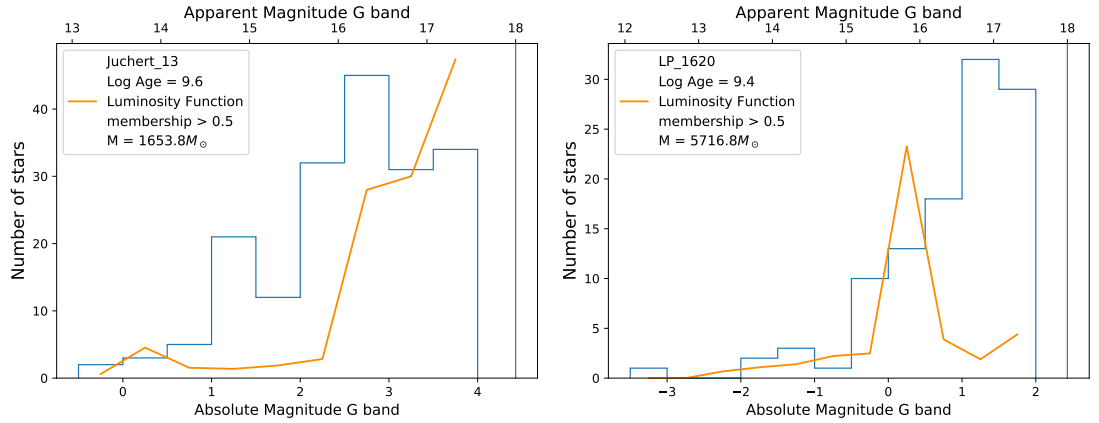

Figure 4.21: Absolute magnitude distribution for Juchert 13 (left) and LP 1620 (right) with luminosity functions scaled to match the observed density distributions. Juchert 13 is classified as M3 and LP 1620 as MX.

4.2.5 Comparison with Piskunov et al. 2007

Assuming that a cluster is in equilibrium and under the influence of the tidal field of the host galaxy, the cluster will experience forces that stretch it in opposite directions along the line towards the centre of the host galaxy. In such case, the tidal radius can be seen as the radius where the gravitational energy of the cluster equals the gravitational pull of the host galaxy. Piskunov et al. (2007) used the tidal radius to determine tidal masses for a sample of 650 open clusters. They obtained the tidal mass by using $M = (4A(A-B)r_t^3)/G$ where A and B are Oort's constants (Piskunov et al., 2006) with values 14.5 $\pm$ 0.8 km/s/kpc and B = -13.0 $\pm$ 1.1 km/s/kpc, respectively and G is the gravitational constant $(4.3 \times 10^{-3}(km/s)^2 pc\ M_{\odot}^{-1})$. The Oort constants express the rotational properties of galactic orbits in the solar neighbourhood, thus also encoding the (tidal) gravitational forces in this region.

As can be seen in Figure 4.22, even though the individual masses are not similar for most clusters, the overall distribution of masses follows the same shape.

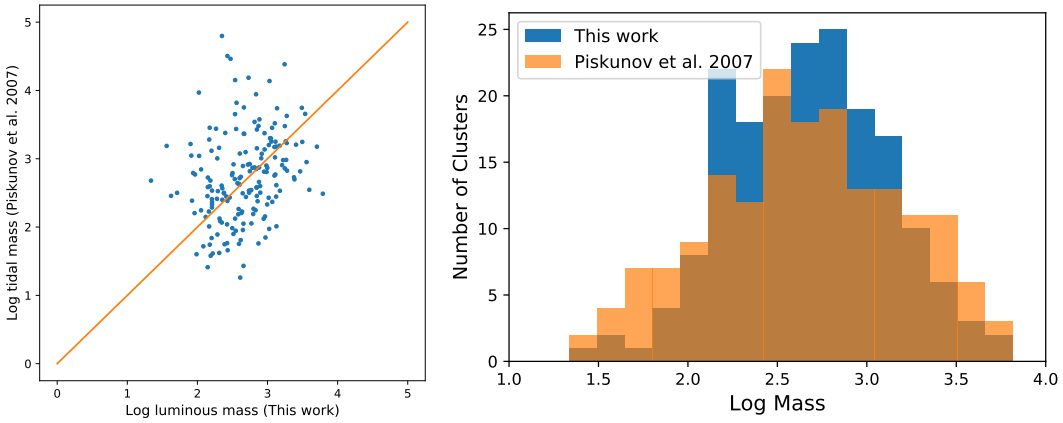

Figure 4.22: Comparison of tidal masses from Piskunov et al. (2007) to the luminous masses determined in this work.

4.3 Sample Selection

To avoid including unreliable results into the sample that we will use to explore the mass loss in open clusters, we selected 2 samples: a stricter high-quality sample, labelled as “gold sample” and another that is less strict and contains results of intermediate to high quality, labelled “silver sample”. The gold sample consists of clusters with only the best classification in the 3 categories: mass (M1), photometry (P1) and radius (R1). The silver sample includes clusters with radius classification R1 and R2, mass classification M1 and M2, and all photometric classifications. In total, the gold sample contains 153 OCs (9%) and the silver sample contains 713 OCs (41%). Note that the gold sample is contained inside the silver sample.

The distributions of age and distance of each sample are plotted in Figure 4.23 as well as their spatial distributions, in Figure 4.24. Both Figures 4.23 and 4.24 indicate that the silver sample is distributed in a way similar to the full sample. No apparent biases arise from the quality cuts to produce the silver sample. As for the gold sample, the number of clusters is manifestly small, displaying a much more limited spatial coverage. For those reasons, we decide not to consider the gold sample as the main sample for our analysis in the following sections, but we will still use it to check the robustness of the results that will be drawn from the silver sample.

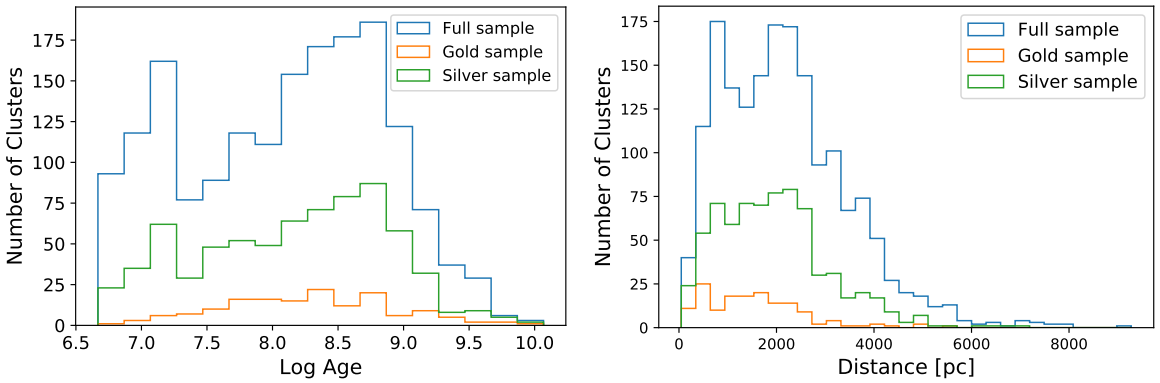

Figure 4.23: Distribution of age (left) and distance (right) for the full sample (blue), gold sample (orange) and silver sample (green).

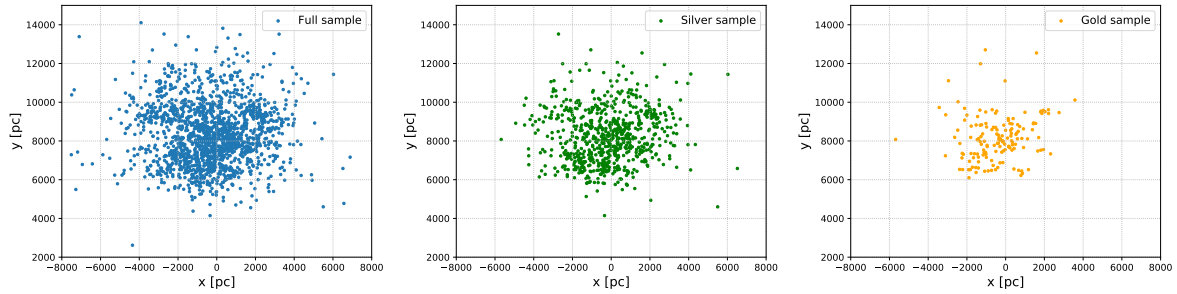

Figure 4.24: Spatial distribution of the full sample (left), silver sample (middle) and the gold sample (right) in the X-Y plane of the Galaxy.

4.3.1 Spatial Completeness

To investigate the spatial completeness, we looked at the density of our silver sample for OCs with ages under 1 Gyr. This age limit was chosen to check the completeness of the sample from where we obtain the mass distribution that will be used to investigate the disruption timescale in the solar neighbourhood.

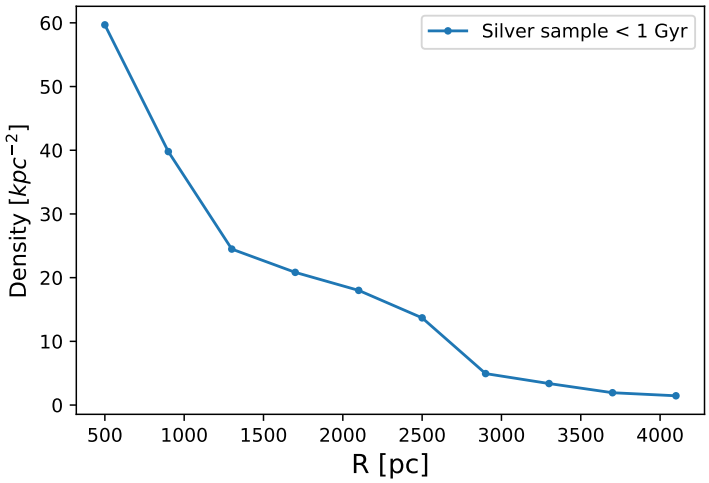

Figure 4.25: Density of the OCs in the silver sample with age under 1 Gyr as a function of the radial distance from the Sun.

In Figure 4.25, we present the density inside concentric rings (with steps of 400 pc) as a function of the distance. Under the assumption that our position in the Galaxy is not special, the density should remain constant in a complete sample. But, as seen in the plot, the density is not constant and decreases with the distance and this is also verified for the full sample. This is somewhat unexpected as some authors have claimed that the sample of known OCs in the solar neighbourhood is complete within 1.8 kpc around the Sun (e.g. Joshi (2005), Bonatto et al. (2006), Buckner and Froebrich (2014), and Kharchenko et al. (2013)).

To conduct a more thorough investigation, we divided the silver sample into 3 subsamples. We separated the clusters by age as young (log(age) $\leq$ 8), intermediate (8 < log(age) < 8.6) and old clusters (log(age) $\geq$ 8.6). In Figure 4.26, we present the spatial distribution of each age subsample with a black circle at 2 kpc. At young ages, the clusters are still near their birthplace in the spiral arms so, their spatial distribution presents a clustered structure around the spiral arms (e.g. Dias and Lépine (2005) and Cantat-Gaudin et al. (2020)). At intermediate and higher ages, the structure is less visible and the distribution appears more homogeneous. However, as seen in Figure 4.27, for every age subsample, the density decreases with the distance.

