Universidad de Concepción

Facultad de Ciencias Físicas y Matemáticas

Programa de Magíster en Astronomía

CHARACTERIZING NGC 6383: A STUDY OF FUNDAMENTAL PROPERTIES, PRE-MAIN SEQUENCE STARS, MASS SEGREGATION, AND AGE USING GAIA DR3 AND 2MASS

Tesis presentada a la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Concepción para optar al grado de Magíster en Astronomía

POR

LUCAS MATIAS PULGAR-ESCOBAR

Profesor Guía: Dr. Ronald Enrique Mennickent Cid

marzo 2026

Concepción, Chile

To all the people who were part of this journey; to my mom, who supported me in many ways; to my friends, who listened, helped, and celebrated every step; to Nicolás, who walked beside me at the beginning of this road, and whose path has since diverged into its own direction; and to Alexis, who arrived in the final stretch, and with whom my path is now, finally, converging.

Acknowledgements

We gratefully acknowledge support by the ANID BASAL project FB210003 and the SOCHIAS GEMINI project 32230014.

L.P. gratefully acknowledges the invaluable comments and guidance of Dr. Robert Mathieu in the preparation of this investigation.

L.P. further acknowledges Dr. Pierluigi Cerulo for his guidance in the integration of machine learning methods into astronomical research, and for introducing the tools and conceptual framework that shaped the data-driven components of this work.

This work has made use of data from the European Space Agency (ESA) mission *Gaia* (`https://www.cosmos.esa.int/gaia`), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, `https://www.cosmos.esa.int/web/gaia/dpac/consortium`). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement.

This publication makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation.

This work made use of Astropy,[1] a community-developed core Python package and an ecosystem of tools and resources for astronomy (The Astropy Collaboration

[1] `http://www.astropy.org`

et al., 2013; The Astropy Collaboration et al., 2018; The Astropy Collaboration et al., 2022).

Table of contents

List of tables

List of figures

Resumen

Presentamos una caracterización bayesiana integral del cúmulo abierto galáctico joven NGC 6383, combinando astrometría y fotometría de *Gaia* DR3 con datos infrarrojos cercanos de 2MASS a través de COSMIC (*Characterisation Of Star clusters using Machine learning Inference and Clustering*), un pipeline de código abierto que integra agrupamiento basado en densidad (HDBSCAN) con modelamiento jerárquico bayesiano mediante PyMC y el muestreador No-U-Turn (NUTS).

A partir de una muestra de trabajo de 15 276 fuentes de *Gaia* DR3 dentro de 40 arcmin del centro del cúmulo, el pipeline identifica 254 miembros probables con un umbral de pseudoprobabilidad $\tilde{p}_i > 0.6$, con un análisis de sensibilidad al umbral que demuestra que la población recuperada varía entre 161 y 321 miembros en cuatro umbrales de probabilidad. Un modelo jerárquico basado en paralajes entrega una distancia $d = 1{,}110 \pm 0{,}060$ kpc, confirmada independientemente por el módulo de distancia isocronal $\mu_{\rm DM} = 10{,}3 \pm 0{,}26$ mag. Los movimientos propios medios son $\overline{\mu_{\alpha*}} = 2{,}540 \pm 0{,}010$ mas año^{-1} y $\overline{\mu_\delta} = -1{,}710 \pm 0{,}009$ mas año^{-1}, con dispersiones intrínsecas de $\sim 0{,}15$ mas año^{-1}.

El ajuste isocronal bayesiano, realizado con ASteCA utilizando isócronas MIST y muestreo NUTS, produce una edad modal $\tau = 3{,}53^{+1{,}40}_{-1{,}00}$ Ma, metalicidad $Z = 0{,}024 \pm 0{,}008$ y extinción visual $A_V = 1{,}24 \pm 0{,}26$ mag. Las clasificaciones de estrellas pre-secuencia principal de Sagitta y el parámetro Q libre de enrojecimiento confirman independientemente un cúmulo muy joven con un episodio extendido de formación estelar que abarca $\sim$1–6 Ma y una fracción elevada de objetos estelares jóvenes $Y_{\rm frac} = 0{,}28$ (IC 95% [0,22; 0,34]). El análisis estructural revela un núcleo compacto ($R_c = 1{,}95 \pm 0{,}19$ arcmin) inmerso en una envolvente extendida

($R_t = 40{,}4 \pm 14{,}3$ arcmin), con concentración $C = 1{,}316$ y 39 potenciales estrellas en escape más allá del radio de Hill que comparten la firma cinemática del cúmulo.

El tiempo de relajación de media masa $t_{\rm rh} = 18{,}4 \pm 2{,}1$ Ma supera la edad del cúmulo por un factor ~ 5 ($\tau/t_{\rm rh} \approx 0{,}19$), situando a NGC 6383 firmemente en el régimen dinámicamente joven. La segregación de masa entre estrellas binarias y de alta masa persiste cuando la muestra se restringe a estrellas cuyas escalas temporales de fricción dinámica exceden la edad del cúmulo ($t_{\rm seg} > \tau$; $m < 8{,}29\ M_\odot$), constituyendo evidencia de segregación de masa primordial o cuasi-primordial consistente con predicciones de simulaciones de N-cuerpos de fusión jerárquica de sub-grupos.

La binaria O7V+O7V HD 159176, históricamente central en los debates sobre la edad de NGC 6383, se demuestra como no-miembro: su movimiento propio se desvía $6{,}6\sigma$ y su paralaje $\sim 60\sigma$ respecto a las distribuciones posteriores del cúmulo. Este resultado resuelve una ambigüedad de décadas en la que la membresía asumida de este sistema inflaba las estimaciones publicadas de edad a 6–10 Ma.

Se identifican cinco contribuciones metodológicas: (i) una arquitectura híbrida de membresía HDBSCAN + pseudoprobabilidad que combina descubrimiento no paramétrico con filtrado iterativo de robustez; (ii) el reemplazo del muestreador por defecto de ASteCA por NUTS, proporcionando distribuciones posteriores exactas y diagnósticos de convergencia geométricos; (iii) una separación formal y doctrinal entre la razón de persistencia de HDBSCAN y las probabilidades de membresía posteriores calibradas; (iv) un análisis de sensibilidad al umbral que cuantifica cómo los resúmenes físicos dependen del corte de membresía; y (v) una disciplina de estatus de afirmaciones que gobierna toda inferencia física, requiriendo condicionamiento explícito sobre el modelo generativo, la especificación de priors y el estado de convergencia. El código fuente de COSMIC, el catálogo de membresía y las muestras posteriores completas se publican abiertamente.

Abstract

We present a comprehensive Bayesian characterisation of the young Galactic open cluster NGC 6383, combining *Gaia* DR3 astrometry and photometry with 2MASS near-infrared data through COSMIC (*Characterisation Of Star clusters using Machine learning Inference and Clustering*), an open-source pipeline that integrates HDBSCAN density-based clustering with hierarchical Bayesian modelling via PyMC and the No-U-Turn Sampler (NUTS).

From a working sample of 15 276 *Gaia* DR3 sources within 40 arcmin of the cluster centre, the pipeline identifies 254 probable members at a pseudoprobability threshold $\tilde{p}_i > 0.6$, with a threshold-sensitivity analysis demonstrating that the recovered population spans 161–321 members across four probability thresholds. A parallax-based hierarchical model yields a distance $d = 1.110 \pm 0.060$ kpc, independently confirmed by the isochronal distance modulus $\mu_{\rm DM} = 10.3 \pm 0.26$ mag. Mean proper motions are $\overline{\mu_{\alpha*}} = 2.540 \pm 0.010$ mas yr^{-1} and $\overline{\mu_\delta} = -1.710 \pm 0.009$ mas yr^{-1}, with intrinsic dispersions of ~ 0.15 mas yr^{-1}.

The Bayesian isochronal fit, performed with ASteCA using MIST isochrones and NUTS sampling, yields a modal age $\tau = 3.53^{+1.40}_{-1.00}$ Myr, metallicity $Z = 0.024 \pm 0.008$, and visual extinction $A_V = 1.24 \pm 0.26$ mag. Sagitta pre-main-sequence classifications and the reddening-free Q-parameter independently confirm a very young cluster with an extended star-formation episode spanning $\sim$1–6 Myr and an elevated young stellar object fraction $Y_{\rm frac} = 0.28$ (95% CI [0.22, 0.34]). The structural analysis reveals a compact core ($R_c = 1.95 \pm 0.19$ arcmin) embedded within an extended envelope ($R_t = 40.4 \pm 14.3$ arcmin), with concentration $C = 1.316$ and 39 potential escapers beyond the Hill radius sharing the cluster's kinematic signature.

The half-mass relaxation time $t_{\rm rh} = 18.4 \pm 2.1$ Myr exceeds the cluster age by a factor of ~ 5 ($\tau/t_{\rm rh} \approx 0.19$), placing NGC 6383 firmly in the dynamically young regime. Mass segregation among binary and high-mass stars persists when the sample is restricted to stars whose dynamical friction timescales exceed the cluster age ($t_{\rm seg} > \tau$; $m < 8.29\ M_{\odot}$), constituting evidence for primordial or near-primordial mass segregation consistent with predictions from N-body simulations of hierarchical sub-clump mergers.

The O7V+O7V binary HD 159176, historically central to age debates about NGC 6383, is demonstrated to be a non-member: its proper motion deviates by 6.6σ and its parallax by $\sim 60\sigma$ from the cluster posteriors. This result resolves a decades-long ambiguity in which assumed membership inflated published age estimates to 6–10 Myr.

Five methodological contributions are identified: (i) a hybrid HDBSCAN + pseudo-probability membership architecture that combines non-parametric discovery with iterative robustness filtering; (ii) the replacement of ASteCA's default sampler with NUTS, providing exact posteriors and geometric convergence diagnostics; (iii) a formal doctrinal separation between the HDBSCAN persistence ratio and calibrated posterior membership probabilities; (iv) a threshold-sensitivity analysis quantifying how physical summaries depend on the membership cut; and (v) a claim-status discipline that governs every physical inference, requiring explicit conditioning on the generative model, prior specification, and convergence status. The COSMIC source code, membership catalogue, and full posterior samples are publicly released. thods: data analysis; techniques: photometric; astrometry

Nomenclature and acronyms

Acronyms and Abbreviations

2MASS Two Micron All Sky Survey.

ADQL Astronomical Data Query Language.

AL/AC Along-scan / Across-scan (Gaia scanning directions).

ASteCA Automated Stellar Cluster Analysis.

BCD Barbier–Chalonge–Divan spectrophotometric system.

BFMI Bayesian Fraction of Missing Information.

CDS Centre de Données astronomiques de Strasbourg.

CI Credible interval.

CMD Colour–magnitude diagram.

COSMIC Characterisation Of Star clusters using Machine learning Inference and Clustering.

CTTS Classical T Tauri star.

DPAC Data Processing and Analysis Consortium (Gaia).

DSS2 Digitized Sky Survey 2.

EDR3 Gaia Early Data Release 3.

EDSD Exponentially Decreasing Space Density (distance prior).

EFF Elson–Fall–Freeman (density profile).

ESA European Space Agency.

ESO European Southern Observatory.

ESS Effective sample size.

FIM Fisher Information Matrix.

Gaia European Space Agency astrometric mission.

Gaia DR3 Gaia Data Release 3.

Gaia RVS Gaia Radial Velocity Spectrometer.

Gaia XP Gaia low-resolution BP/RP spectra.

GMM Gaussian Mixture Model.

HDBSCAN Hierarchical Density-Based Spatial Clustering of Applications with Noise.

HMC Hamiltonian Monte Carlo.

HPD Highest Posterior Density (interval).

IMF Initial mass function.

ISM Interstellar medium.

K–S Kolmogorov–Smirnov (test).

KDE Kernel Density Estimation.

LDB Lithium Depletion Boundary.

LF Luminosity function.

MAP Maximum a posteriori.

MCMC Markov Chain Monte Carlo.

MESA Modules for Experiments in Stellar Astrophysics.

MIST MESA Isochrones and Stellar Tracks.

MOONS Multi-Object Optical and Near-infrared Spectrograph.

MS Main sequence.

NUTS No-U-Turn Sampler.

PARSEC PAdova and TRieste Stellar Evolution Code.

PMS Pre-main sequence.

RUWE Renormalised Unit Weight Error.

SED Spectral energy distribution.

SFE Star formation efficiency.

SIMBAD Set of Identifications, Measurements, and Bibliography for Astronomical Data.

SPOTS Stellar Parameters Of Tracks with Starspots.