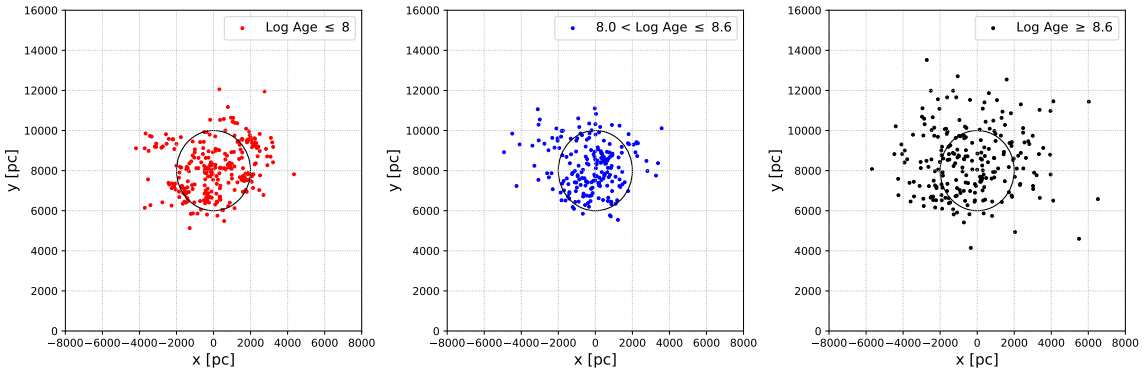

Figure 4.26: Spatial distribution of the silver sample divided into 3 age subsamples with black circle at 2 kpc.

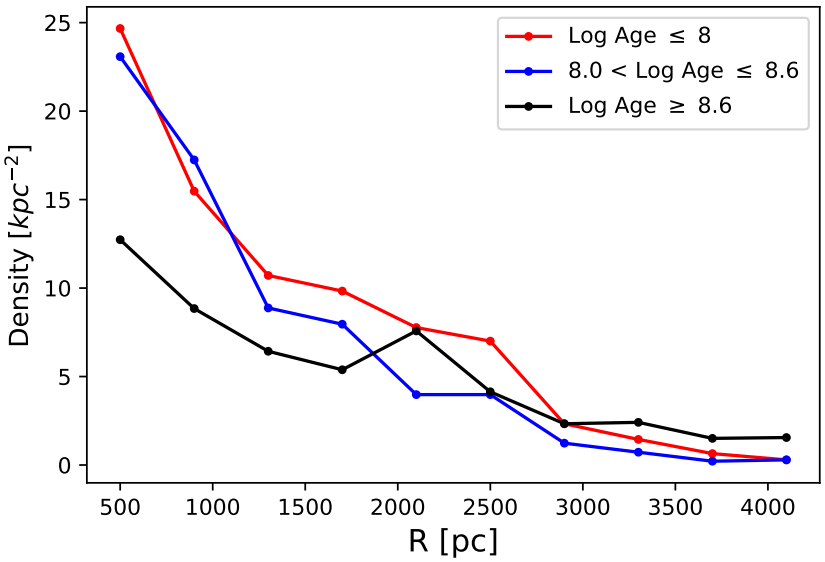

Figure 4.27: Density distribution of each age subsample from the silver sample.

Even though Kharchenko et al. (2013) claims that the Milky Way Star Cluster (MWSC) catalog is complete up to a distance of 1.8 kpc, Joshi et al. (2016), who use that catalog, report a decrease of density for clusters with ages under 700 Myr (log(age) < 8.8). Such decrease is similar to our results from the Dias et al. (2021) catalog. This raises questions on the claims of sample completeness.

For our younger groups, the clumpy structures and irregular density profiles make it hard to assess completeness. However, given the large variations seen in the density profile of each age subsample, the density appears to decrease similarly for every age. This indicates that the selection effects introduced are similar at every age but further investigation on this subject is planned as future work. From the spatial distributions, it is possible to establish a 2 kpc limit after which the distributions are visibly incomplete (at young and intermediate ages). For this reason, we adopt 2 kpc as the spatial limit for our silver sample.

For the gold sample, it is not possible to assess the completeness as there are very few OCs to draw conclusions from. For this reason, we only checked the completeness for the silver sample and assume the same limit for the gold sample. Cleaner CMDs for the clusters are needed to allow more robust determinations of cluster parameters that will increase the size of the gold sample and allow a better completeness analysis.

4.4 Disruption Timescale

As already mentioned in Section 1.3, stellar clusters lose mass through several processes of disruption. In our study, we do not differentiate between the different disruption mechanisms, in the sense that our disruption analysis includes the combined effects of evaporation by two-body relaxation, tidal stripping by the host galaxy and GMC encounters.

In a cluster, the fraction of mass that remains at a certain time t, can be defined as $\mu(t) \equiv M(t)/M_i$, where M_i is the cluster initial mass. Lamers et al. (2005b) considered that the mass loss is a combination of stellar evolution and disruption in such way that $dM/dt = (dM/dt)_{ev} + (dM/dt)_{dis}$ where "ev" stands for stellar evolution and "dis" stands for disruption.

In our model, for the mass loss by stellar evolution, we use the equations derived by Lamers et al. (2005b) who parameterise it as:

$$log\ q_{ev}(t) = (log\ t - a_{ev})^{b_{ev}} + c_{ev},\ for\ t > 12.5\ Myr \tag{4.1}$$

with $q_{ev} \equiv (\triangle M)_{ev}/M_i$ where $(\triangle M)_{ev}$ is the mass lost by stellar evolution and $\mu_{ev}(t) = 1 - q_{ev}(t)$ is the fraction of the initial mass that remains bounded to the cluster at age t, considering only stellar evolution. Values for a_{ev}, b_{ev} and c_{ev} are listed in Lamers et al. (2005b) for different metallicities. As we wish to analyze the mass loss experienced by clusters in the solar neighbourhood, we adopt the values of a_{ev} = 7, b_{ev} = 0.255 and c_{ev} = -1.805, which are valid for the solar neighbourhood.

Regarding the disruption, Lamers et al. (2005b) considered that the disruption time can be defined as $t_{dis}^{-1} = dln(M)/dt$ which describes an exponential decay with a timescale that increases as the mass of the cluster increases. In fact, early works of Boutloukos and Lamers (2003) showed that there is a relation between the disruption time and the mass of a cluster as $t_{dis} = t_0(M/M_\odot)^\gamma$. In this expression, t_0 is the disruption timescale, which depends on the tidal forces of the host galaxy, so it varies between galaxies and is dependent on the position of the cluster in relation to the galaxy. γ describes the mass-dependence of the disruption time and depends on the concentration of the stellar distribution in the cluster. Boutloukos and Lamers (2003) used the age distribution from Wielen (1971) and found γ = 0.6 for clusters within 1 kpc of the Sun. Baumgardt and Makino (2003) performed N-body simulations of clusters in the tidal field of our galaxy at different galactocentric distances, taken into account the mass loss by stellar evolution and by tidal relaxation. Their results show the same mass dependence for the disruption time with γ = 0.62. This allowed to obtain the expression for the mass loss due to disruption for which Lamers et al. (2005b) found an approximated solution that describes the decrease of mass as:

$$\mu(t;M_i) \equiv \frac{M(t)}{M_i} \simeq \left\{ (\mu_{ev}(t))^\gamma - \frac{\gamma t}{t_0}\left(\frac{M_\odot}{M_i}\right)^\gamma \right\}^{1/\gamma} \tag{4.2}$$

This approximation provides excellent agreement (within 0.015 dex) to the numerical solution for 0 < γ < 1, $10^3 \leq M_i/M_\odot \leq 10^6$ and t_0 = [2,30] Myrs.

Using these expressions, Lamers et al. (2005b) compared the predicted age distribution obtained by integrating the previous expressions to the observed age distribution of open clusters in the solar neighbourhood, from the Kharchenko et al. (2005) catalogue, and obtained $t_0 = 3.3^{+1.5}_{-1.0}$ Myr and γ = 0.62. It was pointed out by the authors that ideally one would use both the mass and age distributions to compare against predictions however, a mass catalogue was not available at the time. This has now changed with *Gaia*, which has allowed us to build the catalogue of masses presented in Section 4.2. It is now possible to use the observed mass distribution for constraining the modelled disruption timescale of open clusters in the solar neighbourhood.

In this study, we made use of the code developed by our group at CENTRA-FCUL (Silva (2016) and Moreira (in prep.)) that simulates the build up and mass evolution of a population of OCs along time. The model generates clusters at a constant rate (10 clusters at each timestep of 0.2 Myr), with initial masses drawn from a Cluster Initial Mass Function (CIMF), and includes the cluster mass loss as prescribed in Equation 4.2. The CIMF used, which describes the number of clusters that is initially formed of each mass, is the one universally adopted determined using embedded clusters, which is described with a power law of the type (Lada and Lada, 2003; Gieles, 2009):

$$\frac{dN}{dM} = M_{min} M^{-\alpha} \tag{4.3}$$

with $\alpha \sim 2$, $M_{min} = 100\ M_{\odot}$ and $M_{max} = 3 \times 10^4\ M_{\odot}$ as suggested by Lamers and Gieles (2006).

The model has two parameters, t_0 and γ, that are varied in the simulations, producing different cluster mass and age distributions to be compared to the observations. By changing t_0 and γ in the model, so that the age and mass distribution of the generated clusters matches the observed distributions, in the solar neighbourhood, we obtain the disruption timescale (t_0) and γ. In our work, as clusters are within a few kpc around the Sun, we assume the same tidal influence for all clusters, independent of their location in the galaxy. In the future, we plan to lift this assumption by considering the tidal field dependence on the location. We also do not consider the differences in the stellar concentration in the clusters, so our parameter γ will be the same for all clusters.

To quantify the quality of the agreement between the model and observations for the mass and age distributions, the root mean square (RMS) between the observations and normalized predictions was calculated. The normalization was done so that the total number of clusters in the simulation is the same as the number of observed open clusters. This means that the code works for any cluster formation rate and the value of 10 clusters per timestep was an arbitrary choice that allowed to have enough simulated clusters to make a robust comparison.

4.4.1 Results with power law CIMF

We ran our simulations for 1 Gyr and compared the final distribution of age and mass of the simulated clusters to the distribution of clusters from the silver sample with age $\leq$ 1 Gyr. We chose to run our simulations for 1 Gyr since, for older ages, the OC sample in the solar neighbourhood is known to be incomplete (Moitinho (2010), Kharchenko et al. (2013), and others). We also limit our sample to a sphere of 2 kpc around the Sun as we consider that to be the completeness limit of our silver sample, as discussed in Section 4.3.1.

Taking the values of t_0 and γ from Lamers et al. (2005a) as initial references, we ran simulations for a large range of these parameters on a grid of 8x5 with t_0 between 0 and 35 Myrs (steps of 5 Myr) and γ between 0.2 and 1.0 (steps of 0.2). The initial value of t_0 was replaced by 1 Myr because the equations used are not valid for $t_0 = 0$.

In Figure 4.28, we present the RMS of the difference between the simulated and observed age and mass distributions (top panels). The RMS considering the combination of the age and mass distributions was obtained by multiplying the RMS from the age to the RMS from the mass fit and is displayed in Figure 4.28 on the bottom. In the green filled area, we represent the contour of the 1.5σ region, considering σ as the minimum value of the RMS, with a smoothing factor of 3. This gives an idea of the optimal area for the parameters t_0 and γ.

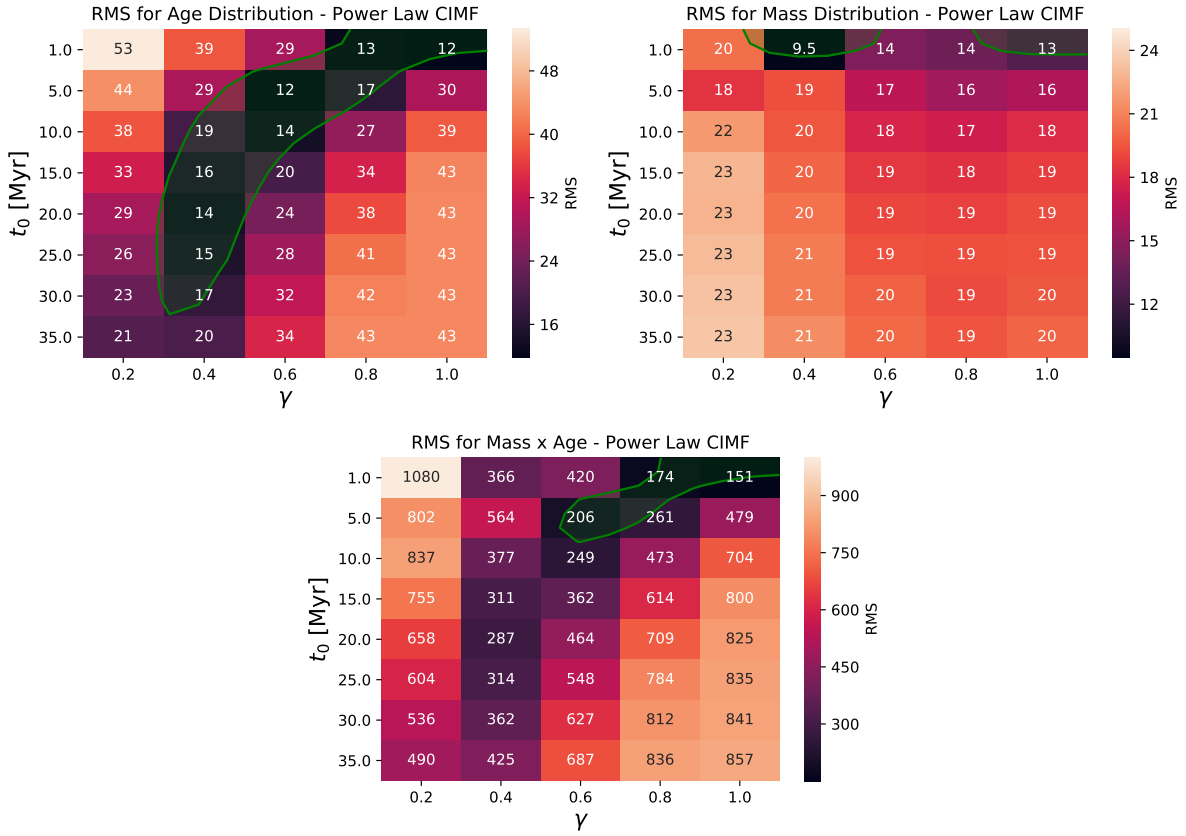

Figure 4.28: Heatmaps of the RMS between the simulated and observed age (top left), mass (top right) and combination of age and mass (bottom) distributions for the silver sample within 2 kpc around the Sun. The green area represents the contour of the 1.5σ region, considering σ as the minimum value of the RMS.

Notice that considering the age distribution (top left panel of Figure 4.28) we detect an optimal parameter range inside a curved shape region for γ between 0.4 and 1 and t_0 between 1 and 30 Myr. The best value is at $\gamma = 0.6$ and $t_0 = 5$ Myr, consistent with the values suggested by Lamers et al. (2005b) and also for $t_0 = 1$ Myr and $\gamma = 0.8$ and 1, which seems to be a degenerate area that extends into the upper right part of the grid (for lower t_0 and higher γ). However, using the mass distribution (top right panel of Figure 4.28) it is not possible to isolate an optimal range for t_0 and γ but the heatmap shows that changing the value of γ leads to significant differences in the agreement between observed and simulated mass distributions with higher values of γ being favoured, i.e., the quality of the agreement increases for higher values of γ (for a fixed t_0). The disruption timescale t_0 has a minor effect on the simulated mass distribution but lower values of t_0 are favoured (for the same γ).

When combining the two effects, the optimal set of parameters is within two regions. One at $t_0 = 5$ Myr and $\gamma = 0.6$ and other at $t_0 = 1$ Myr and γ of 0.8 and 1. However, although these two regions provide the best fit inside this grid, the observed and simulated mass distributions do not show a good match, as seen in Figures 4.29 and 4.30.

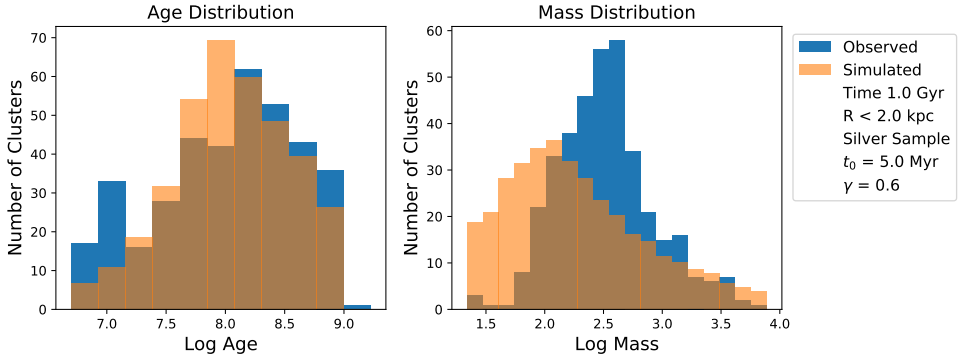

Figure 4.29: Comparison of the distribution of age (left) and mass (right) between the simulations (orange) and the observations (blue) considering $t_0 = 5$ Myr and $\gamma = 0.6$, using the silver sample within 2 kpc around the Sun.

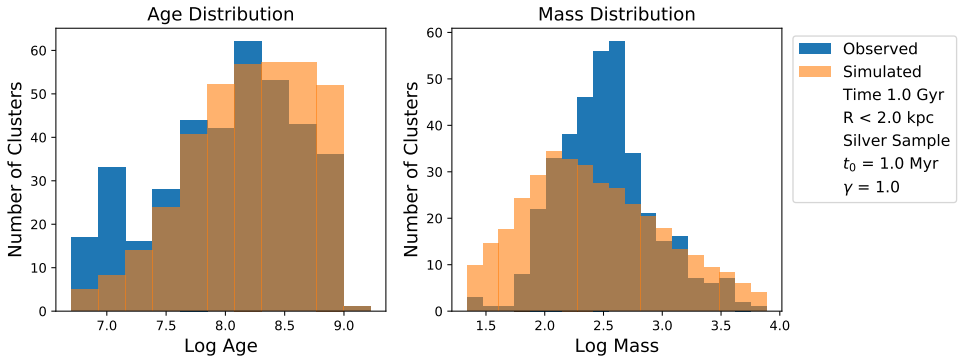

Figure 4.30: Comparison of the distribution of age (left) and mass (right) between the simulations (orange) and the observations (blue) considering $t_0 = 1$ Myr and $\gamma = 1$, using the silver sample within 2 kpc around the Sun.

In fact, using the values found in the literature (determined using the age distribution only) of $\gamma = 0.62$ and $t_0 = 3.3$ Myrs (Lamers et al., 2005a), the observed mass distribution does not match the simulated mass distribution (Figure 4.31). The simulated distribution peaks at lower masses than the observations for any value of t_0 in the intervals considered in the literature. This incompatibility came as a surprise and was not suspected in previous works (e.g. Lamers and Gieles (2006) and Lamers et al. (2005a)). It has been revealed by our newly determined masses and could be indicating that additional physics needs to be considered. To explore the origin of this mass distribution discrepancy, we consider, in the next section, the possibility that the CIMF of non embedded clusters (i.e., clusters that survived emerging from their parent molecular cloud) may be different from the CIMF of embedded clusters, which is the one universally adopted.

The heatmaps of the RMS determined using the silver sample but only within 1.5 kpc is shown in Figure 4.32. Since the RMS was calculated using a different number of bins, comparing the value of the RMS directly with Figure 4.32 is not possible but the results show, in general, the same areas for the optimal parameter combinations.

These results are also verified when using the gold sample instead of the silver sample, considering the 2 kpc radial limit (Figure 4.33).

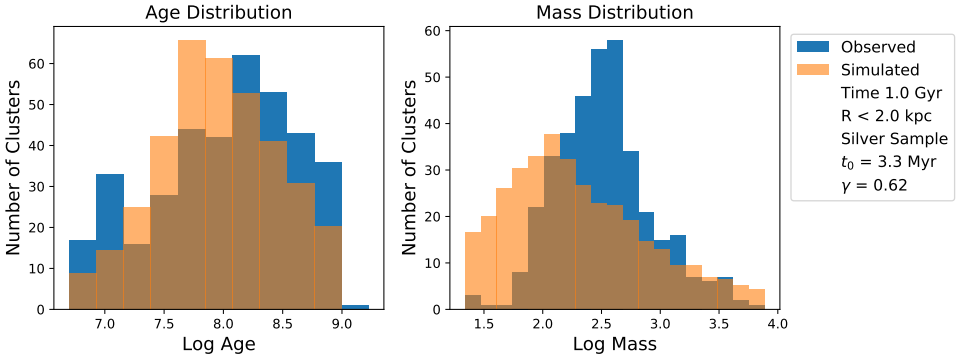

Figure 4.31: Comparison of the distribution of age (left) and mass (right) between the simulations (orange) and the observations (blue) considering $t_0 = 3.3$ Myr and $\gamma = 0.62$ (Lamers et al., 2005), using the silver sample within 2 kpc around the Sun.

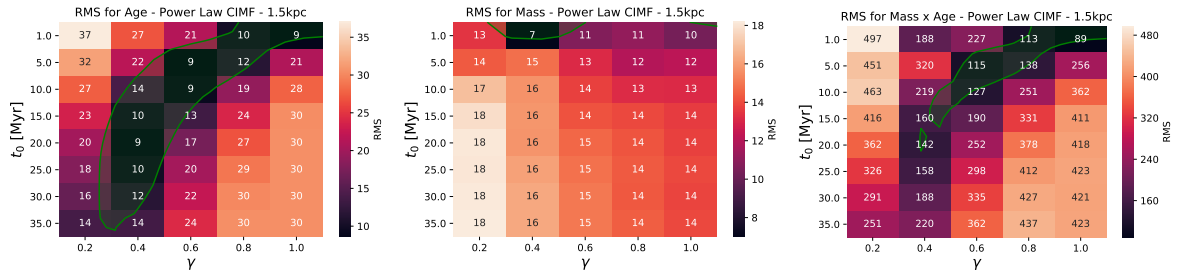

Figure 4.32: Heatmaps of the RMS between the simulated and observed age (left), mass (center) and combination of age and mass (right) distributions for the silver sample within 1.5 kpc around the Sun. The green area represents the contour of the 1.5σ region, considering σ as the minimum value of the RMS.