TORCH Theoretical Open cluster Radial velocity CHallenge.

VISTA Visible and Infrared Survey Telescope for Astronomy.

VLT Very Large Telescope.

VPHAS+ VST Photometric Hα Survey of the Southern Galactic Plane.

VVV VISTA Variables in the Vía Láctea.

VVVX VVV eXtended survey.

YSO Young stellar object.

ZAMS Zero-age main sequence.

Software Packages

ArviZ Bayesian model diagnostics and visualisation.

ASteCA Automated Stellar Cluster Analysis (Perren, Vázquez, and Piatti, 2015).

AstroPy Core Python astronomy package (The Astropy Collaboration et al., 2022).

gaiadr3_zeropoint Gaia DR3 parallax zero-point correction (Lindegren et al., 2021a).

HDBSCAN Density-based clustering (Campello, Moulavi, and Sander, 2013; McInnes, Healy, and Astels, 2017).

PyMC Probabilistic programming and MCMC sampling.

Sagitta PMS classification neural network (McBride et al., 2021).

scikit-learn Machine learning library (Pedregosa et al., 2012).

Mathematical Symbols

Astrometric parameters

α, δ Right Ascension and Declination (deg).

ℓ, b Galactic longitude and latitude (deg).

ϖ	Parallax (mas).
$\varpi_{\rm corr}$	Zero-point corrected parallax (mas).
$\varpi_{\rm zp}$	Parallax zero-point offset (mas).
$\mu_{\alpha*}$	Proper motion in RA, $\mu_\alpha \cos\delta$ (mas yr^{-1}).
μ_δ	Proper motion in Declination (mas yr^{-1}).
σ_μ	Proper-motion dispersion (mas yr^{-1}).
v_r	Radial velocity (km s^{-1}).
$v_{\rm proj}$	Projected velocity on the plane of the sky (mas yr^{-1}).
d	Distance (kpc).
$f_{\rm fid}$	Astrometric fidelity score.

Photometric parameters

G, $G_{\rm BP}$, $G_{\rm RP}$	Gaia broadband magnitudes (mag).
J, H, K_s	2MASS near-infrared magnitudes (mag).
M_G	Absolute G-band magnitude (mag).
A_V	Visual extinction (mag).
$E(B-V)$	Colour excess (mag).
R_V	Total-to-selective extinction ratio.
$\mu_{\rm DM}$	Distance modulus (mag).
Q	Reddening-free photometric parameter (mag).

Cluster physical parameters

τ	Cluster age (Myr).
Z	Metallicity (mass fraction).
m_c	Total cluster mass ($M_\odot$).
$\langle m \rangle$	Mean stellar mass ($M_\odot$).
$R_{\rm GC}$	Galactocentric distance (kpc).
$Y_{\rm frac}$	Young stellar object fraction.

Structural parameters

R_c	Core radius, King model (arcmin / pc).
R_t	Tidal radius, King model (arcmin / pc).
$R_{\rm hl}$	Half-light radius (arcmin / pc).
$R_{\rm hm}$	Half-mass radius (arcmin / pc).
$R_{\rm Hill}$	Hill (Jacobi) radius (arcmin).
$R_{\rm bound}$	Gravitationally bound radius (arcmin).
k	King profile central density normalisation (stars arcmin^{-2}).
b	King profile background density (stars arcmin^{-2}).
C	Concentration parameter, $\log_{10}(R_t/R_c)$.

Dynamical parameters

$t_{\rm rh}$	Half-mass relaxation time (Myr).
$t_{\rm cr}$	Crossing time (Myr).
$t_{\rm seg}$	Mass segregation timescale (Myr).
$\eta = \tau/t_{\rm rh}$	Dynamical age parameter.

Membership and inference

$\tilde{p}_i$	Membership pseudoprobability of star i.
$p_{\rm PMS}$	Pre-main-sequence probability (Sagitta).
f_i	HDBSCAN iteration fraction.
$p_{{\rm HDBSCAN},i}$	HDBSCAN λ-persistence value.
$m_{\rm cl}$	HDBSCAN minimum cluster size parameter.
λ	HDBSCAN inverse mutual reachability distance.
D_{AB}	Kolmogorov–Smirnov test statistic.

Isochronal model parameters

$\alpha_{\rm MF}$	Multiplicity fraction (Offner et al., 2022).
$\beta_{\rm MF}$	Multiplicity fraction slope.
q	Binary mass ratio (m_2/m_1).
σ	ASteCA CMD dispersion parameter.

Astronomical Objects

NGC 6383 Young open cluster in Scorpius; primary subject of this thesis.

HD 159176 O6.5 V + O7 V double-lined spectroscopic binary; projected near the cluster centre.

NGC 6383 22 λ Bootis star; non-member passing through the cluster field.

Sh 2-012 H II region surrounding NGC 6383.

Sgr OB1 Sagittarius OB1 association; parent complex of NGC 6383.

Chapter 1

Conceptual framework and scientific context

Young open clusters occupy a distinctive interpretive regime in which the imprint of formation, the effects of early dynamical processing, and the consequences of observational selection remain entangled in the data. This chapter establishes the conceptual framework within which the physical analysis of NGC 6383 will be conducted and the scientific questions that the thesis is designed to address. We begin by reviewing the role of open clusters as laboratories for hierarchical star formation, early gas expulsion, and initial mass function inference, with emphasis on why the observable present-day structure of a young cluster cannot be read as a transparent record of its birth conditions (Section 1.1). The dynamical youth of such systems—incomplete relaxation, ambiguous virial states, and the contested causal origins of mass segregation—is examined in Section 1.2. Section 1.3 frames the membership problem as it stands in the Gaia era: the passage from high-dimensional phase-space data to a probabilistic population census, and the consequences of that construction for every downstream physical claim. The broader methodological transition from cluster detection to physical inference,

including the role of Bayesian modelling and integrated pipelines, is developed in Section 1.4. NGC 6383 is then introduced not as an isolated target but as a particularly instructive testbed in which nine decades of parameter disagreement, environmental complexity, and the historical role of HD 159176 make the thesis-level tensions unusually visible (Section 1.5). The chapter closes by converting these tensions into a finite scientific programme: the three central questions and supporting problems that define the scope of the dissertation (Section 1.6).

1.1 Open clusters as dynamical laboratories

1.1.1 Hierarchical star formation

Open clusters are physically informative not because they preserve a pristine record of their birth conditions, but because their present-day stellar distributions remain shaped by the coupled effects of formation, feedback, and early dynamical evolution. In young systems, the relevant initial state is not a smooth equilibrium configuration. Rather, both theory and observation support a picture in which stellar structure emerges within a multiscale, turbulence-shaped gaseous hierarchy, so that the observable cluster is the processed stellar imprint of a larger ISM geometry (Menon et al., 2021; Hota, Grijs, and Subramaniam, 2025). The key problem is therefore not simply whether stars form in clusters, but how hierarchical gas structure is converted into a stellar system with some degree of dynamical coherence, and how much memory of that hierarchy survives into the observable phase.

This hierarchical view is supported across both simulations and resolved stellar surveys. Young stellar structures exhibit scale-free or fractal-like clustering statistics consistent with turbulence-regulated assembly rather than with formation in

isolated, monolithic units (Menon et al., 2021; Hota, Grijs, and Subramaniam, 2025). Theoretical frameworks differ in how they interpret the subsequent evolution of that hierarchy: some emphasize continuous global hierarchical collapse in converging flows (Vázquez-Semadeni et al., 2019), whereas others parameterise cluster formation efficiency and early environmental disruption separately (Pfeffer et al., 2018). The thesis-relevant consequence is that the stellar system inherited from star formation is expected to be structured, time-dependent, and not straightforwardly reducible to a relaxed equilibrium precursor. That inherited structure must then be reworked further by early gas loss and subsequent dynamical evolution.

This caution is reinforced by work on massive protoclusters and embedded star-forming regions. ALMA-based studies and related models increasingly favour filament-fed, time-dependent accretion networks over a simple one-core/one-star mapping, implying that compact structures in the gas need not correspond to long-lived, isolated stellar building blocks (Sandoval-Garrido et al., 2025; Padoan et al., 2020; Sanhueza et al., 2019; Sabatini et al., 2022). In parallel, coupled gas-plus-stellar simulations show that hierarchical stellar assembly cannot be understood through collisionless arguments alone: close interactions, subcluster merging, and gas-regulated dynamics all influence whether hierarchical structure is erased, partially retained, or converted into higher-order multiples and long-lived kinematic signatures (Grudić et al., 2021).

Observationally, the erasure of hierarchy is neither instantaneous nor universal. Gaia-era kinematics suggest that any large-scale coalescence must often be largely complete by the end of the embedded phase, while chronometric studies continue to indicate spatially organized age spreads and non-instantaneous assembly histories (Kuhn et al., 2019; Getman et al., 2018; Getman et al., 2022). More generally, hierarchical spatial structure can persist for tens of Myr, with the exact decay

timescale depending on environment and shear (Menon et al., 2021). Young clusters should therefore not be read as transparent birth relics, but as dynamically processed outcomes of hierarchical formation whose observable structure already folds together star formation, feedback, and early stellar dynamics.

1.1.2 Early gas expulsion

If hierarchical assembly sets the initial stellar structure, early gas removal is the main process that rapidly reorganizes it. Gas expulsion can leave behind a bound remnant, produce a largely unbound expanding association, or generate an intermediate state in which a compact core coexists with escaping material. Because this mass loss can occur on timescales comparable to or shorter than the crossing time, the relevant regime is intrinsically non-equilibrium (Dinnbier and Kroupa, 2020; Dinnbier and Walch, 2020). The scientific issue is therefore not whether gas matters in a generic sense, but how the timing and spatial coupling of gas removal condition later survivability, expansion, and structural interpretation.

Idealized N-body models established the baseline point that cluster survival after gas loss is not controlled by a universal star-formation-efficiency threshold. Smooth models show strong sensitivity to the gas-removal timescale and to how the stellar mass is distributed relative to the gas, with centrally concentrated configurations surviving more easily than diffuse ones (Baumgardt and Kroupa, 2007; Shukirgaliyev et al., 2019). More recent work sharpens the same conclusion in a different language: when primordial substructure and non-equilibrium kinematics are included, survivability depends strongly on local density enhancements, correlated motions, and the pre-expulsion dynamical state (Domínguez et al., 2021; Grudić et al., 2021). What matters for the NGC 6383 analysis is that present-day structure is shaped not only by how much stellar mass formed, but by how that mass was arranged in phase space when the potential changed—a phase-space

inheritance that Chapter 7 will attempt to recover.

This is why gas-expulsion signatures remain astrophysically relevant in the Gaia era. Proper-motion studies show that expansion is common among $\sim$ 1–5 Myr systems, supporting the view that many young clusters and associations are observed during or after a phase of active unbinding (Kuhn et al., 2019). At the same time, simulations indicate that escaping populations and tidal debris can retain information about the expulsion history and initial virial state (Dinnbier and Kroupa, 2020). Even among coeval systems, divergent outcomes are possible: some retain a compact, re-virializing core, whereas others evolve rapidly toward dissolution (Pang et al., 2020). Early gas loss therefore does not merely perturb an otherwise simple cluster; it helps determine which stars remain in the system, how the outer structure should be interpreted, and how later dynamical diagnostics must be read.

Recent work has sharpened this picture further. Wright et al. (2024) showed that all 18 young groups in the Gaia-ESO Survey exhibit supervirial velocity dispersions when only stellar mass is considered, confirming that expansion signatures are ubiquitous in systems younger than $\sim$ 10 Myr. At the simulation frontier, Ćalović et al. (2025) challenged the classical gas-expulsion framework itself: when stars inherit realistic velocities from hydrodynamic formation simulations rather than being initialised in idealised virial equilibrium, the post-gas-removal evolution proceeds closer to virial balance than the instantaneous-expulsion models predict. If confirmed, this result implies that the dynamical state of young clusters like NGC 6383 may require reinterpretation beyond the classical SFE-dependent survival framework of Baumgardt and Kroupa (2007). These developments motivate the careful dynamical assessment undertaken in Chapter 7.

1.1.3 Clusters and the IMF

Open clusters are also a major empirical route to IMF inference, but in young systems that inference is inseparable from dynamical and observational processing. The present-day stellar census is shaped by incompleteness, multiplicity, mass segregation, and early mass loss, so the observed mass distribution is not automatically equivalent to the birth IMF (Hopkins, 2018). The relevant doctrine is that cluster mass-function inference is conditional on how the stellar population has been selected, retained, and modelled.