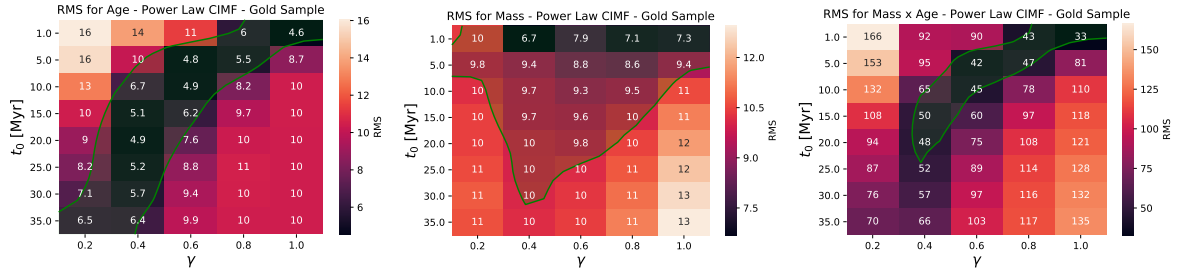

Figure 4.33: Heatmaps of the RMS between the simulated and observed age (left), mass (center) and combination of age and mass (right) distributions for the gold sample within 2 kpc around the Sun. The green area represents the contour of the 1.5σ region, considering σ as the minimum value of the RMS.

4.4.2 Modified CIMF

For the purpose of this work, the CIMF employed in the simulations should reflect the mass distribution of OCs just after emerging from their molecular clouds. This is when our simulations start. We note that our model does not include the mass loss and disruption that happen with the emergence process. However, the CIMFs found in the literature and used in previous similar works (e.g. Lamers et al. (2005a)) as well as in the previous sections have been determined for embedded clusters. While determining the CIMF of OCs is beyond the scope of this work, we inspect the mass distribution of young OCs from the silver sample within 2 kpc (complete and high quality) to get some insight.

We considered the logarithmic mass distribution for OCs with age ≤ 20 Myr and 50 Myr but, as the sample with OCs younger than 20 Myr has only 59 clusters, we opted to use only the sample with ages ≤ 50 Myr, which has 108 clusters. The shape of the logarithmic mass distribution appears to roughly follow a gaussian shape, so we adopted a heuristic approach by fitting a Gaussian with Equation 4.4. We

obtained x_0 and σ by taking the mean and standard deviation of the logarithm of the mass distribution.

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x_0-\mu}{\sigma})^2} \tag{4.4}$$

The values of x_0 and σ obtained are 2.5 and 0.4, respectively. Using this Gaussian functional as the CIMF in our model instead of the previous power law shape is expected to have a visible impact in the results. Using this Gaussian toy CIMF translates into drawing a random sample from the Gaussian, and then convert those logarithmic values into linear mass. That will be the initial mass of each generated cluster. Our choice to use a Gaussian is meant as a quick test to check the results. Other distributions (for example, log-normal) could have been used and we plan to explore, in depth, the CIMF as future work.

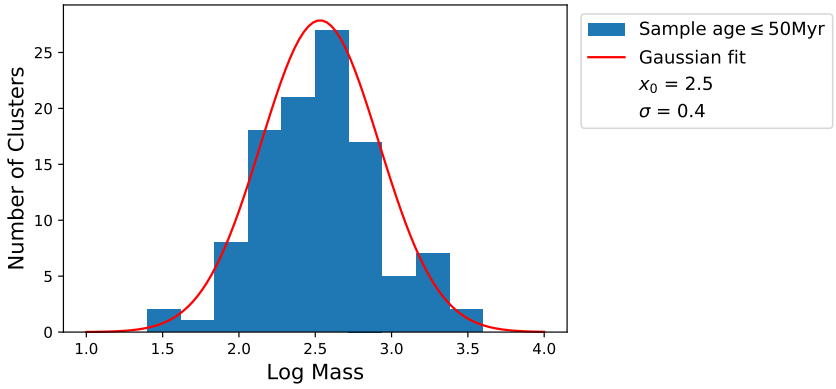

Figure 4.34: Distribution of the logarithm of mass for OCs with age under 50 Myr with fitted Gaussian with x_0 = 2.5 and σ = 0.4.

4.4.2.1 Effect of Spatial Completeness on the modified CIMF

We explored the effect of spatial completeness in the values of x_0 and σ by determining the Gaussian parameters using OCs within 1, 1.5 and 3 kpc to compare to the values determined using OCs within 2 kpc. The sample of OCs under 50 Myr within 1 kpc contains very few clusters so it's not possible perform to fit a Gaussian to the mass distribution. As for the 1.5 and 3 kpc, as shown in Table 4.6, the value of x_0 increases with the size of the sample. The larger sample (3 kpc) may suffer from incompleteness, therefore artificially decreasing the proportion of low mass clusters, which increases the mean mass. On the smaller sample (1.5 kpc) there is a smaller number of high mass clusters as we are sampling a smaller portion of space, leading to a lower mean mass. This indicates that a compromise is necessary so we adopt the values determined using OCs within 2 kpc.

Table 4.6: Values of the fitted Gaussian parameters (x_0 and σ) for OCs with age under 50 Myrs in the silver sample, considering a radial distance of 1.5, 2 and 3 kpc around the Sun.

	1.5 kpc	2 kpc	3 kpc
x_0	2.4	2.5	2.6
σ	0.4	0.4	0.4

4.4.3 Results with modified CIMF and comparison with power law CIMF

The comparison between simulated data and observations using the modified CIMF is seen in Figure 4.35 for the optimal t_0 and γ values obtained previously (t_0 = 5 Myr and $\gamma = 0.6$). When compared with Figure 4.29, it shows that the peak of the mass distribution is slightly closer to the peak of the observed mass distribution and the age distribution peaks for the same age as the observations.

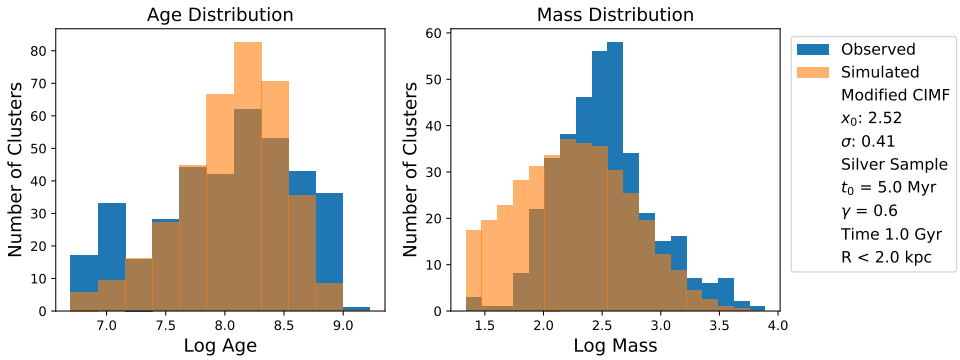

Figure 4.35: Comparison of the distribution of age (left) and mass (right) between the simulations (orange) and the observations (blue) considering t_0 = 5 Myr and $\gamma = 0.6$, using the silver sample within 2 kpc around the Sun and a modified CIMF.

To obtain the optimal parameter combination considering the modified CIMF, we ran the whole grid of simulations with the new CIMF. As shown in Figure 4.36, the new results show a better agreement between simulated and observed mass distributions, confirmed by the lower RMS for each set of t_0 and γ in the top right panel. The effect of changing γ is much more pronounced in the heatmap of the RMS from the mass distribution using the modified CIMF than with the power-law CIMF. When combining both the age and mass distributions, we recover the same optimal areas as previously seen.

The age and mass distributions for the optimal combinations (t_0, γ) = (5, 0.6); (1, 0.8) and (1, 1) are presented in Figure 4.35 and 4.37.

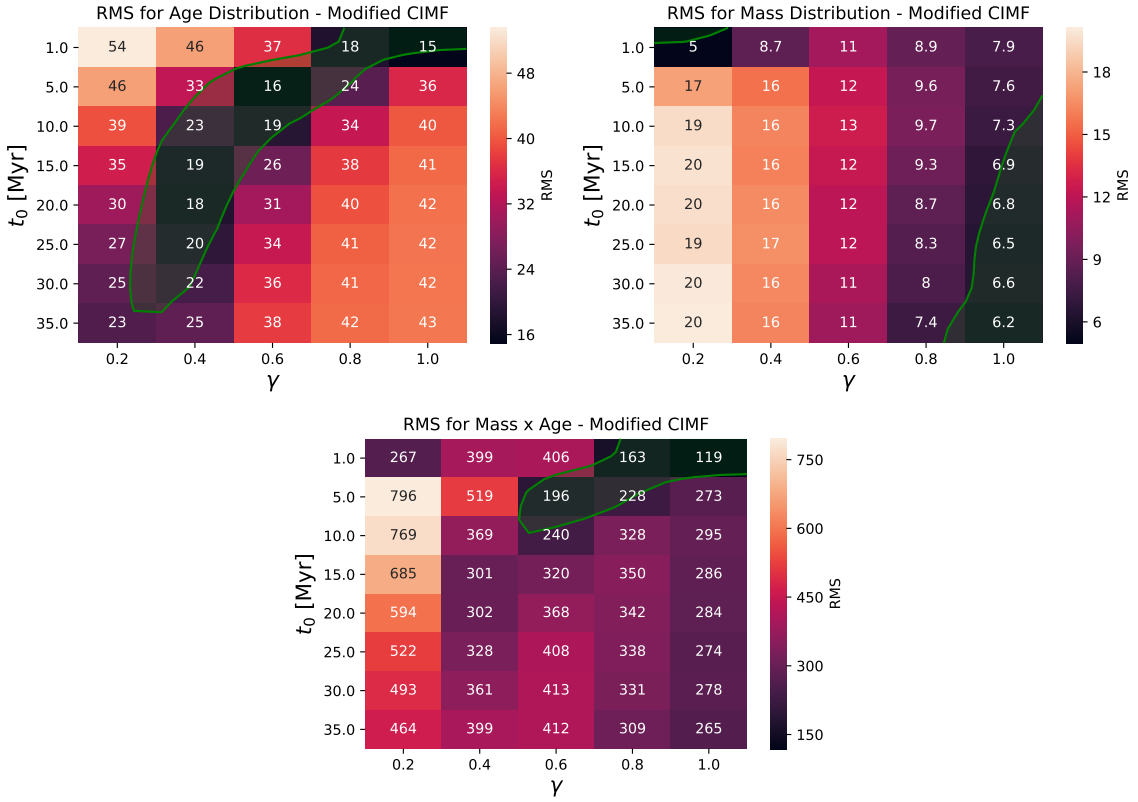

Figure 4.36: Heatmaps of the RMS between the simulated and observed age (left), mass (center) and combination of age and mass (right) distributions for the silver sample within 2 kpc around the Sun and a modified CIMF. The green area represents the contour of the 1.5σ region, considering σ as the minimum value of the RMS.

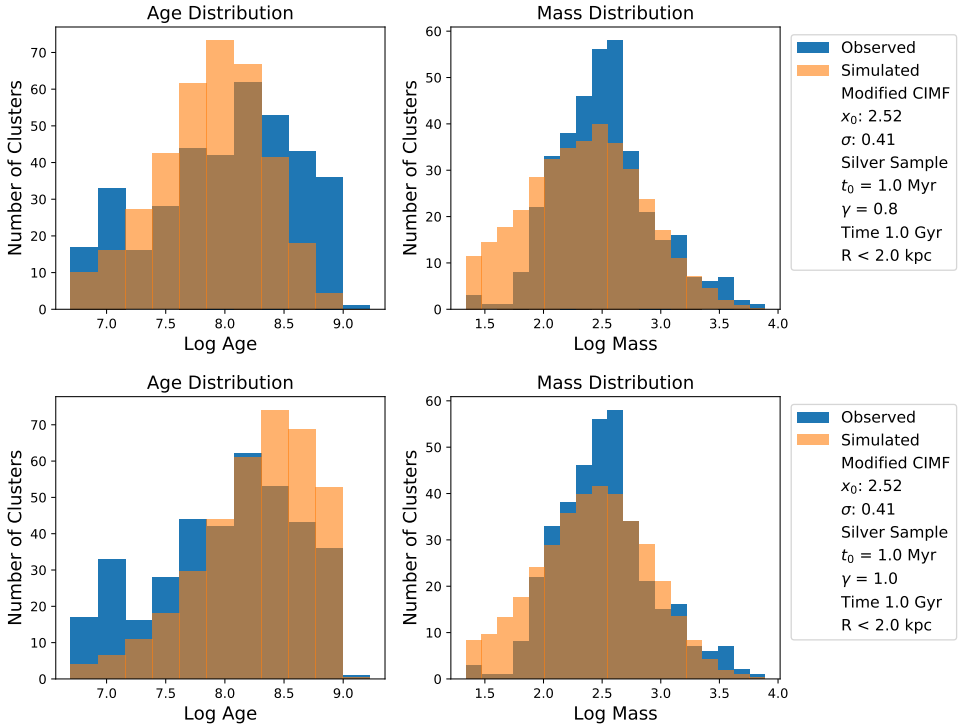

Figure 4.37: Comparison of the distribution of age (left) and mass (right) between the simulations (orange) and the observations (blue) considering t_0 = 1 Myr and $\gamma = 0.8$ (top) and t_0 = 1 Myr and $\gamma = 1$ (bottom), using the silver sample within 2 kpc around the Sun and the modified CIMF.

For the OCs in the silver sample within 1.5 kpc, the results show the exact same trends as well as the results with OCs from the gold sample. In Appendix B, we show the results considering t_0 = 5 Myr and $\gamma = 0.6$, using the silver sample within 1.5 kpc and the gold sample, both for the power law and modified CIMF.

As seen in our results, the modified CIMF that was implemented as a "quick test" leads to a better agreement between the simulated and observed mass distributions. This indicates that the CIMF commonly used may not be the same for embedded and non-embedded clusters. We suggest that it has a "breaking" point where it deviates from the power law form. Further investigation on this is planed for future work.

Chapter 5

Conclusions

In this study, we determined the luminous masses for a sample of 1724 OCs by comparing their luminosity distribution to the theoretical luminosity function, taken from Padova stellar evolution models, building the first catalogue of masses for a large sample in the solar neighbourhood using *Gaia* Data Release 2 (Gaia Collaboration et al., 2018). Our luminous mass distribution peaks at log(M) = 2.7 dex and is compatible with the tidal mass distribution determined by Piskunov et al. (2007).

To determine the masses, it was required that we started by determining the radii of each cluster in our sample. The tidal and core radii were determined by fitting the King density profile to the observed density profile of each cluster using LMFIT python package (Newville et al., 2014). Our values of core radii are consistent with the values found by Tarricq et al. (2021) however, we obtained a different distribution of tidal radii which is more consistent with the one found by Piskunov et al. (2007). Besides that, the results show large uncertainties for the tidal radii that suggest that determining tidal radii is very challenging. However, we verified that despite the large uncertainties in the tidal radii, it does not have a significant impact on our luminous mass determinations. In the future, we plan to explore other density profiles to obtain more robust determinations and assess the effect of choosing a particular profile.

To study the disruption process that leads to the loss of mass in open clusters, we simulated the build up and mass evolution of a population of clusters following the equations derived by Lamers et al. (2005a). Comparing the observed age and, for the first time, the mass distributions to the simulations, we were able to constrain the disruption experienced by open clusters in the solar neighbourhood. Our results recover the parameters obtained in the literature for disruption timescales of a few million years (Lamers et al., 2005a) as well as an additional optimal area for smaller disruption timescales, t_0 (related to the tidal forces of the host galaxy) and larger values of γ. However, when using the parameters (t_0 and γ) presented in the literature determined using the age distribution only, the simulated mass distribution does not agree with the observations. As our model starts from the moment clusters emerge from the GMC, the Cluster Initial Mass Function (CIMF) should reflect the mass distribution of young clusters so we modified the CIMF to match the distribution of mass of OCs with ages under 50 Myrs. In a heuristic approach, we use this toy CIMF as a quick test to check the results. In the future, we plan to explore, in depth, the functional form of the CIMF. The results showed a simulated mass distribution that peaks closer to the peak of the observations, while still preserving the good agreement with the age distribution. These results are also verified for a sample with only clusters with high quality determinations as well as a smaller sample within 1.5 kpc around the Sun. This indicates that the CIMF universally used, which is determined for embedded clusters, may not be the same for non embedded clusters.

In the future, we plan to determine dynamical masses which will be compared to the luminous ones found in this study to assess the common assumption of virial equilibrium in open clusters. The tidal field

dependence on the location will also be explored in future work. In the next years, with missions such as *Gaia* which are still collecting data, cleaner CMDs will be made available for more clusters which will allow more precise determinations of clusters' parameters. That will increase the number of clusters with high quality determinations, increasing the size of our gold sample to provide better constraints on the processes that drive the disruption of OCs.

Bibliography

Alexander, Poul E. R. and Gieles (June 2012). "A prescription and fast code for the long-term evolution of star clusters". In: *Monthly Notices of the Royal Astronomical Society* 422. ADS Bibcode: 2012MNRAS.422.3415A, pp. 3415–3432. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2012.20867.x. URL: https://ui.adsabs.harvard.edu/abs/2012MNRAS.422.3415A.

Angelo, M. S. et al. (Jan. 2021). "Characterizing dynamical states of Galactic open clusters with Gaia DR2". In: *Monthly Notices of the Royal Astronomical Society* 500. ADS Bibcode: 2021MNRAS.500.4338A, pp. 4338–4353. ISSN: 0035-8711. DOI: 10.1093/mnras/staa3192. URL: https://ui.adsabs.harvard.edu/abs/2021MNRAS.500.4338A.

Baumgardt, Holger and Junichiro Makino (Mar. 2003). "Dynamical evolution of star clusters in tidal fields". In: *Monthly Notices of the Royal Astronomical Society* 340. ADS Bibcode: 2003MNRAS.340..227B, pp. 227–246. ISSN: 0035-8711. DOI: 10.1046/j.1365-8711.2003.06286.x. URL: https://ui.adsabs.harvard.edu/abs/2003MNRAS.340..227B.

Binney, J. and S. Tremaine (2011). *Galactic Dynamics: Second Edition*. Princeton Series in Astrophysics. Princeton University Press. ISBN: 9781400828722. URL: https://books.google.pt/books?id=6mF4CKxlbLsC.

Bisht, D. et al. (May 2020). "A comprehensive study of open clusters Czernik 14, Haffner 14, Haffner 17 and King 10 using multicolour photometry and Gaia DR2 astrometry". In: *Monthly Notices of the Royal Astronomical Society* 494. ADS Bibcode: 2020MNRAS.494..607B, pp. 607–623. ISSN: 0035-8711. DOI: 10.1093/mnras/staa656. URL: https://ui.adsabs.harvard.edu/abs/2020MNRAS.494..607B.

Bobylev, V. V. and A. T. Bajkova (Jan. 2014). "The Milky Way spiral structure parameters from data on masers and selected open clusters". In: *Monthly Notices of the Royal Astronomical Society* 437. ADS Bibcode: 2014MNRAS.437.1549B, pp. 1549–1553. ISSN: 0035-8711. DOI: 10.1093/mnras/stt1987. URL: https://ui.adsabs.harvard.edu/abs/2014MNRAS.437.1549B.

Bonatto, C. et al. (Jan. 2006). "Probing disk properties with open clusters". In: 446.1, pp. 121–135. DOI: 10.1051/0004-6361:20053573. arXiv: astro-ph/0509804 [astro-ph].

Bossini, D. et al. (Mar. 2019). "Age determination for 269 $Gaia$ DR2 Open Clusters". In: *Astronomy & Astrophysics* 623. arXiv:1901.04733 [astro-ph], A108. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/201834693. URL: http://arxiv.org/abs/1901.04733.

Boutloukos, S. G. and H. Lamers (Jan. 2003). "Star cluster formation and disruption time-scales - I. An empirical determination of the disruption time of star clusters in four galaxies". In: *Monthly Notices of the Royal Astronomical Society* 338. ADS Bibcode: 2003MNRAS.338..717B, pp. 717–732. ISSN: 0035-8711. DOI: 10.1046/j.1365-8711.2003.06083.x. URL: https://ui.adsabs.harvard.edu/abs/2003MNRAS.338..717B.

Bressan, Alessandro et al. (Nov. 2012). "PARSEC: stellar tracks and isochrones with the PAdova and TRieste Stellar Evolution Code". In: *Monthly Notices of the Royal Astronomical Society* 427. ADS

Bibcode: 2012MNRAS.427..127B, pp. 127–145. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2012.21948.x. URL: https://ui.adsabs.harvard.edu/abs/2012MNRAS.427..127B.

Buckner, Anne S. M. and Dirk Froebrich (Oct. 2014). "Properties of star clusters - II. Scaleheight evolution of clusters". In: 444.1, pp. 290–302. DOI: 10.1093/mnras/stu1440. arXiv: 1407.4611 [astro-ph.SR].

Cantat-Gaudin, T. et al. (Oct. 2018). "A Gaia DR2 view of the open cluster population in the Milky Way". en. In: *Astronomy and Astrophysics* 618, A93. ISSN: 0004-6361. DOI: 10.1051/0004-6361/201833476. URL: https://ui.adsabs.harvard.edu/abs/2018A&A...618A..93C/abstract.

Cantat-Gaudin, T. et al. (Aug. 2020). "Painting a portrait of the Galactic disc with its stellar clusters". en. In: *Astronomy and Astrophysics* 640, A1. ISSN: 0004-6361. DOI: 10.1051/0004-6361/202038192. URL: https://ui.adsabs.harvard.edu/abs/2020A&A...640A...1C/abstract.