That caution is now especially clear in Gaia-era work. Modern cluster analyses can derive apparently stable IMF-like slopes under homogeneous membership and binary-aware mass estimation (Alfonso, García-Varela, and Vieira, 2024), yet both extreme environments and individual massive clusters continue to show either meaningful deviations or internally inconsistent slopes depending on the mass range, completeness, and treatment of multiplicity (Hosek Jr. et al., 2019; Lim et al., 2013). Unresolved binaries are a central part of this problem, since they broaden the CMD and bias inferred masses high if they are treated deterministically or ignored (Alexander and Albrow, 2024). More generally, the literature shows that apparent IMF variation may reflect a combination of birth physics, multiplicity, selection, and dynamical processing rather than a uniquely identifiable change in the underlying stellar mass spectrum (Hopkins, 2018; Alexander and Albrow, 2024).

Cluster populations are dynamically and observationally processed, and even apparently basic population inferences are conditional on membership construction, multiplicity treatment, and early evolution.

1.2 Dynamical youth and non-equilibrium states

1.2.1 Virial states

Young clusters are frequently observed in a regime where neither their phase-space structure nor their potential is stationary. In that setting, the central problem is not simply whether a system should be labelled bound or unbound, but what dynamical interpretation is warranted when expansion, residual substructure, and incomplete mixing remain important. Virial language is still useful, but only if applied cautiously.

The literature shows that survivability after rapid gas loss depends more strongly on the instantaneous pre-expulsion dynamical state than on any single global efficiency proxy. In particular, the virial ratio and the local coupling between stars and gas are more informative than a global star-formation-efficiency number alone (Farias et al., 2015). This is complicated further by the fact that hierarchically assembled young regions can oscillate between subvirial and supervirial states on short timescales, so the timing of gas removal relative to those oscillations can materially affect the outcome (Dominguez et al., 2017). The relevant lesson is not that one scalar fully diagnoses the problem, but that virial inference in young systems is phase-dependent and intrinsically non-equilibrium.

Gaia-era observations confirm that expansion signatures are common, but they also show why raw kinematic patterns cannot be translated directly into internal virial states. Proper-motion analyses recover widespread expansion in young populations (Kuhn et al., 2019), yet intrinsic dispersions depend on how one models geometry, perspective effects, correlated astrometric errors, and binaries (Wright et al., 2024; Cantat-Gaudin, 2022). Binary contamination is especially important because unresolved orbital motion can inflate dispersions enough to bias virial classifications if left untreated (Jackson et al., 2021; Hunt and Reffert, 2024;

Pianta, Capuzzo-Dolcetta, and Carraro, 2022). Binaries also matter physically, not merely observationally, since early hardening can inject energy into the system and influence later expansion (Torniamenti et al., 2021; Karam and Sills, 2022).

Apparent supervirial behaviour in a young cluster therefore cannot be read at face value: it may reflect genuine expansion, a compact core embedded in an extended unbound component, measurement geometry, unresolved binaries, or some combination of these (Meingast, Alves, and Rottensteiner, 2021). The formal framework for reading such signals is developed in Chapter 2.

1.2.2 Relaxation timescales

The persistence of non-equilibrium structure also reflects the fact that relaxation operates on multiple spatial and dynamical scales. Dense cores, sub-clumps, extended halos, and tidally affected outskirts do not evolve on the same timescale, so a young cluster cannot usually be assigned a single global dynamical age in any simple sense. Classical half-mass relaxation arguments remain important, but only as part of a stratified picture in which local and global evolution can differ sharply (Baumgardt, 2001).

This multi-timescale view is reinforced by both internal and external effects. External tidal shocks do not simply heat and dissolve clusters monotonically; they can preferentially remove loosely bound outer layers while leaving a denser, more rapidly relaxing remnant (Gieles and Renaud, 2016). At the same time, classical equipartition expectations do not hold in realistic multi-mass systems. Full equipartition is dynamically inaccessible under common cluster conditions (Spitzer Jr., 1969; Bianchini et al., 2016b), and late evolution can remain only partially thermalized because massive stars, remnants, and hard binaries act as persistent internal heat sources (Trenti and Marel, 2013; Spera, Mapelli, and

Jeffries, 2016; Vishniac, 1978). The practical implication is that relaxation-based reasoning is informative, but it does not license simple equilibrium readings.

Observationally, this leads naturally to radially stratified interpretation. Nearby clusters can host relatively mature inner regions while retaining extended, anisotropic, or evaporating outer populations (Meingast, Alves, and Rottensteiner, 2021). Local primordial substructure may also relax faster than the cluster as a whole, so early dynamical processing can occur in sub-clumps well before the global system approaches any meaningful equilibrium (Arnold and Wright, 2024; Kuhn et al., 2019). Relaxation is therefore neither global nor single-timescale, and any later inference about dynamical state or segregation must respect that stratification.

1.2.3 Mass segregation origins

Mass segregation is one of the clearest examples of why young-cluster interpretation is regime-dependent. A centrally concentrated population of massive or binary systems may reflect primordial or near-primordial inheritance, rapid local dynamical processing in dense substructure, later secular evolution, or selective depletion of the low-mass halo. The signal is therefore not self-interpreting.

Methodologically, this problem is already visible in the diversity of segregation diagnostics. Geometry-robust approaches such as minimum spanning tree statistics avoid the need for a unique cluster centre and show that strong segregation can arise on short timescales in substructured, subvirial systems (Allison et al., 2009). More realistic simulations broadly support the same cautionary point: early concentration of massive stars can emerge rapidly without requiring global primordial segregation by construction (Polak et al., 2025; Guszejnov et al., 2022a). At the same time, local density ratios and related multi-statistic approaches

show that local formation bias need not imply global central concentration, so segregation can differ by spatial scale and by metric (Dib, Schmeja, and Parker, 2018; Buckner et al., 2019).

The dichotomy between primordial and dynamical mass segregation is itself dissolving. TORCH radiation-hydrodynamic simulations by Polak et al. (2025) demonstrate that massive star clusters develop significant mass segregation within ~ 2 Myr from initially non-segregated conditions, driven by dynamical friction during the collapse of substructure, a process that the embedded gas dynamics does not inhibit rather than by long-term two-body relaxation. Observational support for a primordial channel has also strengthened: Motherway et al. (2024) showed that binaries in the ~ 200 Myr cluster M35 remain more centrally concentrated than single stars even when the analysis is restricted to masses whose segregation timescales exceed the cluster age, providing direct evidence for primordially segregated binary populations—a finding that echoes the binary segregation signature detected in NGC 6383 (Chapter 7).

Population-level Gaia studies extend the problem rather than resolving it. Segregation appears widespread among open clusters and generally strengthens with age, but the observed radial distributions are also shaped by tidal stripping and halo depletion (Tarricq et al., 2022). In that sense, apparent segregation is not only an internal-relaxation signature; it can also be amplified by the preferential tidal removal of low-mass stars from the outskirts (Baumgardt and Makino, 2003). Additional ambiguity arises because different initial conditions can evolve into similar present-day density structures, so morphology alone does not uniquely recover causal origin (Daffern-Powell and Parker, 2020).

Segregation is therefore an interpretive problem, not a pre-classified result—physically informative under caution, but not causally closed by Chapter 1 alone. Its full reading requires the dynamical and membership framework developed in

Chapters 2–4.

1.3 The membership problem in the Gaia era

1.3.1 High-dimensional phase space

Gaia transformed cluster membership from a largely visual or low-dimensional identification problem into a statistical separation problem in high-dimensional measurement space. Cluster populations are now identified primarily through their occupation of compact loci in astrometric phase space, not through projected overdensity alone (Cantat-Gaudin and Casamiquela, 2024; Gaia Collaboration et al., 2023). This shift is scientifically powerful, but it also means that membership depends on how one represents the data geometry.

That geometry is not Euclidean in any naive sense. Gaia astrometry is heteroscedastic, correlated, and affected by structured systematics tied to the scanning law and calibration model (Lindegren et al., 2021a). Proper-motion precision and parallax quality vary strongly with magnitude, and credible membership pipelines must therefore treat the covariance matrix as part of the data rather than as an afterthought. This is not a merely technical nicety: the way uncertainties are represented conditions which sources remain compatible with the cluster locus and which are excluded as interlopers (Deb, Baruah, and Kumar, 2022; Lindegren et al., 2021a).

The extension from 5D to 6D phase space adds real discriminating power where radial velocities are available, especially in complex or distant fields (Katz et al., 2023; Carrera et al., 2019a; Jackson et al., 2021). But that additional dimension is incomplete survey-wide, so most membership problems remain partially observed. The more general point is that dimensionality helps, but it does not eliminate

ambiguity. Membership remains conditioned by measurement geometry, catalogue selection, and the available tracer set.

This is why the Gaia-era membership problem cannot be separated from selection effects. Completeness depends on magnitude, crowding, extinction, and local background structure, so the observable member distribution is itself survey-conditioned (Hunt et al., 2025; Gaia Collaboration et al., 2023). Membership in the Gaia era is therefore an inference problem over structured, uncertain, and selection-limited data.

1.3.2 Field contamination

Greater dimensionality does not turn contamination into a solved problem. It changes the problem from one of random background noise into one of structured dynamical superposition. The Galactic field contains streams, moving groups, extinction-modulated backgrounds, and young extended populations that can mimic cluster-like structure in either astrometric space or the CMD (Bonaca and Price-Whelan, 2025; Cantat-Gaudin, 2022; Queiroz et al., 2023; Helmi, 2020). As a result, membership must be evaluated against a structured background, not a featureless field.

This helps explain why apparently coherent kinematic discoveries can weaken under improved diagnostics. Gaia-era overdensities or filamentary structures may look compelling in one projection and then fail under radial-velocity, dynamical-consistency, or improved astrometric tests (Kounkel and Covey, 2019; Kounkel, Covey, and Stassun, 2020). The general lesson is thesis-relevant and simple: kinematic coherence can be necessary for membership, but it is not always sufficient.

Contamination therefore forces an explicit completeness–purity trade-off. Relaxing

thresholds improves recovery of diffuse halos and outskirts, but it also raises the number of interlopers, often sharply because the background itself is structured (Hunt and Reffert, 2023). This means that membership thresholds are not merely technical settings; they are objective-function choices tied to the scientific aim. A purity-dominated study and a completeness-dominated study do not recover the same operational member list, even in the same field (Tarricq et al., 2022).

Photometric and chemical information provide important secondary constraints, but they must be used carefully. CMD coherence can validate astrometric candidates without necessarily dictating the clustering stage (Hunt and Reffert, 2023; Hao et al., 2022), while chemical tagging or precise radial velocities can adjudicate the hardest phase-space degeneracies (Carrera et al., 2019b; Helmi, 2020). For this chapter, the point is not to prescribe a universal pipeline, but to establish that contamination is local, structured, and inseparable from the inferential meaning of membership.

1.3.3 Deterministic vs probabilistic criteria

These conditions make purely deterministic membership criteria inadequate. Hard cuts in parallax and proper-motion space assume symmetric distributions and homogeneous uncertainties, whereas real cluster populations can be asymmetric, extended, or partially overlapping with the field, and Gaia errors are heteroscedastic and correlated (Cantat-Gaudin et al., 2018). The result is not just classificatory inconvenience. Hard boundaries can exclude low-mass members preferentially, bias the observed mass function, distort binary demographics, and misrepresent the uncertainty attached to the adopted sample (Monteiro et al., 2020; Donada et al., 2023; Dias et al., 2021).

Probabilistic frameworks address this by treating the observed sample as a mixture

of cluster and field populations and assigning soft membership probabilities rather than fixed labels (Dias et al., 2021). Different implementations vary in topology and assumptions, but they share the crucial advantage that they retain ambiguity in a form that remains scientifically meaningful for later physical interpretation (Deb, Baruah, and Kumar, 2022; Agarwal et al., 2021). The point is not to choose a universal winner among GMMs, deconvolution methods, density-based clustering, or hybrids.

The same logic applies to binaries. Strict quality cuts or deterministic boundaries often reject unresolved binaries because their astrometry or photometry deviates from a single-star locus, yet those systems are astrophysically real and materially affect mass, dispersion, and CMD inference (Belokurov et al., 2020; Monteiro et al., 2020; Yalyalieva et al., 2022). Probabilistic frameworks can model that additional structure rather than forcing it into noise.

The broader conclusion is therefore foundational to the thesis: Gaia did not reduce membership to a cleaner version of classical catalogue filtering. It exposed membership as a coupled inference problem involving measurement geometry, contamination, selection, and astrophysical structure.