Cantat-Gaudin, Tristan (Feb. 2022). "Milky Way Star Clusters and Gaia: A Review of the Ongoing Revolution". In: *Universe* 8.2, p. 111. DOI: 10.3390/universe8020111.

Carlos, Edward Stafford (June 2014). *The Sidereal Messenger of Galileo Galileiand a Part of the Preface to Kepler's Dioptrics Containing the Original Account of Galileo's Astronomical Discoveries*. Trans. by Edward Stafford Carlos. DOI: https://doi.org/10.1007/978-3-642-00792-7.

Castro-Ginard, A. et al. (Mar. 2020). "Hunting for open clusters in Gaia DR2: 582 new open clusters in the Galactic disc". en. In: *Astronomy and Astrophysics* 635, A45. ISSN: 0004-6361. DOI: 10.1051/0004-6361/201937386. URL: https://ui.adsabs.harvard.edu/abs/2020A&A...635A..45C/abstract.

Chen, Yang et al. (Sept. 2015). "PARSEC evolutionary tracks of massive stars up to 350 $M_{\odot}$ at metallicities $0.0001 \leq Z \leq 0.04$". In: 452.1, pp. 1068–1080. DOI: 10.1093/mnras/stv1281. arXiv: 1506.01681 [astro-ph.SR].

D'Onofrio, Mauro and C. Burigana (Jan. 2009). *Questions of Modern Cosmology: Galileo's Legacy*. DOI: 10.1007/978-3-642-00792-7.

Dias, W. S. et al. (July 2002). "New catalogue of optically visible open clusters and candidates". en. In: *Astronomy and Astrophysics* 389, pp. 871–873. ISSN: 0004-6361. DOI: 10.1051/0004-6361:20020668. URL: https://ui.adsabs.harvard.edu/abs/2002A&A...389..871D/abstract.

Dias, W. S. et al. (June 2021). "Updated parameters of 1743 open clusters based on Gaia DR2". In: *Monthly Notices of the Royal Astronomical Society* 504. ADS Bibcode: 2021MNRAS.504..356D, pp. 356–371. ISSN: 0035-8711. DOI: 10.1093/mnras/stab770. URL: https://ui.adsabs.harvard.edu/abs/2021MNRAS.504..356D.

Dias, Wilton S. and J. R. D. Lépine (Aug. 2005). "Direct Determination of the Spiral Pattern Rotation Speed of the Galaxy". In: 629.2, pp. 825–831. DOI: 10.1086/431456. arXiv: astro-ph/0503083 [astro-ph].

Ferreira, F. A. et al. (Mar. 2021). "New star clusters discovered towards the Galactic bulge direction using Gaia DR2". In: *Monthly Notices of the Royal Astronomical Society* 502. ADS Bibcode: 2021MNRAS.502L..90F, pp. L90–L94. ISSN: 0035-8711. DOI: 10.1093/mnrasl/slab011. URL: https://ui.adsabs.harvard.edu/abs/2021MNRAS.502L..90F.

Foreman-Mackey, Daniel et al. (Mar. 2013). "emcee: The MCMC Hammer". In: *Publications of the Astronomical Society of the Pacific* 125. ADS Bibcode: 2013PASP..125..306F, p. 306. ISSN: 0004-6280. DOI: 10.1086/670067. URL: https://ui.adsabs.harvard.edu/abs/2013PASP..125..306F.

Gaia Collaboration et al. (Nov. 2016). “Gaia Data Release 1. Summary of the astrometric, photometric, and survey properties”. In: 595, A2, A2. DOI: 10.1051/0004-6361/201629512. arXiv: 1609.04172 [astro-ph.IM].

Gaia Collaboration et al. (Aug. 2018). “Gaia Data Release 2. Summary of the contents and survey properties”. In: 616, A1, A1. DOI: 10.1051/0004-6361/201833051. arXiv: 1804.09365 [astro-ph.GA].

Gaia Collaboration et al. (May 2021). “Gaia Early Data Release 3. Summary of the contents and survey properties”. In: 649, A1, A1. DOI: 10.1051/0004-6361/202039657. arXiv: 2012.01533 [astro-ph.GA].

Gaia Collaboration et al. (July 2022). “Gaia Data Release 3: Summary of the content and survey properties”. In: *arXiv e-prints*, arXiv:2208.00211, arXiv:2208.00211. arXiv: 2208.00211 [astro-ph.GA].

Gao, Xinhua, ShouKun Xu, and Lei Xue (June 2021). “The spatial structure and dynamical state of the open cluster NGC 2112”. In: *Publications of the Astronomical Society of Japan* 73. ADS Bibcode: 2021PASJ...73..652G, pp. 652–659. ISSN: 0004-6264. DOI: 10.1093/pasj/psab027. URL: https://ui.adsabs.harvard.edu/abs/2021PASJ...73..652G.

Gieles, M. (Apr. 2009). “The early evolution of the star cluster mass function”. In: *Monthly Notices of the Royal Astronomical Society* 394.4. arXiv:0901.0830 [astro-ph], pp. 2113–2126. ISSN: 00358711, 13652966. DOI: 10.1111/j.1365-2966.2009.14473.x. URL: http://arxiv.org/abs/0901.0830.

Gieles et al. (Jan. 2014). “A prescription and fast code for the long-term evolution of star clusters - II. Unbalanced and core evolution”. In: *Monthly Notices of the Royal Astronomical Society* 437. ADS Bibcode: 2014MNRAS.437..916G, pp. 916–929. ISSN: 0035-8711. DOI: 10.1093/mnras/stt1980. URL: https://ui.adsabs.harvard.edu/abs/2014MNRAS.437..916G.

Goodman, Jonathan and Jonathan Weare (Jan. 2010). “Ensemble samplers with affine invariance”. In: *Communications in Applied Mathematics and Computational Science* 5.1, pp. 65–80. DOI: 10.2140/camcos.2010.5.65.

Harris, William E. (Oct. 1996). “A Catalog of Parameters for Globular Clusters in the Milky Way”. In: 112, p. 1487. DOI: 10.1086/118116.

Joshi, Y. C. (Oct. 2005). “Interstellar extinction towards open clusters and Galactic structure”. In: 362.4, pp. 1259–1266. DOI: 10.1111/j.1365-2966.2005.09391.x. arXiv: astro-ph/0507069 [astro-ph].

Joshi, Y. C. et al. (Oct. 2016). “Study of open clusters within 1.8 kpc and understanding the Galactic structure”. en. In: *Astronomy & Astrophysics, Volume 593, id.A116, <NUMPAGES>13</NUMPAGES> pp.* 593, A116. ISSN: 0004-6361. DOI: 10.1051/0004-6361/201628944. URL: https://ui.adsabs.harvard.edu/abs/2016A%26A...593A.116J/abstract.

Kharchenko, N. V. (Oct. 2001). “All-sky compiled catalogue of 2.5 million stars”. In: *Kinematika i Fizika Nebesnykh Tel* 17. ADS Bibcode: 2001KFNT...17..409K, pp. 409–423. ISSN: 0233-7665. URL: https://ui.adsabs.harvard.edu/abs/2001KFNT...17..409K.

Kharchenko, N. V. et al. (Aug. 2005). “Astrophysical parameters of Galactic open clusters”. In: 438.3, pp. 1163–1173. DOI: 10.1051/0004-6361:20042523. arXiv: astro-ph/0501674 [astro-ph].

Kharchenko, N. V. et al. (Oct. 2013). “Global survey of star clusters in the Milky Way. II. The catalogue of basic parameters”. en. In: *Astronomy and Astrophysics* 558, A53. ISSN: 0004-6361. DOI: 10.

1051/0004-6361/201322302. URL: https://ui.adsabs.harvard.edu/abs/2013A&A...558A..53K/abstract.

King, Ivan (Oct. 1962). "The structure of star clusters. I. an empirical density law". In: *The Astronomical Journal* 67. ADS Bibcode: 1962AJ.....67..471K, p. 471. ISSN: 0004-6256. DOI: 10.1086/108756. URL: https://ui.adsabs.harvard.edu/abs/1962AJ.....67..471K.

Knuth, Kevin H. (May 2006). "Optimal Data-Based Binning for Histograms". In: *arXiv e-prints*. ADS Bibcode: 2006physics...5197K, physics/0605197. URL: https://ui.adsabs.harvard.edu/abs/2006physics...5197K.

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

Kruijssen, J. M. Diederik et al. (June 2011). "Modelling the formation and evolution of star cluster populations in galaxy simulations". In: *Monthly Notices of the Royal Astronomical Society* 414. ADS Bibcode: 2011MNRAS.414.1339K, pp. 1339–1364. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2011.18467.x. URL: https://ui.adsabs.harvard.edu/abs/2011MNRAS.414.1339K.

Küpper, Andreas H. W. et al. (Oct. 2010). "Peculiarities in velocity dispersion and surface density profiles of star clusters". In: *Monthly Notices of the Royal Astronomical Society* 407. ADS Bibcode: 2010MNRAS.407.2241K, pp. 2241–2260. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2010.17084.x. URL: https://ui.adsabs.harvard.edu/abs/2010MNRAS.407.2241K.

Lada, Charles J. and Elizabeth A. Lada (Jan. 1991). "The nature, origin and evolution of embedded star clusters." In: *The Formation and Evolution of Star Clusters*. Ed. by Kenneth Janes. Vol. 13. Astronomical Society of the Pacific Conference Series, pp. 3–22. URL: https://ui.adsabs.harvard.edu/abs/1991ASPC...13....3L.

— (Sept. 2003). "Embedded Clusters in Molecular Clouds". In: *Annual Review of Astronomy and Astrophysics* 41.1. arXiv:astro-ph/0301540, pp. 57–115. ISSN: 0066-4146, 1545-4282. DOI: 10.1146/annurev.astro.41.011802.094844. URL: http://arxiv.org/abs/astro-ph/0301540.

Lamers, Holger Baumgardt, and Gieles (July 2010). "Mass loss rates and the mass evolution of star clusters". en. In: DOI: 10.1111/j.1365-2966.2010.17309.x. URL: https://arxiv.org/abs/1007.1078v1.

Lamers, H., N. Bastian, and M. Gieles (Dec. 2004). "The disruption time scales of clusters in different galaxies". In: 322. Conference Name: The Formation and Evolution of Massive Young Star Clusters ADS Bibcode: 2004ASPC..322..409L, p. 409. URL: https://ui.adsabs.harvard.edu/abs/2004ASPC..322..409L.

Lamers, H. and M. Gieles (Aug. 2006). "Clusters in the solar neighbourhood: how are they destroyed?" en. In: *Astronomy and Astrophysics* 455.2, pp. L17–L20. ISSN: 0004-6361. DOI: 10.1051/0004-6361:20065567. URL: https://ui.adsabs.harvard.edu/abs/2006A&A...455L..17L/abstract.

Lamers, H., M. Gieles, and S. F. Portegies Zwart (Jan. 2005a). "Disruption time scales of star clusters in different galaxies". en. In: *Astronomy and Astrophysics* 429, pp. 173–179. ISSN: 0004-6361. DOI: 10.1051/0004-6361:20041476. URL: https://ui.adsabs.harvard.edu/abs/2005A&A...429..173L/abstract.