1.4 From detection to physical inference in Gaia-era open clusters

1.4.1 Clustering as density estimation

The Gaia-era detection problem is best understood not as the search for geometric balls in parameter space, but as the identification of persistent overdensities in a heterogeneous and anisotropic density field. In practice, this is why different clustering approaches recover different cluster morphologies, halos, and false

positives in the same data (Hunt and Reffert, 2020; Rodríguez-Casal and Saavedra-Nieves, 2022). The scientific consequence is that a detected cluster is already an inference product conditioned by metric choice, density definition, and local background structure.

This matters because single-scale methods and multiscale methods encode different assumptions about what should count as a real structure. Variable-density environments, extended outskirts, and tidal-like morphologies are not handled equally well by algorithms built around one global density threshold (Campello, Moulavi, and Sander, 2013; Hunt and Reffert, 2020). Covariance-aware distances provide a complementary correction by reshaping the geometry of the problem in a way that respects Gaia measurement uncertainties (Alfonso, García-Varela, and Vieira, 2024). Parametric mixture models, in turn, trade topological flexibility for soft assignments and explicit generative structure (Jaehnig, Bird, and Holley-Bockelmann, 2021; Küpper et al., 2010). Cluster detection is therefore not observationally transparent, and different operational definitions of structure can produce materially different candidate populations.

This becomes even more important once selection effects are admitted. Detection efficiency depends on intrinsic cluster richness, measurement precision, background density, and kinematic contrast with the field (Hunt et al., 2026). Cluster catalogues are therefore conditional products of both the discovery algorithm and the survey selection function. That is precisely why the chapter must move beyond detection alone.

1.4.2 Bayesian inference as uncertainty quantification

Once a candidate population has been identified, physical inference becomes an uncertainty-propagation problem rather than a matter of reading single best-

fit parameters from the data. Ages, distances, extinction, and related cluster properties are all affected by non-linear degeneracies, uncertain membership, survey truncation, and unresolved multiplicity (Bossini et al., 2019; Perren, Vázquez, and Piatti, 2015; Cerviño and Luridiana, 2004). Bayesian inference is useful here not because it is fashionable, but because it makes those degeneracies explicit and forces the model to represent uncertainty as an output rather than as an afterthought.

Hierarchical formulations sharpen that logic by coupling star-level latent quantities to population-level hyperparameters, allowing heterogeneous uncertainties and partial missingness to propagate coherently (Moya et al., 2022; Hippel et al., 2006). Priors are not cosmetic regularisers in this setting; they encode physically necessary constraints when the data do not by themselves identify the model uniquely. Binary fractions, magnitude limits, and incompleteness are especially important because if they are ignored, apparent CMD broadening or population imbalance can be misread as age structure, mass variation, or distance effects (Perren, Vázquez, and Piatti, 2015; Cerviño and Luridiana, 2004).

Cluster parameters are therefore not direct observables but posterior quantities conditioned by membership, priors, selection, and model assumptions. The operational implementation of that doctrine belongs to Chapters 4 and 6.

1.4.3 Integrated pipelines for bias control and reproducibility

The move from detection to physical inference also exposes a structural weakness in the classical two-stage workflow. If membership is fixed first and the resulting labels are then treated as ground truth during physical modelling, uncertainty is artificially compressed and selection-induced bias is introduced (Vasiliev, 2019; Sollima, 2020). Marginal members, outskirts, and contamination structure are

then removed from the inference precisely where they matter most for structural and dynamical interpretation.

The remedy is conceptual more than technological: membership, contamination, and physical parameters should be treated as coupled components of a single inferential problem whenever the science requires it (Vasiliev, 2019; Pang et al., 2020). Whether that is implemented through a fully unified graphical model or through a carefully layered probabilistic workflow is secondary at the level of Chapter 1. The key point is that physical interpretation should be conditioned on membership uncertainty rather than insulated from it. Otherwise, uncertain sample boundaries are converted into fixed truth, and the resulting ages, structures, or dynamical readings can be biased precisely by the membership ambiguity that the analysis claims to control.

Automated and homogeneous pipelines are useful insofar as they reduce irreproducible hand-tuning and make cluster analyses comparable at catalogue scale (Perren, Vázquez, and Piatti, 2015; Dias et al., 2021). But in this thesis they remain operational means, not scientific ends. The chapter therefore retains only the part of the pipeline argument that matters downstream: integrated inference is needed because the scientific quantities of interest are conditional on how detection, contamination control, and uncertainty propagation are connected.

1.5 NGC 6383 as a testbed

NGC 6383 becomes relevant at this point because it concentrates within a single field many of the chapter's central tensions: youth, mixed stellar populations, environmental complexity, strong contamination, and method-dependent disagreement over the basic physical characterization of the system (Kalari, 2019; Paunzen, Netopil, and Zwintz, 2007; Rauw et al., 2003). Located within Sh 2-012 and

historically associated with the massive binary HD 159176, the cluster is observed in a regime where stellar feedback, residual gas structure, and incomplete dynamical processing remain entangled (De Becker et al., 2004; Rauw, Manfroid, and De Becker, 2010). It is therefore not introduced here as an object dossier, but as a case in which membership, structure, age, and interpretation cannot be cleanly separated.

A limited subset of the observational picture is sufficient to justify that role. NGC 6383 contains both an upper main-sequence population and a substantial pre-main-sequence component, with Hα emitters, infrared excess sources, X-ray-selected PMS stars, and related youth indicators showing that contraction and activity remain important (Zwintz et al., 2005; Kalari, 2019; Paunzen, Netopil, and Zwintz, 2007; Rauw et al., 2003). At the same time, the surrounding region contains bright rims, dark clouds, and star-formation activity beyond the immediate core (Rauw, Manfroid, and De Becker, 2010; Kalari, 2019). This places the system in an especially useful intermediate regime: it is not so embedded that Gaia-era astrometry becomes largely unusable, but it is not dynamically old enough for present structure to be read as a simple secular endpoint.

Its diagnostic value is strengthened by the fact that the literature disagreements follow methodological fault lines that mirror the thesis questions. Published distances and ages have varied substantially, from early photographic estimates (Trumpler, 1930) through photoelectric work (Fitzgerald et al., 1978) to modern spectrophotometric (Aidelman et al., 2018) and Gaia-era determinations (Cantat-Gaudin et al., 2018; Cantat-Gaudin et al., 2020), and Gaia-era work reduces but does not erase that instability. The point is not merely that the literature once disagreed. It is that the disagreement itself shows that cluster properties are conditioned by sample definition and tracer choice. In that sense, NGC 6383 is already an example of membership-conditioned physical inference before the later

Table 1.1: Historical review of observed parameters for NGC 6383. Columns list the number of members, core radius R_c, tidal radius R_t, parallax ϖ, distance d, colour excess $E(B-V)$, distance modulus $(m-M)$, and age. Entries marked $\cdots$ indicate that the parameter was not reported in the original study.

Reference	Members	R_c	R_t	ϖ	d	$E(B-V)$	$(m-M)$	Age
		arcmin	arcmin	mas	kpc	mag	mag	Myr
Trumpler (1930)	···	···	···	···	2.13	···	···	···
Sanford (1949)	···	···	···	···	0.76	···	···	···
Eggen (1961)	···	···	···	···	···	0.30	···	···
Garrison (1967)	74	···	···	···	1.38	···	10.7 ± 0.5	···
Evans (1968)	≥ 21	···	···	···	1.30	0.35	10.6	5.0
Lindoff (1968)	···	···	···	···	1.25	···	10.5	~ 20
Becker and Fenkart (1971)	···	···	···	···	1.07	0.26	10.9	···
Fitzgerald et al. (1978)	···	1.25	···	···	1.50 ± 0.20	0.33 ± 0.02	10.9	1.7 ± 0.4
Lloyd Evans (1978)	···	···	···	···	1.35	0.35	10.6	≤ 5.0
The et al. (1985)	···	···	···	···	1.40 ± 0.15	0.30 ± 0.01	···	···
Pandey et al. (1989)	···	···	···	···	···	0.35	11.7	~ 4.5
Battinelli and Capuzzo-Dolcetta (1991)	27	···	···	···	1.38	0.34	···	~ 20
Feinstein (1994)	···	···	···	···	1.40	0.33	···	···
Rauw et al. (2003)	···	···	···	···	···	0.33	10.7	···
Kharchenko et al. (2005)	13	4.80	15.0	···	0.99	0.30	10.9	5.0
Paunzen, Netopil, and Zwintz (2007)	···	···	···	···	1.70 ± 0.30	0.29 ± 0.05	···	< 4.0
Piskunov et al. (2008)	···	···	29.0 ± 6.6	···	0.99	0.30	10.9	5.0
Rauw, Manfroid, and De Becker (2010)	···	···	···	···	···	0.32	···	2–3
Aidelman et al. (2018)	···	···	···	···	0.83 ± 0.16	0.51 ± 0.03	9.61 ± 0.38	~ 3–10
Kalari (2019)	≥ 55	···	···	···	···	···	···	2.8 ± 1.6
Jaehnig, Bird, and Holley-Bockelmann (2021)	284	···	···	0.93 ± 0.09	1.07	···	···	···
He et al. (2022)	···	···	···	0.88 ± 0.05	···	0.45	···	3.5
Hunt and Reffert (2024)	322	2.49	22.7	0.89 ± 0.08	1.10	0.38	10.2	~ 4.0
Pulgar-Escobar et al. (2024)	254	1.95	40.4	0.91 ± 0.06	1.11 ± 0.06	0.40 ± 0.08	10.3 ± 0.26	$3.53^{+1.40}_{-1.00}$

chapters formalize it more explicitly. The historical evolution of published cluster parameters is summarised in Table 1.1.

The same is true of membership and structural extent. Earlier work often identified the cluster through compact radial selections or photometric overdensities in a contaminated Galactic-plane field (Fitzgerald et al., 1978; Kharchenko et al., 2005; Paunzen, Netopil, and Zwintz, 2007), whereas Gaia-era studies recover a more extended and method-dependent system with a sparse halo and ambiguous periphery (Cantat-Gaudin et al., 2020; Hunt and Reffert, 2020; Tarricq et al., 2022). That difference is scientifically substantive. A compact photometric core embedded in a broader and uncertain outer population is exactly the kind of object for which both false positives and false negatives can materially alter the interpretation of age, structure, and segregation.

NGC 6383 also remains dynamically and environmentally young enough for non-equilibrium caution to matter. Its broader setting within Sh 2-012, together with evidence for preserved substructure and ongoing environmental complexity, argues

against treating it as a simple relaxed open cluster (Hetem and Gregorio-Hetem, 2019; Cartwright and Whitworth, 2004; Rauw, Manfroid, and De Becker, 2010). Yet the object should not be over-converted into a universal template. It does not by itself settle whether apparent PMS broadening reflects extended star formation, extinction, accretion, unresolved multiplicity, or model dependence (Paunzen, Netopil, and Zwintz, 2007; Aidelman et al., 2018). Nor does the historical role of HD 159176 justify allowing that star to dominate the thesis architecture. Its importance is subordinate and corrective: it illustrates how inherited physical narratives can become unstable once membership itself is reconsidered.

Despite these tensions, NGC 6383 is not merely a difficult case; it is a specifically *tractable* one. Three features make it informative rather than intractable for the present investigation. First, the membership census at $\tilde{p} > 0.6$ yields 254 stars—large enough for statistically meaningful structural, kinematic, and population inference, yet small enough for individual-star treatment within a hierarchical Bayesian framework. Second, at $d \approx 1.1$ kpc the cluster lies within the regime where individual Gaia DR3 parallaxes carry useful per-star information (fractional errors below ~ 10 per cent for most members brighter than $G \approx 18$ mag), so that distance inference is not reduced to a purely population-level constraint. Third, the coexistence of an upper main-sequence population and a substantial pre-main-sequence component provides independent age cross-checks that are absent in either purely embedded or dynamically old systems. These properties do not resolve the interpretive tensions catalogued above, but they define the observational envelope within which those tensions can be addressed quantitatively.

NGC 6383 is therefore a testbed in the strict chapter-level sense. It concentrates the thesis-relevant tensions of youth, structured environment, membership ambiguity, parameter instability, and inference-sensitive interpretation. That is enough to motivate its central role in the dissertation without turning this section into a

condensed paper introduction.

1.6 Scientific questions

The preceding sections have established that young open clusters cannot be read straightforwardly from a single membership list, a single structural profile, or a single age indicator. Rather, they are dynamically young, inference-sensitive systems in which contamination, tracer choice, incomplete mixing, environmental complexity, and model dependence remain entangled with the physical conclusions drawn from the data. Within that broader framework, NGC 6383 is relevant not as an isolated curiosity, but as a particularly instructive Gaia-era testbed in which these difficulties become unusually visible. The purpose of this section is therefore not to recapitulate the conceptual material already developed in Sections 1.1–1.5, nor to anticipate the formal detail of Chapter 2, but to convert those tensions into a finite scientific programme: to define what this thesis is actually trying to decide, what it seeks only to constrain or clarify, and what must remain open until the later physical and empirical analysis is brought to bear.

A first central question concerns *membership-conditioned physical inference*. In the Gaia era, membership cannot be treated as a merely preparatory filtering step whose only role is to clean the sample prior to later analysis. Rather, membership construction conditions the very population to which age, structure, kinematics, and segregation are subsequently attributed. For a system such as NGC 6383, whose published parameters have varied substantially across studies, this is not a narrowly technical matter about sample selection, nor a question of algorithmic preference. The scientific issue is not only whether a probabilistic census can be obtained, but how strongly the adopted inferential construction of membership reshapes the physical interpretation that may legitimately be assigned to the

cluster. In a case of this kind, different membership constructions may select materially different stellar samples, with correspondingly different inferred cluster properties and physical interpretations, and revised membership may alter not only the recovered stellar content and structural extent of the system, but also the status of inherited narratives anchored to historically prominent tracers. In that sense, the thesis is not centred on membership as a technical preprocessing problem, but on membership-conditioned physical inference as a foundational scientific problem.

A second central question concerns the *dynamical-state interpretation* of a young, structurally complex system. Chapter 1 has already shown that youth, early expansion, incomplete phase-space mixing, and environmental coupling make equilibrium-style readings hazardous when applied too quickly to clusters near their formation epoch. The issue for this thesis is therefore not to impose a simple classificatory label at the outset, nor to collapse the problem into a single diagnostic. The real scientific question is what kind of dynamical interpretation is warranted once NGC 6383 is treated under the full caution demanded by its youth, by the uncertainties affecting its membership and structure, and by the possibility that some aspects of its present configuration remain non-equilibrium in origin. The Gaia DR3 radial velocity coverage for a cluster of this distance and richness is known *a priori* to be insufficient for a definitive scalar virial ratio; the thesis therefore does not promise a single classificatory label, but rather asks what physical claims about the present dynamical state of NGC 6383 remain defensible from proper-motion kinematics, structural diagnostics, and relaxation-timescale arguments, and what must remain inferentially open until three-dimensional kinematic data become available.

A third central question follows directly from that problem and concerns the *interpretation of segregation under timescale and regime tension.* The literature

on young clusters makes clear that central concentration of massive or binary populations is not self-interpreting. It may reflect primordial or near-primordial inheritance, rapid local processing in dense substructure, hierarchical assembly effects, later secular evolution, or some combination of these channels. For NGC 6383, this issue is especially important because any reported concentration must be read against the broader context of youth, incomplete relaxation, and the limited authority of simple global timescale arguments in a dynamically young regime. The scientific issue is therefore not whether one explanatory label can be imposed in advance, but how any segregation signal should be interpreted when canonical relaxation-based reasoning may be informative yet insufficient, and when multiple physically plausible pathways remain open. In this thesis, segregation is treated not as a pre-classified result, but as an interpretive problem whose causal reading requires restraint.

These three central questions define the main scientific core of the dissertation. They are complemented by a second layer of *supporting questions* that remain important but do not govern the section at the same level. One concerns the apparent spread of pre-main-sequence ages in NGC 6383: whether the broadened chronology suggested by the PMS population should be read as evidence of extended or multi-episode star formation, as the product of extinction, accretion, variability, unresolved multiplicity, and model dependence, or as some combination of physical history and inferential degeneracy. A second concerns the physical meaning of the outer structure and extended populations of the cluster, since the literature does not justify a simple equivalence between observational extent, fitted structural scale, and a physically meaningful boundary. A third concerns what NGC 6383, as a difficult Gaia-era testbed, can clarify about inference-conditioned physical interpretation in young open clusters more broadly, without treating a single object as a definitive arbiter of wider astrophysical debates. These questions

remain scientifically substantive, but they are supporting rather than central, and their role is to refine the interpretive agenda rather than redefine it.

The next chapter provides the physical formalism needed to treat these questions with greater rigour. In particular, the issues raised here require a more careful framework for the interpretation of dynamical state in young non-equilibrium systems, for the use and limits of relaxation-based reasoning, for the reading of segregation under multiple plausible regimes, and for the physical meaning of structural scales, outskirts, and boundary language. Chapter 2 therefore supplies not a results preview, but the conceptual and formal groundwork required before the scientific questions posed here can be addressed responsibly. Section 1.6 thus marks the point at which the conceptual architecture of Chapter 1 is converted into a controlled scientific agenda.

Chapter 2

Physical formalism of stellar systems

The conceptual framework of Chapter 1 identified dynamical youth, incomplete relaxation, and non-equilibrium caution as recurring themes in the interpretation of young open clusters. The present chapter develops the physical formalism required to give those themes quantitative content. We derive the virial theorem in the form relevant to stellar dynamics, establish the conditions under which virial equilibrium may or may not be assumed, and identify the boundary, environmental, and non-equilibrium terms that complicate its application to young systems embedded in active star-forming environments (Section 2.1). The granular two-body relaxation process is formalised through the half-mass relaxation time and the mass-dependent segregation timescale, together with the conditions under which simple timescale arguments break down (Section 2.2). Section 2.3 reviews the structural models—King profiles, alternative lowered-isothermal families, and the problem of extended envelopes and potential escapers—that will be fitted to NGC 6383 in Chapter 6 and interpreted in Chapter 7. The Galactic tidal field and its role in defining cluster boundaries, regulating escape, and coupling internal

dynamics to the external environment are treated in Section 2.4. Each formalism is presented with its assumptions stated explicitly, so that the domain of validity of later physical claims can be traced to the formal tools on which they depend.

2.1 The virial theorem in stellar dynamics

2.1.1 Formal basis and the conditions for virial equilibrium

The virial theorem enters stellar dynamics as a moment relation of the collisionless Boltzmann equation rather than as an empirical scaling law. In the standard derivation, the Jeans equations are obtained from velocity moments of the distribution function and integrated over the system to yield the tensor virial theorem; the scalar form follows only after taking the trace and therefore discarding directional information about geometry and the velocity field (Jeans, 1915; Binney and Tremaine, 2008; Chandrasekhar, 1970; Heggie and Hut, 2012). This distinction matters because the scalar theorem summarizes the global energy budget, but cannot by itself distinguish isotropic pressure support from ordered rotation or anisotropic stress (Lee and Hennebelle, 2016).

The usual equilibrium statement is correspondingly narrower than is often assumed. For an isolated stellar system in strict dynamical equilibrium, the second time derivative of the moment of inertia vanishes and the stationary virial relation follows (Spitzer, 2014; Heggie and Hut, 2012). But a vanishing long-time average of that term is not equivalent to instantaneous equilibrium (Chandrasekhar, 1970). In cluster applications this matters because the theorem remains formally valid even while collisional evolution, structural readjustment, or bulk expansion is still ongoing (Spitzer, 2014; Binney and Tremaine, 2008). An instantaneous passage through $Q \simeq 1/2$ is therefore not, by itself, evidence that the system is

dynamically settled: collapsing or rebounding clusters can cross that value while their large-scale structure is still evolving (Parker et al., 2014; Allison et al., 2009; Smith et al., 2011).

Boundary and environmental terms make the same point more sharply. A finite integration volume introduces a surface-pressure term, while an external field enters through the tidal tensor and may either deepen or weaken the effective binding depending on geometry (McKee and Zweibel, 1992; Renaud, Gieles, and Boily, 2011). In embedded or pressurized environments, apparent excess kinetic support does not necessarily imply that the stellar system is unbound if boundary pressure contributes materially to the virial balance (Ballesteros-Paredes, 2006; Dib, Kim, and Shadmehri, 2007; Adamo et al., 2020). Apparent super-viriality is therefore ambiguous unless the surface and external terms can be shown to be dynamically negligible in the regime under discussion (McKee and Zweibel, 1992; Dib, Kim, and Shadmehri, 2007; Renaud, Gieles, and Boily, 2011).

The scalar virial theorem for a self-gravitating stellar system in the absence of external fields and surface pressure takes the form

$$\frac{1}{2}\,\ddot{I} \;=\; 2T + W\,, \tag{2.1}$$

where I is the scalar moment of inertia, T the total kinetic energy, and W the gravitational potential energy. In strict dynamical equilibrium, $\ddot{I} = 0$ and the familiar stationary relation $2T + W = 0$ follows. The virial ratio

$$Q \;\equiv\; \frac{T}{|W|} \tag{2.2}$$

then equals $1/2$ in equilibrium; $Q > 1/2$ implies net expansion or super-virial kinematics, while $Q < 1/2$ implies contraction (Portegies Zwart, McMillan, and Gieles, 2010; Gieles and Portegies Zwart, 2011). When an external tidal field or a

boundary pressure term contributes, Eq. (2.1) acquires additional terms that may either deepen or weaken the effective binding (McKee and Zweibel, 1992; Renaud, Gieles, and Boily, 2011).

The dynamical timescale against which equilibrium readings must be assessed is the crossing time,

$$t_{\rm cr} \;=\; \frac{2\,r_{\rm hm}}{\sigma_v}\,, \tag{2.3}$$

where $r_{\rm hm}$ is the half-mass radius and σ_v the three-dimensional velocity dispersion. A system younger than a few crossing times has not generally had enough time to erase large-scale structural perturbations, so a near-equilibrium virial ratio does not secure a settled dynamical interpretation (Gieles and Portegies Zwart, 2011; Krumholz, McKee, and Bland-Hawthorn, 2019).

The theorem is thus best treated as a formal constraint on the global energy budget, not as a self-sufficient classification tool. In young or structurally complex clusters, the relevant interpretive question is not whether the scalar relation can be written down, but whether it is being applied to the appropriate dynamical components, over an appropriate boundary, and on a timescale long enough for equilibrium-style reasoning to be meaningful.

2.1.2 Virial ratio, crossing times, and the meaning of "virialized"

In observational and simulation work, cluster dynamical state is often compressed into the virial ratio, which measures kinetic support relative to gravitational binding (Portegies Zwart, McMillan, and Gieles, 2010; Gieles and Portegies Zwart, 2011). In the idealised limit of an isolated steady-state system, the canonical equilibrium value follows from the scalar theorem (Heggie and Hut, 2012). Ratios above that value are commonly labelled super-virial and associated with expansion,

whereas lower values are associated with collapse; if the kinetic term becomes sufficiently dominant, the total energy is positive and the system is unbound (Bastian, Covey, and Meyer, 2010; Portegies Zwart, McMillan, and Gieles, 2010; Vesperini et al., 2021). This language is useful only if it is not mistaken for a diagnosis of relaxation.

The central interpretive error is to treat an instantaneous virial ratio near the equilibrium value as evidence that the system is virialized. That equivalence does not hold. A cluster can pass through $Q = 1/2$ during collapse, rebound, or violent relaxation while still being far from steady state (Krumholz, McKee, and Bland-Hawthorn, 2019; Parker et al., 2014; Allison et al., 2009; Smith et al., 2011). The asymmetry is simple: a stationary virialized state implies a specific energetic relation, but the converse does not follow from a single snapshot. The virial ratio is an instantaneous scalar descriptor; virialization is a statement about the damping of large-scale structural accelerations.

The physically relevant timescale is therefore the crossing time. Before any equilibrium-style reading is invoked, the cluster age must be compared to the timescale on which stars communicate dynamically across the system (Gieles and Portegies Zwart, 2011). Systems younger than only a few crossing times have not generally had enough time to erase large-scale perturbations, so a non-zero structural acceleration is expected rather than pathological (Portegies Zwart, McMillan, and Gieles, 2010; Bastian, Covey, and Meyer, 2010; Krumholz, McKee, and Bland-Hawthorn, 2019). This is especially important for young clusters, whose ages may be comparable to their dominant internal mixing timescales. In that regime, a near-equilibrium virial ratio does not secure a settled dynamical interpretation.