Lamers, H. et al. (Oct. 2005b). "An analytical description of the disruption of star clusters in tidal fields with an application to Galactic open clusters". en. In: *Astronomy and Astrophysics* 441.1, pp. 117–129. ISSN: 0004-6361. DOI: 10.1051/0004-6361:20042241. URL: https://ui.adsabs.harvard.edu/abs/2005A&A...441..117L/abstract.

Liu, Lei and Xiaoying Pang (Dec. 2019). “A Catalog of Newly Identified Star Clusters in Gaia DR2”. In: *The Astrophysical Journal Supplement Series* 245. ADS Bibcode: 2019ApJS..245...32L, p. 32. ISSN: 0067-0049. DOI: 10.3847/1538-4365/ab530a. URL: https://ui.adsabs.harvard.edu/abs/2019ApJS..245...32L.

Maíz Apellániz, J. and M. Weiler (Nov. 2018). “Reanalysis of the Gaia Data Release 2 photometric sensitivity curves using HST/STIS spectrophotometry”. In: 619, A180, A180. DOI: 10.1051/0004-6361/201834051. arXiv: 1808.02820 [astro-ph.IM].

Marigo, P. et al. (May 2008). “Evolution of asymptotic giant branch stars. II. Optical to far-infrared isochrones with improved TP-AGB models”. In: 482.3, pp. 883–905. DOI: 10.1051/0004-6361:20078467. arXiv: 0711.4922 [astro-ph].

Marigo, P. et al. (Sept. 2013). “Evolution of thermally pulsing asymptotic giant branch stars - I. The COLIBRI code”. In: 434.1, pp. 488–526. DOI: 10.1093/mnras/stt1034. arXiv: 1305.4485 [astro-ph.SR].

Minniti, D. et al. (July 2010). “VISTA Variables in the Via Lactea (VVV): The public ESO near-IR variability survey of the Milky Way”. In: *New Astronomy* 15.5, pp. 433–443. DOI: 10.1016/j.newast.2009.12.002. URL: https://doi.org/10.10162Fj.newast.2009.12.002.

Moitinho, A. et al. (May 2006). “Spiral structure of the third galactic quadrant and the solution to the Canis Major debate”. In: *Monthly Notices of the Royal Astronomical Society* 368. ADS Bibcode: 2006MNRAS.368L..77M, pp. L77–L81. ISSN: 0035-8711. DOI: 10.1111/j.1745-3933.2006.00163.x. URL: https://ui.adsabs.harvard.edu/abs/2006MNRAS.368L..77M.

Moitinho, André (Jan. 2010). “Observational properties of the open cluster system of the Milky Way and what they tell us about our Galaxy”. In: 266. Conference Name: Star Clusters: Basic Galactic Building Blocks Throughout Time and Space Place: eprint: arXiv:0911.1459 ADS Bibcode: 2010IAUS..266..106M, pp. 106–116. DOI: 10.1017/S1743921309990949. URL: https://ui.adsabs.harvard.edu/abs/2010IAUS..266..106M.

Monteiro, H. et al. (Feb. 2017). “The OPD photometric survey of open clusters II. robust determination of the fundamental parameters of 24 open clusters”. In: *New Astronomy* 51. ADS Bibcode: 2017NewA...51...15M, pp. 15–27. ISSN: 1384-1076. DOI: 10.1016/j.newast.2016.08.001. URL: https://ui.adsabs.harvard.edu/abs/2017NewA...51...15M.

Monteiro, H. et al. (Dec. 2020). “Fundamental parameters for 45 open clusters with Gaia DR2, an improved extinction correction and a metallicity gradient prior”. In: *Monthly Notices of the Royal Astronomical Society* 499. ADS Bibcode: 2020MNRAS.499.1874M, pp. 1874–1889. ISSN: 0035-8711. DOI: 10.1093/mnras/staa2983. URL: https://ui.adsabs.harvard.edu/abs/2020MNRAS.499.1874M.

Monteiro, H. et al. (2021). “The Distribution of Open Clusters in the Galaxy”. In: *Frontiers in Astronomy and Space Sciences* 8. ISSN: 2296-987X. URL: https://www.frontiersin.org/articles/10.3389/fspas.2021.656474.

Newville, Matthew et al. (Sept. 2014). *LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python*. Version 0.8.0. DOI: 10.5281/zenodo.11813. URL: https://doi.org/10.5281/zenodo.11813.

Olivares, J. et al. (Apr. 2018). “The seven sisters DANCe. III. Projected spatial distribution”. en. In: *Astronomy & Astrophysics, Volume 612, id.A70, <NUMPAGES>39</NUMPAGES> pp.* 612, A70. ISSN: 0004-6361. DOI: 10.1051/0004-6361/201731996. URL: https://ui.adsabs.harvard.edu/abs/2018A%26A...612A..70O/abstract.

Piskunov, A. E. et al. (Jan. 2006). “Revisiting the population of Galactic open clusters”. In: 445.2, pp. 545–565. DOI: 10.1051/0004-6361:20053764. arXiv: astro-ph/0508575 [astro-ph].

Piskunov, A. E. et al. (June 2007). “Towards absolute scales for the radii and masses of open clusters”. en. In: *Astronomy & Astrophysics* 468.1. Number: 1 Publisher: EDP Sciences, pp. 151–161. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361:20077073. URL: https://www.aanda.org/articles/aa/abs/2007/22/aa7073-07/aa7073-07.html.

— (Jan. 2008). “Tidal radii and masses of open clusters”. en. In: *Astronomy and Astrophysics* 477.1, pp. 165–172. ISSN: 0004-6361. DOI: 10.1051/0004-6361:20078525. URL: https://ui.adsabs.harvard.edu/abs/2008A&A...477..165P/abstract.

Silva, André Filipe Vitorino da (2016). *Tracing the scale height evolution of the galactic disk with open clusters*. DOI: http://hdl.handle.net/10451/35673.

Sim, Gyuheon et al. (Oct. 2019). “207 New Open Star Clusters within 1 kpc from Gaia Data Release 2”. In: *Journal of Korean Astronomical Society* 52. ADS Bibcode: 2019JKAS...52..145S, pp. 145–158. ISSN: 1225-4614. DOI: 10.5303/JKAS.2019.52.5.145. URL: https://ui.adsabs.harvard.edu/abs/2019JKAS...52..145S.

Skrutskie, M. F. et al. (Feb. 2006). “The Two Micron All Sky Survey (2MASS)”. In: 131.2, pp. 1163–1183. DOI: 10.1086/498708.

Tarricq, Y. et al. (Mar. 2021). “Structural parameters of 389 local open clusters”. en. In: *Astronomy and Astrophysics* 659, A59. ISSN: 0004-6361. DOI: 10.1051/0004-6361/202142186. URL: https://ui.adsabs.harvard.edu/abs/2022A&A...659A..59T/abstract.

Vázquez, Ruben A. et al. (Jan. 2008). “Spiral Structure in the Outer Galactic Disk. I. The Third Galactic Quadrant”. In: *The Astrophysical Journal* 672. ADS Bibcode: 2008ApJ...672..930V, pp. 930–939. ISSN: 0004-637X. DOI: 10.1086/524003. URL: https://ui.adsabs.harvard.edu/abs/2008ApJ...672..930V.

Ven, G. van de et al. (Jan. 2006). “The dynamical distance and intrinsic structure of the globular cluster ω Centauri”. en. In: *Astronomy and Astrophysics* 445.2, pp. 513–543. ISSN: 0004-6361. DOI: 10.1051/0004-6361:20053061. URL: https://ui.adsabs.harvard.edu/abs/2006A&A...445..513V/abstract.

Wielen, R. (July 1971). “The Age Distribution and Total Lifetimes of Galactic Clusters”. In: 13, pp. 309–322.

Yeh, Fu Chi et al. (Mar. 2019). “Ruprecht 147: A Paradigm of Dissolving Star Cluster”. In: *The Astronomical Journal* 157. ADS Bibcode: 2019AJ....157..115Y, p. 115. ISSN: 0004-6256. DOI: 10.3847/1538-3881/aaff6c. URL: https://ui.adsabs.harvard.edu/abs/2019AJ....157..115Y.

Appendices

Appendix A

Distribution of parameters N_0 and c

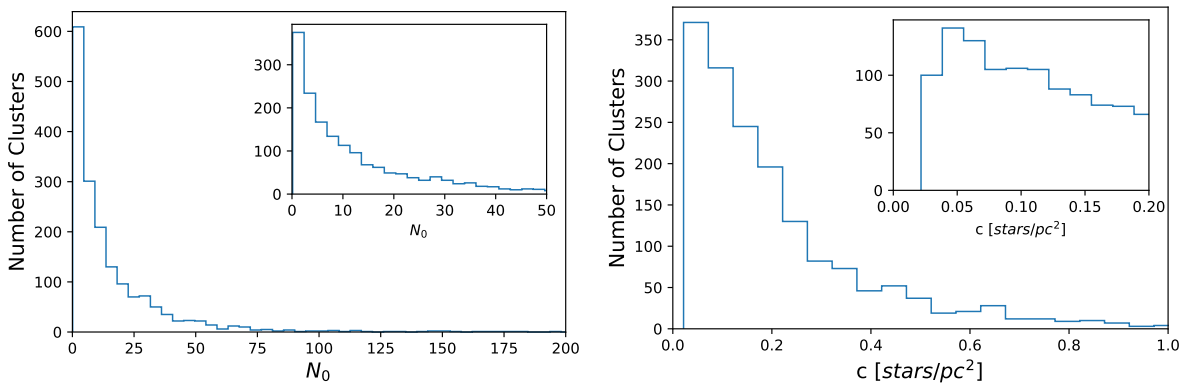

Figure A.1: Distribution of N_0 (left) and parameter c (right) for 1724 OCs.

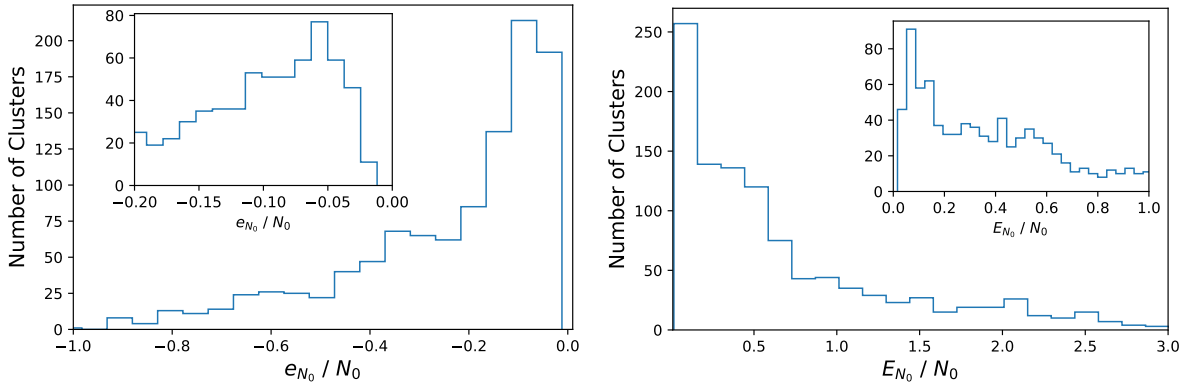

Figure A.2: Distribution of the lower and upper uncertainty of N_0, considering 1724 OCs.