Non-equilibrium evolution further complicates the mapping between intrinsic state and observables. Hierarchical assembly, escaping stars, and incomplete truncation

can all bias the measured virial ratio, while background potentials, gas, or tides can alter the structural coefficient relevant to virial inference (Farias et al., 2015; Gavagnin et al., 2017; Parker and Goodwin, 2015). The important point is not to catalogue every such mechanism, but to note that virial-ratio interpretation is conditional on the boundary, tracer population, and physical regime being modelled consistently.

The virial ratio should therefore be treated as a useful instantaneous diagnostic, not as a sufficient certificate of dynamical settledness. Any claim that a cluster is "virialized" requires more than $Q \approx 1/2$: it requires evidence that the system is older than its relevant mixing timescale and that the measured kinematics are not dominated by collapse, rebound, escaping stars, or unmodeled environmental forcing (Gieles and Portegies Zwart, 2011; Krumholz, McKee, and Bland-Hawthorn, 2019; Parker and Goodwin, 2015).

2.1.3 Non-equilibrium regimes: gas, substructure, feedback, and tides

The strongest departures from ideal virial interpretation occur during cluster formation, when the potential is not stellar-only. In the embedded phase, stellar motions respond to the combined star–gas potential rather than to stellar self-gravity alone (McKee and Zweibel, 1992; Dib, Kim, and Shadmehri, 2007; Baumgardt and Kroupa, 2007). Stellar-only virial estimates can therefore exaggerate the bound stellar mass or falsely label an embedded system as super-virial when it remains bound within the full multi-component potential (Dib, Kim, and Shadmehri, 2007; Pang et al., 2020).

Gas removal introduces a second non-equilibrium transition. In the classical impulsive picture, rapid gas loss changes the potential before stellar velocities

can adjust, leading naturally to expansion or partial dissolution at low effective star-formation efficiency (Hills, 1980; Goodwin and Bastian, 2006; Baumgardt and Kroupa, 2007). But later work shows that the post-expulsion response depends strongly on the pre-expulsion dynamical state, local stellar fraction, and substructure, not only on the global gas fraction (Goodwin, 2009; Smith et al., 2011; Pang et al., 2020). The result is that gas removal cannot be interpreted through a single threshold logic alone.

Hierarchical assembly makes the same point from a different direction. Young clusters may form as clumpy or fractal systems in which local force balance inside substructures coexists with global non-equilibrium (Parker et al., 2014; Allison et al., 2009; Smith et al., 2011; Krause et al., 2020). A global velocity dispersion can then mix organized infall with internal random motion, while early concentration patterns may reflect local processing in dense sub-clumps rather than relaxation across the scale of the final assembled cluster (Parker and Goodwin, 2015; Lee and Hennebelle, 2016). In such cases, local dynamical maturity and global equilibrium are not the same observable.

Feedback and tides add further time dependence. Feedback-regulated gas removal, transient tidal forcing, and escaping stars can all perturb the kinetic budget in ways not captured by a stationary stellar-only virial picture (Grudić et al., 2021; Guszejnov et al., 2022b; Claydon, Gieles, and Zocchi, 2017). The exact mechanism varies with regime, but the interpretive consequence is stable: apparent virial balance in a young cluster is conditional on what physical components and boundaries have been included.

Non-equilibrium in clusters is not a single phenomenon: it can arise from gas-dominated potentials, hierarchical assembly, feedback-shaped emergence, or external tides. Virial language in young clusters is therefore regime-dependent, and its application to NGC 6383 in Chapter 7 will be conditioned accordingly.

2.1.4 Observational mass estimators, biases, and inference-level remedies

Observational cluster dynamics often inverts the virial theorem into a mass estimator built from a projected size, a line-of-sight velocity dispersion, and a structural coefficient calibrated for an assumed density profile and orbital structure (Spitzer, 2014; Heggie and Hut, 2012; Sollima et al., 2015). The attraction of this estimator is obvious, but so are its assumptions: spherical symmetry, near-stationarity, isotropy, and a tracer population that samples the same dynamical structure as the underlying mass.

Several failure modes follow directly. Velocity anisotropy biases isotropic virial inversion (Wolf et al., 2010; Fleck et al., 2006), unresolved binaries can dominate the apparent dispersion in dynamically cold systems (Kouwenhoven and Grijs, 2007; Cottaar, Meyer, and Parker, 2012), and tracer selection can shift the inferred mass if the observed population is not dynamically representative of the underlying mass distribution (Bovy, 2015; Gieles et al., 2010; Webb et al., 2014). Mass segregation is particularly relevant because luminous centrally concentrated tracers can produce an underestimate of the total mass, whereas binaries can inflate the measured dispersion and bias the estimate upward. The sign of the error is therefore population-dependent, not merely statistical.

Membership uncertainty and spatial incompleteness add another layer. Crowding, field contamination, and incomplete sampling of low-mass or peripheral stars can all bias virial inference if they are treated as upstream cleaning problems rather than as part of the inference itself (Kamann et al., 2018; Tarricq et al., 2021; Galli et al., 2021; Wright et al., 2024). For that reason, recent work increasingly treats membership uncertainty, structure, and kinematics jointly within discrete-likelihood or hierarchical frameworks (Hénault-Brunet et al., 2019; Eadie, Webb,

and Rosenthal, 2022; Wen et al., 2023). These approaches do not remove model dependence, but they make the assumptions explicit and handle several classical failure modes more transparently.

Observational virial estimates are therefore not direct measurements of dynamical mass but assumption-heavy inferences whose reliability depends on equilibrium status, tracer representativeness, binaries, anisotropy, and membership.

2.2 Two-body relaxation

2.2.1 Half-mass relaxation time and the granular basis of collisional evolution

The distinction between collisionless and collisional stellar dynamics rests on whether discrete gravitational encounters perturb stellar orbits on an astrophysically relevant timescale. In real clusters, cumulative small-angle deflections drive diffusion in velocity space, so two-body relaxation is controlled not by rare direct collisions but by the integrated effect of many weak encounters (Trenti et al., 2007; Spitzer, 2014). The relevant question is therefore whether the accumulated variance in the velocity perturbations becomes comparable to the original squared velocity of a star.

This diffusion picture leads to the standard relaxation-time estimates, in which the Coulomb logarithm absorbs the range of dynamically effective impact parameters (Trenti et al., 2007; Giersz and Heggie, 1994). That term is physically necessary but not uniquely fixed: both the lower and upper cutoffs depend on modelling choices, and calibrated comparisons show that effective values differ between equal-mass and unequal-mass systems (Giersz et al., 2013; Gieles et al., 2010). The normalization of the relaxation clock is therefore model-dependent even before

observational uncertainties are introduced.

The half-mass radius enters because local core timescales are unstable under secular evolution, whereas the half-mass scale remains a more useful tracer of the bulk dynamical state (Spitzer, 2014; Trenti et al., 2007). The corresponding half-mass relaxation time is thus the standard global diagnostic. But it should not be mistaken for a completed dynamical diagnosis. It is a heuristic clock that links collisional evolution to the system size, mass, and particle number under specific structural assumptions.

Formally, the half-mass relaxation time for a system of N equal-mass stars is

$$t_{\rm rh} \;=\; \frac{0.17\,N}{\ln(\lambda N)}\sqrt{\frac{r_{\rm hm}^3}{G\,M}}\,, \tag{2.4}$$

where M is the total cluster mass and $\lambda \approx 0.11$ a calibration constant that absorbs the Coulomb logarithm (Binney and Tremaine, 2008; Giersz and Heggie, 1994). In a multi-mass system, the segregation timescale for a star of mass m is

$$t_{\rm seg} \;=\; \frac{\langle m\rangle}{m}\,t_{\rm rh}\,, \tag{2.5}$$

where $\langle m\rangle$ is the mean stellar mass (Spitzer, 2014). This classical scaling implies that the most massive stars segregate first, but it presupposes that full equipartition is dynamically attainable—a condition examined in Section 2.2.2.

That caution matters because the relaxation clock is itself entangled with structural evolution and escape physics. Expansion can lengthen the relaxation time, and in tidally limited systems the lifetime does not scale linearly with t_{rh} because stars that exceed the escape threshold are not removed instantaneously (Baumgardt, 2001; Baumgardt and Makino, 2003; Fukushige and Heggie, 2000). Potential escapers can remain dynamically coupled to the cluster for finite times, so even at the level of idealised theory, relaxation time and dissolution time are not

interchangeable.

Observationally, the inferred half-mass relaxation time is further biased if half-light radii are substituted for half-mass radii, if the low-mass population is incomplete, or if binaries distort the inferred mass and velocity scales (Shanahan and Gieles, 2015; Sollima et al., 2015; Sollima, Baumgardt, and Hilker, 2019). The conclusion required for this thesis is therefore limited but important: the half-mass relaxation time remains the most useful global collisional timescale, but it is a model-dependent heuristic clock, not a proof that the system has already attained the dynamical condition that the clock nominally measures.

2.2.2 Segregation timescales, partial equipartition, and regime sensitivity

In multi-mass systems, two-body relaxation drives differential orbital evolution through dynamical friction and energy exchange. In the classical picture, massive stars lose kinetic energy, drift inward, and segregate on a timescale shorter than the global relaxation time by a factor set by the stellar mass ratio (Spitzer, 2014; Trenti et al., 2007). This gives the familiar expectation that the most massive objects should concentrate first in the deepest part of the potential well.

That classical scaling is useful but incomplete because it assumes that full equipartition is dynamically attainable. Modern work instead shows that realistic clusters generally approach only partial equipartition, with the inner regions relaxing more efficiently than the halo (Bianchini et al., 2016b; Gieles and Zocchi, 2015; Spera, Mapelli, and Jeffries, 2016). Partial equipartition is therefore not merely incomplete progress toward a classical endpoint, but evidence that the fully equipartitioned endpoint is itself generally not the dynamically realized state. The consequence is interpretive rather than semantic: a segregation timescale derived

from classical equipartition logic should be treated as a heuristic guide, not as a guaranteed achieved state. Classical mass-scaled segregation times therefore indicate at best an ordering tendency in collisional transport, not a self-certifying causal clock. Segregation may be dynamically informative without implying that the system has fully relaxed in the classical thermodynamic sense.

This limitation becomes more severe in young or structurally complex systems. Massive stars can appear centrally concentrated at ages shorter than the global relaxation time because early concentration may arise from local processing in dense substructure, hierarchical assembly, or other non-global channels (Portegies Zwart, McMillan, and Gieles, 2010; Lee and Hennebelle, 2016; Parker et al., 2014; Parker and Goodwin, 2015). In such regimes, the relevant contrast is not simply "primordial" versus "dynamical," but global relaxation versus regime-dependent early transport. A single global t_{rh} cannot by itself decide among these possibilities.

For the present thesis, that is the crucial point. Segregation is not self-interpreting. Relaxation-based timescales remain informative, but their explanatory authority depends on whether the cluster is close enough to the regime in which global collisional reasoning is a good leading-order description. In young, substructured, or only partially mixed systems, central concentration must be interpreted cautiously and without premature causal closure.

2.2.3 Inference, calibration, and breakdown regimes

The translation from relaxation theory to observed cluster dynamics is unavoidably inferential. Structural radii, particle numbers, binary content, and membership all enter the relaxation estimate, so the corresponding timescale inherits their uncertainties and biases (Vasiliev and Baumgardt, 2021; Krone-Martins and

Moitinho, 2014; Dalessandro et al., 2015; Sollima, Baumgardt, and Hilker, 2019). Probabilistic membership and latent-variable approaches are therefore preferable to hard-cut pipelines when the structural and kinematic posteriors are coupled.

Numerical calibration sharpens but does not remove this point. Comparisons between direct N-body and Monte Carlo methods show that classical formulae often require recalibration rather than literal use, especially in unequal-mass systems (Giersz et al., 2013). The value of those calibrations is not that they deliver a universal constant, but that they make explicit how model assumptions enter the relaxation clock and whether that clock is being applied within its leading-order dynamical domain of validity.

There are also physically important regimes in which ordinary two-body relaxation is not the sole or dominant control variable. Rotation, anisotropy, strong tidal forcing, and other collective effects can alter segregation or mass-loss pathways (Tiongco, Vesperini, and Varri, 2017; Livernois et al., 2021; Gieles and Renaud, 2016; Webb, Reina-Campos, and Kruijssen, 2019). For the purposes of this thesis, however, the detailed taxonomy of such breakdowns is less important than the general doctrine they establish: relaxation-based reasoning remains useful only after a regime audit. Once the cluster departs too far from the weakly rotating, weakly anisotropic, quasi-equilibrium picture in which the classical timescale was derived, the timescale becomes interpretive support rather than dynamical closure.

Two-body relaxation remains the backbone of collisional stellar dynamics, and the half-mass relaxation time its most useful global clock. But relaxation time is not equivalent to achieved relaxation, and segregation timescales are conditional rather than self-certifying. For NGC 6383, with $\tau/t_{\rm rh} \approx 0.19$ (Chapter 7), this distinction is not academic: the cluster has completed only a fraction of its first relaxation cycle, so any apparent segregation must be read against both primordial and early-transport channels.

2.3 Structural models

2.3.1 From projected profiles to dynamical structure

Structural modelling connects the observed distribution of stars on the plane of the sky to an inferred three-dimensional stellar system. The central distinction is therefore not between fitting functions as such, but between the projected surface-density profile that is measured and the underlying structure that is inferred (Binney and Tremaine, 2008; Heggie and Hut, 2012). Because the observable is a line-of-sight projection, structural inference is an inverse problem rather than a purely descriptive exercise.

A second distinction separates empirical profile laws from models tied to explicit distribution functions. Empirical laws can fit projected data flexibly, but they do not guarantee that the implied three-dimensional density and kinematic structure correspond to a dynamically admissible equilibrium configuration (Baumgardt and Hilker, 2018; Gieles and Zocchi, 2015). DF-based models are more restrictive, but that restriction is their physical strength: they connect density and kinematics within a single framework (King, 1966; Gieles and Zocchi, 2015). The contrast is therefore not simplicity versus sophistication, but descriptive adequacy in projection versus dynamical self-consistency.

This distinction becomes especially important in the outskirts. A fitted outer cutoff may be interpreted as a tidal boundary, yet observationally it may also mark only the radius at which the cluster profile becomes indistinguishable from the field (Baumgardt and Hilker, 2018; Sollima et al., 2015). Those interpretations are not equivalent. A physical boundary constrains escape and survival arguments; an observational cutoff may reflect background contrast or model misspecification. Structural radii therefore do not interpret themselves.

Modern inference makes this even clearer by showing that outer-profile parameters are often entangled with the contamination model. Background misestimation can shift the fitted truncation scale and thereby propagate directly into later mass, relaxation, and boundary interpretations (Eadie, Webb, and Rosenthal, 2022; Sollima et al., 2015). Structural modelling is therefore a major inferential layer rather than a neutral descriptive prelude. Before one interprets a core radius, outer scale, or concentration, one must ask what physical assumptions made those quantities meaningful.

2.3.2 King profiles and the logic of lowered isothermal models

King-type models remain historically important because they encode the idea of a finite cluster extent. The empirical law of King (1962) provided a practical profile with a core scale and an outer truncation, while King (1966) introduced a self-consistent lowered-isothermal distribution-function framework in which density, potential, and kinematics are linked (King, 1962; King, 1966). These two formulations are related but not interchangeable: one is an observationally convenient projected law, the other a dynamical equilibrium model (Baumgardt and Hilker, 2018; Gieles and Zocchi, 2015).

This distinction matters because a fitted King-like profile may summarize projected structure adequately without justifying a strong dynamical interpretation. Classical lowered-isothermal models assume approximate equilibrium, spherical symmetry, and a particular truncation law at the escape boundary. Under those conditions, they remain useful summaries of finite, quasi-relaxed stellar systems (Gieles and Zocchi, 2015; Baumgardt and Hilker, 2018). But when the system is young, anisotropic, weakly filling, or surrounded by a more extended outer

population, the same formalism becomes interpretively fragile.

The empirical projected density law of King (1962) takes the form

$$\Sigma(r) \;=\; k\left[\frac{1}{\sqrt{1+(r/R_c)^2}}-\frac{1}{\sqrt{1+(R_t/R_c)^2}}\right]^2+b\,, \tag{2.6}$$

where R_c is the core radius, R_t the tidal (limiting) radius, k a central surface-density normalisation, and b the background level. The concentration parameter

$$C \;\equiv\; \log_{10}\!\left(\frac{R_t}{R_c}\right) \tag{2.7}$$

encodes the ratio of outer to inner scale and serves as a compact structural descriptor (Peterson and King, 1975). Equation (2.6) is empirically convenient and widely used, but it is not derived from a distribution function; the self-consistent lowered-isothermal models of King (1966) are related but not interchangeable (Gieles and Zocchi, 2015).

Generalized lowered-isothermal families make this point explicit by relaxing the rigidity of the classical cutoff and, in some cases, allowing anisotropy to vary separately from truncation sharpness (Gieles and Zocchi, 2015; Zocchi et al., 2016; Breen, Varri, and Heggie, 2017). For the present thesis, however, the essential lesson is simpler: King-like models are most useful when the goal is to connect projected structure to a finite equilibrium-like configuration. They become less reliable when the cluster outskirts are dynamically mixed, incompletely relaxed, or only weakly represented by a sharp classical truncation. The minimal contrasts retained here matter only because different model logics imply different physical readings of whether the outskirts should be treated as truncated, softened, or extended, not because the thesis requires a survey of model families.

2.3.3 Alternative profile families and the problem of extended envelopes

Not all clusters are well described by a flat core plus sharp cutoff. Some model families are better understood not as entries in a taxonomy, but as examples of different assumptions about the outskirts.

The Plummer model remains analytically convenient and often pedagogically useful, but its infinite extent and fixed halo shape limit its direct interpretive authority in real cluster outskirts (Plummer, 1911; Binney and Tremaine, 2008). EFF-type profiles, by contrast, are useful descriptive models for young systems with extended power-law envelopes and no obvious tidal edge (Elson, Fall, and Freeman, 1987; Cuevas-Otahola et al., 2020). Wilson-like models occupy an intermediate position: they retain an equilibrium-style framework but allow a softer outer taper than classical King models (Wilson, 1975; Barmby et al., 2007; Gieles and Zocchi, 2015).

The important point is not to map a full ecology of profile families, but to recognise the inferential distinction they encode. A model with a sharp truncation, a softened equilibrium halo, or an untruncated power-law envelope is not merely fitting the data differently; it is assigning different meaning to the outer structure. In that sense, the choice among these families is physically consequential because different fitted outer scales answer different questions about structure, truncation, and boundary.

2.3.4 Outer boundaries, potential escapers, and the breakdown of classical truncation

The outer envelope of a cluster is not populated solely by stars that are cleanly bound in the classical lowered-isothermal sense. Potential escapers are energetically unbound stars that can remain trapped inside or near the Jacobi surface for extended periods because escape through the tidal boundary is dynamically constrained (Küpper et al., 2010; Claydon, Gieles, and Zocchi, 2017; Boer et al., 2019). Their presence weakens the classical assumption that the phase-space density must vanish sharply at the escape energy.

This matters because outer-halo structure then ceases to map cleanly onto a classical tidal cutoff. Potential escapers can flatten outer kinematic profiles, populate mixed outskirts, and generate fitted boundaries that differ depending on whether the model allows a non-zero population at the escape threshold (Claydon, Gieles, and Zocchi, 2017; Boer et al., 2019; Sollima et al., 2023). The disagreement among fitted outer scales is therefore physical, not merely parametric: different models treat the meaning of the escape boundary differently.

Contamination and tidal filling amplify the same ambiguity. In strongly filling systems, extra-tidal material can elevate the apparent background and bias one-dimensional radial fits, while the transition between cluster outskirts and field population may require wider spatial domains than a compact classical fit would imply (Balbinot and Gieles, 2018; Wan et al., 2022). The practical conclusion is that fitted outer scale and physical boundary are not equivalent. The outskirts are a mixed dynamical zone in which contamination, escape physics, and structural modelling are inseparable.

2.3.5 Multimass structure, anisotropy, and the limits of simple fits

Single-mass spherical models assume that the observed tracer population is an adequate proxy for the total mass distribution. In segregated systems, that approximation can fail because bright stars, low-mass stars, and dark remnants occupy different spatial distributions (Hénault-Brunet et al., 2019; Dickson et al., 2023). The resulting error is not cosmetic: if the light does not trace the mass, then inferred half-mass radii, central densities, and relaxation times can all be biased.

Multimass models address this problem by allowing distinct mass components to share a common potential while occupying different spatial distributions (Dickson et al., 2023; Dickson et al., 2024). The thesis does not require a full survey of that literature, but it does require the main doctrinal point: half-light and half-mass structure are not automatically interchangeable, and structural inferences based on a single effective tracer can be physically misleading in segregated systems.

A related caution applies to anisotropy and non-sphericity. Rotation and anisotropy need not leave the same signature, and flattening cannot be attributed to a single mechanism without an appropriately matched model (Jeffreson et al., 2017; Bianchini et al., 2018; Martens et al., 2023). For the later empirical chapters, however, the needed lesson is modest: deviations from simple spherical isotropic structure are another reason why fitted structural parameters should be interpreted as model-conditioned rather than as direct physical observables.

2.3.6 Inference strategy, identifiability, and model choice

The statistical problem of structural modelling cannot be separated from the physical one. Binned radial fitting compresses point-process information and can

obscure covariances between centre position, background level, and outer scale, whereas unbinned and hierarchical approaches retain more of the information actually present in the data (Eadie and Jurić, 2019; Eadie, Webb, and Rosenthal, 2022; Wen et al., 2023). Yet improved inference does not eliminate misspecification. It only makes the assumptions more explicit.

The main identifiability problem is that truncation scale and contamination level are often covariant, especially in shallow or low-contrast data (Eadie, Webb, and Rosenthal, 2022; Vasiliev, 2019). Field-star decontamination is therefore not a cosmetic preprocessing step but part of the structural inference itself. This is especially relevant when the cluster boundary is not sharply defined.

Model choice is therefore an assumption audit rather than a competition among interchangeable fitting functions. We adopt a King-type profile for NGC 6383 (Chapter 6), which is appropriate insofar as the system is approximately finite and admits an equilibrium-like inner structure, but the fitted outer scale R_t must be interpreted with the caveats established in Sections 2.3.4 and 2.4.

2.4 Galactic tidal effects

2.4.1 Jacobi radius, tidal geometry, and the meaning of a cluster boundary

The first difficulty in discussing galactic tidal effects is definitional. The empirical limiting radius introduced in profile fitting is not the same object as the Jacobi radius derived from the rotating-frame dynamics of a cluster embedded in an external field (King, 1962; Hoerner, 1957; Binney and Tremaine, 2008). The former is a fitted structural cutoff in projected density; the latter is a dynamical scale associated with the topology of the effective potential. Treating them as

interchangeable risks turning a fitted profile parameter into a physical boundary by assumption.

Even the Jacobi radius is only straightforward under restrictive conditions. The classical derivation assumes a circular orbit, a time-independent axisymmetric host potential, and a point-like cluster mass (Spitzer, 2014; Heggie and Hut, 2012). More general formulations embed the boundary in the local tidal tensor and therefore make it explicitly dependent on the local galactic geometry and orbital state (Just et al., 2009; Renaud, Gieles, and Boily, 2011). The Jacobi scale is therefore best interpreted as an instantaneous, model-dependent tidal boundary, not as a permanently fixed edge.

This matters because outer envelopes need not remain fixed at a single orbital phase. Time-dependent studies show that the response of the outskirts can vary along eccentric or otherwise non-stationary orbits, and that stars displaced beyond an instantaneous boundary need not be lost irreversibly in the simplest static sense (Baumgardt and Makino, 2003; Webb et al., 2014; Renaud, Gieles, and Boily, 2011). The main doctrinal lesson is that different radii encode different physical questions: fitted structural extent, instantaneous tidal balance, and effective outer material need not coincide.

For a cluster of mass m_c on a circular orbit at galactocentric distance $R_{\rm gc}$ within a host galaxy of enclosed mass $M_{\rm gc}(R_{\rm gc})$, the Jacobi (Hill) radius is

$$R_{\rm Hill} \;=\; R_{\rm gc}\left(\frac{m_c}{3\,M_{\rm gc}}\right)^{1/3}, \tag{2.8}$$

which defines the instantaneous tidal boundary under the stated symmetry and stationarity assumptions (Binney and Tremaine, 2008; Just et al., 2009). An

alternative estimator, the gravitationally bound radius

$$r_{\rm bound} = \left(\frac{G\, m_c}{2\,(A-B)^2} \right)^{1/3}, \tag{2.9}$$

where A and B are the Oort constants, provides a complementary tidal scale tied to local differential rotation (Zwart et al., 1998). Neither quantity coincides automatically with the fitted structural truncation R_t of Eq. (2.6).