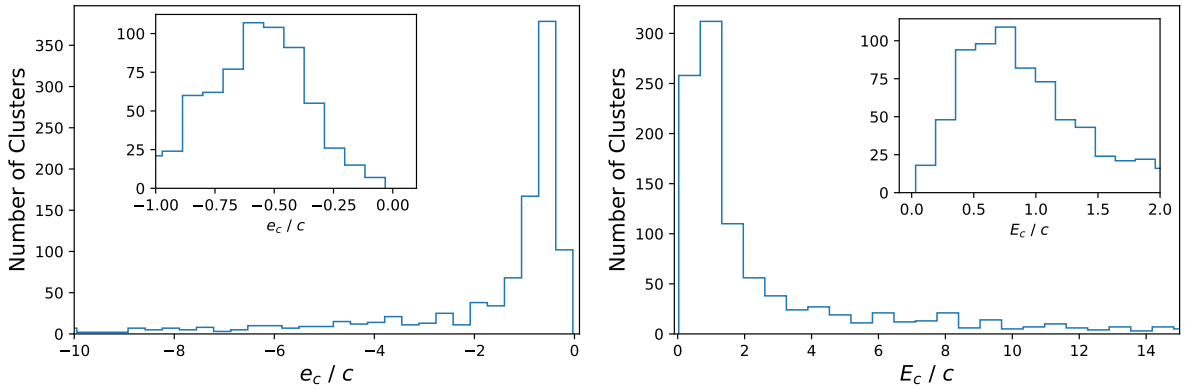

Figure A.3: Distribution of the lower and upper uncertainty of parameter c, considering 1724 OCs.

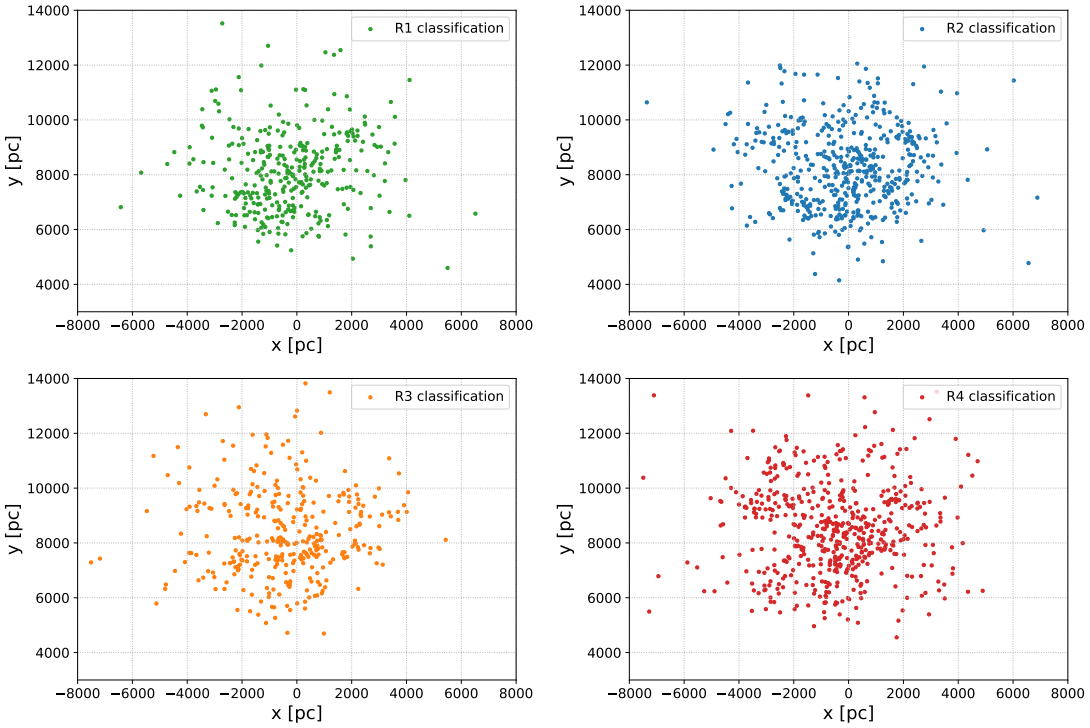

Figure A.4: Spatial distribution in the X-Y plane for the clusters with R1 (best quality), R2, R3 (worst quality) and R4 (unreliable) classification.

Appendix B

Simulation results for the gold sample and a smaller (1.5 kpc) sample

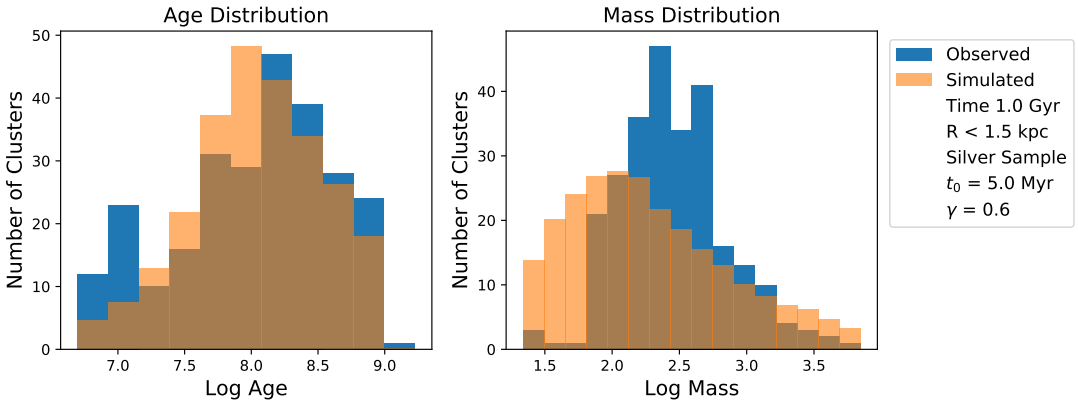

Figure B.1: Comparison of the distribution of age (left) and mass (right) between the simulations (orange) and the observations (blue) considering $t_0 = 5$ Myr and $\gamma = 0.6$, using the silver sample within 1.5 kpc around the Sun and a power law CIMF.

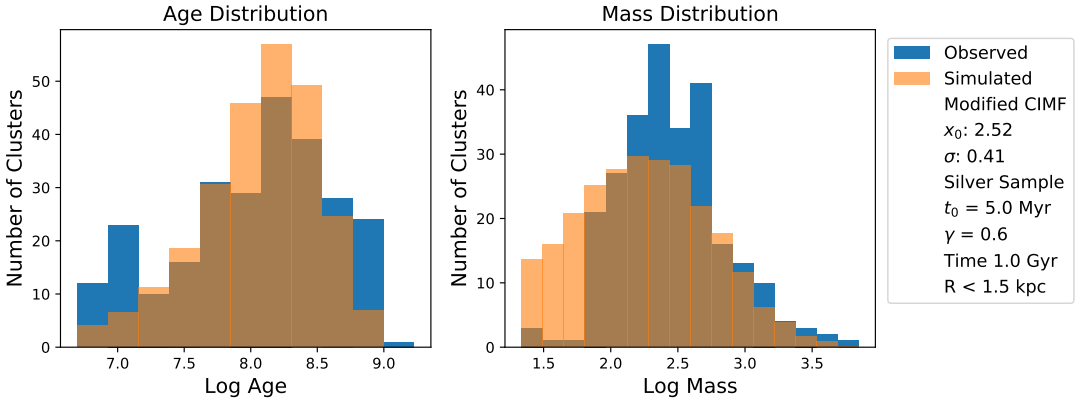

Figure B.2: Comparison of the distribution of age (left) and mass (right) between the simulations (orange) and the observations (blue) considering $t_0 = 5$ Myr and $\gamma = 0.6$, using the silver sample within 1.5 kpc around the Sun and the implemented modified CIMF.

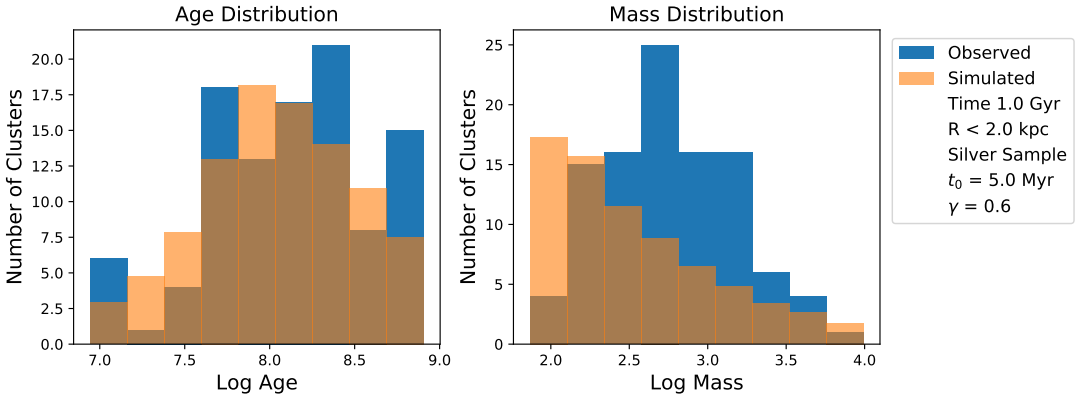

Figure B.3: Comparison of the distribution of age (left) and mass (right) between the simulations (orange) and the observations (blue) considering $t_0 = 5$ Myr and $\gamma = 0.6$, using the gold sample within 2 kpc around the Sun and a power law CIMF.

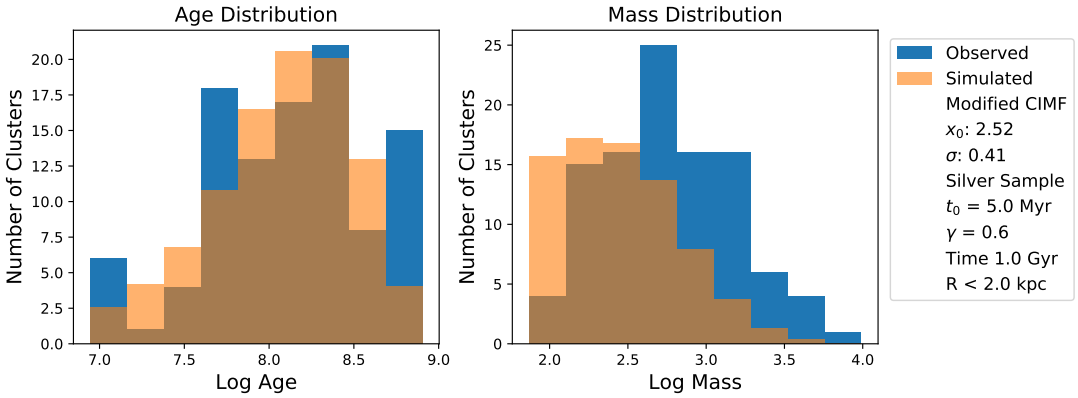

Figure B.4: Comparison of the distribution of age (left) and mass (right) between the simulations (orange) and the observations (blue) considering $t_0 = 5$ Myr and $\gamma = 0.6$, using the gold sample within 2 kpc around the Sun and the implemented modified CIMF.