For the purposes of this thesis, that distinction is enough. The physical meaning of a cluster boundary cannot be read directly from a single radius estimate, and any later interpretation of outer structure must keep separate the empirical limiting scale and the dynamical Jacobi scale.

2.4.2 Bound fraction, potential escapers, and delayed escape

The bound fraction is often treated as a simple measure of cluster survival, but its meaning depends on both epoch and criterion. The fraction retained immediately after gas expulsion is not the same physical quantity as the fraction remaining after long-term relaxation-driven escape in a tidal field (Smith et al., 2011; Farias et al., 2015). These regimes differ not only in timescale but in controlling dynamics: the former concerns violent post-expulsion readjustment, whereas the latter concerns delayed escape and secular depletion in an external tidal field. Merging these regimes into a single survival number obscures the fact that early-time violent readjustment and late-time secular depletion are dynamically distinct processes.

In the secular regime, the crucial correction is that energetic unbinding does not imply immediate escape. Clusters host potential escapers: stars whose energies exceed the nominal threshold but which remain trapped inside or near the Jacobi surface for extended periods (Küpper et al., 2010; Daniel, Heggie, and Varri, 2017; Ernst et al., 2007; Zotos, 2014). Escape is therefore controlled not only by the

energy criterion, but also by access to the relevant phase-space channels through the Lagrange-point bottlenecks. Energetic unbinding is necessary for escape, but not sufficient.

This delayed-escape picture has direct consequences for interpretation. A simple spatial cut does not cleanly separate bound from unbound populations, and outer-halo material may still contribute to the observable density and velocity field even after exceeding the nominal escape energy (Daniel, Heggie, and Varri, 2017). Bound fraction is therefore not a direct observable but an inference conditioned by the adopted energetic criterion, the escape-time physics, and the internal or external heating processes included in the model.

The same ambiguity appears observationally in extra-tidal populations and tidal tails. Gaia-era studies have made such structures much easier to detect, but not necessarily easier to interpret, because tail members become progressively harder to distinguish from the field and because delayed escape blurs the transition between cluster outskirts and detached debris (Meingast, Alves, and Rottensteiner, 2021; Bianchini et al., 2016a; Risbud, Jadhav, and Kroupa, 2025; Tarricq et al., 2022). The relevant conclusion is therefore restrained: outer populations cannot be read straightforwardly as either securely bound or cleanly escaped without a model of delayed escape and observational selection.

2.4.3 Environmental dependence and observational ambiguity

Galactic tidal effects are conditioned by environment, but the present thesis does not require a broad survey of every environmental regime. What is needed is the narrower point that tidal forcing can be time-dependent, orbit-dependent, and observationally difficult to read. Repeated shocks, varying orbital phase,

and host-galaxy structure can all alter the effective tidal field experienced by the cluster (Gnedin and Ostriker, 1997; Gieles and Renaud, 2016; Martinez-Medina et al., 2018). The practical consequence is that a static tidal picture is often only an approximation.

This time dependence matters because outer structures such as halos or tails are not shaped solely by internal escape physics. They also depend on the external forcing history and on how that history couples to the internal response time of the cluster (Kruijssen et al., 2011; Gieles and Renaud, 2016). For later interpretation, however, the essential message is not cosmological breadth but observational caution: outer asymmetries and extended populations are selection-sensitive and environment-sensitive, so they should not be overread as uniquely diagnostic of a single dynamical mechanism.

The observational literature reinforces this caution. Gaia has revealed a wide diversity of extra-tidal structures, but projection effects, incompleteness, and clustering bias can distort both the apparent morphology and the inferred dynamical state (Hunt and Reffert, 2024; Buckner et al., 2022). Tail-like features are therefore not self-interpreting, and apparent asymmetries should not be converted directly into strong physical claims without forward treatment of contamination, selection, and projection. Galactic tides therefore matter not only because they limit clusters dynamically, but because they complicate the meaning of boundaries, escape, and outer structure. Fitted extent, tidal radius, and outer populations are not interchangeable—a distinction that becomes concrete when interpreting the NGC 6383 structural profile in Section 7.3.

Chapter 3

Statistical philosophy and model assumptions

The physical formalism of Chapter 2 provides the equations; this chapter establishes the inferential framework that determines how those equations may be connected to observational data and what status the resulting claims may legitimately carry. We begin from the measurement model that separates latent astrophysical quantities from their noisy observational projections, and characterise the structured uncertainty architecture of Gaia DR3—heteroscedasticity, intra-source correlation, and spatially correlated systematics—that conditions every downstream inference (Section 3.1). Membership is then recast as a problem of probabilistic population inference rather than deterministic sample selection, formalising the sense in which the stellar population of NGC 6383 is not a fixed list but a posterior-dependent inferential object (Section 3.2). The sensitivity of posterior inference to prior specification is then developed as a general principle of inferential fragility, showing how apparently modest regularisation choices become substantive when the likelihood is weak or flat (Section 3.3). The domain of interpretive validity is defined next: the catalogue-conditioned population recov-

ered from Gaia DR3 is not the unrestricted physical cluster, and that restriction propagates into every downstream physical summary (Section 3.4). The chapter concludes by establishing the claim-status rule that will govern every physical result reported in Chapters 7–8: the requirement that each claim is stated together with the generative model, prior, and convergence status on which it is conditioned (Section 3.5).

3.1 Measurement model and inferential status of observables

3.1.1 Latent and observed quantities

The analysis of NGC 6383 does not begin from directly readable physical quantities but from catalogued observables whose inferential status must be established before any cluster-level claim can be made. In the standard measurement-error formulation, the observed datum for star i is not the underlying astrophysical state but a noisy realisation of an unobserved latent quantity (Fuller, 1987; Kelly, 2007). We write the generic measurement model as

$$\boldsymbol{y}_i \;=\; \boldsymbol{x}_i + \boldsymbol{\varepsilon}_i \,, \tag{3.1}$$

where $\boldsymbol{x}_i$ is the true (latent) astrometric–photometric state of the source and $\boldsymbol{\varepsilon}_i$ is the measurement error, drawn from a source-specific distribution with covariance $\boldsymbol{C}_i$ that may be non-diagonal and magnitude-dependent (Lindegren et al., 2021a; Hogg, Myers, and Bovy, 2010). For Gaia DR3, $\boldsymbol{y}_i$ includes the five-parameter astrometric solution $(\alpha, \delta, \varpi, \mu_{\alpha*}, \mu_\delta)$ together with broadband photometry $(G, G_{\rm BP}, G_{\rm RP})$; the catalogue is therefore not a direct tabulation of

physical stellar states but an observational projection of them through a measurement process whose error structure is itself part of the data (Gaia Collaboration et al., 2023; Riello et al., 2021).

This distinction is especially consequential for parallax, which enters physical interpretation only after a nonlinear transformation. Because $d = 1/\varpi$ is a convex function of the observable, uncertainty does not transfer symmetrically from parallax space into distance space, and a direct inversion of noisy parallaxes produces biased and asymmetric distance estimates (Luri et al., 2018; Bailer-Jones, 2015). The passage from $\varpi_i \pm \sigma_{\varpi,i}$ to a distance-relevant posterior therefore requires an explicit prior and a generative model rather than naive algebraic manipulation (Bailer-Jones et al., 2018; Bailer-Jones et al., 2021). This point is not confined to distance: any population-level summary computed from the observed $\{\boldsymbol{y}_i\}$ mixes the intrinsic population structure of $\{\boldsymbol{x}_i\}$ with the measurement process of Eq. (3.1).

3.1.2 Structured uncertainty in Gaia DR3

The measurement model of Eq. (3.1) is only useful if $\boldsymbol{C}_i$ is taken seriously. In Gaia DR3, three properties of the error structure condition every downstream inference and will be referenced throughout this thesis rather than restated in each section.

Heteroscedasticity. Gaia uncertainties vary substantially across sources as a function of magnitude, colour, and observing conditions, so stars do not contribute equally precise information (Gaia Collaboration et al., 2023; Fabricius et al., 2021). Bright sources encounter calibration floors, while faint sources are noise-dominated. The observed scatter in any astrometric or photometric coordinate therefore mixes instrumental degradation with genuine population-level variation,

and any attempt to read raw dispersion as physical structure risks attributing observational artefacts to the cluster (Carroll and Ostlie, 2006; Heyl, Caiazzo, and Richer, 2022).

Intra-source correlation. For individual stars, the five astrometric parameters are reported with a non-diagonal covariance matrix because the scanning-law geometry and global astrometric solution couple the uncertainty on ϖ, $\mu_{\alpha*}$, and μ_δ (Lindegren et al., 2021a). When these correlations are ignored, the apparent geometry of the observed distribution in phase space can be distorted: error ellipses elongated along specific axes are misread as intrinsic anisotropy (Robotham and Obreschkow, 2015; Bovy, Hogg, and Roweis, 2011).

Spatially correlated systematics. Beyond the single-source level, Gaia astrometry exhibits shared systematic structure on the sky, including parallax zero-point behaviour that varies with magnitude, colour, and position, and proper-motion residuals correlated on angular scales comparable to cluster diameters (Lindegren et al., 2021a; Vasiliev, 2019; Zinn, 2021; Groenewegen, 2021). These systematics arise from survey-level calibration and attitude residuals rather than from independent stellar fluctuations. The practical consequence is that neighbouring sources are not statistically independent: a precision floor exists that does not decrease under simple averaging, and coherent local offsets can mimic ordered kinematic patterns such as expansion, rotation, or tidal structure (Vasiliev and Baumgardt, 2021; Maíz Apellániz, 2022; Hunt and Reffert, 2023).

These three properties—heteroscedasticity, intra-source correlation, and spatially correlated systematics—define the measurement structure that every later inferential step must respect. We refer to them collectively as the *Gaia uncertainty architecture*; when later sections invoke this term, the full content of the present subsection is implied.

3.1.3 Population-level inference requires forward modelling

In the presence of structured uncertainty, population-level quantities cannot be extracted by naive inversion of the catalogue. Cluster-level statements concern the distribution of the latent $\{\boldsymbol{x}_i\}$, which is accessed only through the observed $\{\boldsymbol{y}_i\}$ via Eq. (3.1). The observed marginal distribution is therefore a convolution,

$$p(\boldsymbol{y}) \;=\; \int p(\boldsymbol{y} \mid \boldsymbol{x})\, p(\boldsymbol{x} \mid \boldsymbol{\Theta})\, \mathrm{d}\boldsymbol{x}\,, \tag{3.2}$$

where $p(\boldsymbol{y} \mid \boldsymbol{x})$ encodes the measurement model and $p(\boldsymbol{x} \mid \boldsymbol{\Theta})$ is the latent population model parameterised by $\boldsymbol{\Theta}$ (Kelly, 2007; Bovy, Hogg, and Roweis, 2011). Recovering $\boldsymbol{\Theta}$ from $\{\boldsymbol{y}_i, \boldsymbol{C}_i\}$ is an inverse problem that requires an explicit forward model rather than a direct reading of the catalogue.

Observed scatter in parallax, proper motion, or photometry is therefore a mixture of two distinct inferential objects: the intrinsic dispersion of the latent population and the measurement noise specific to each source (Carroll and Ostlie, 2006; Fuller, 1987). When these contributions are not separated through a proper forward model, apparent population-level properties—internal velocity dispersion, parallax depth, CMD width—are conflated with instrumental broadening. For a young system such as NGC 6383, where unresolved multiplicity can inject additional non-Gaussian scatter into both astrometric and photometric observables (Holl et al., 2023; El-Badry, 2024; Penoyre et al., 2020), naive catalogue reading is therefore insufficient. The analysis must instead proceed through the generative framework formalised in Chapter 4.

3.2 Membership as probabilistic population inference

3.2.1 Fixed-list membership is inferentially insufficient

In cluster studies, membership selection is often treated as a preliminary cleaning step that merely removes obvious contaminants before the physical analysis begins. That picture is inferentially misleading. Any hard cut in parallax, proper motion, or photometric space does not reveal a pre-existing population; it actively defines the population that will later be analysed. A deterministic boundary replaces continuous uncertainty with a binary decision, discards the evidential role of ambiguous objects, and entangles astrophysical variation with source-specific measurement geometry under the Gaia uncertainty architecture of Section 3.1.2 (Loredo, 2004; Trotta, 2008; Cantat-Gaudin et al., 2020). Because $\boldsymbol{C}_i$ varies strongly across the sample, stars with intrinsically similar latent states $\boldsymbol{x}_i$ can occupy very different observed positions $\boldsymbol{y}_i$, so a fixed boundary in the observed space cuts through the latent population in a membership-biased way.

The same problem applies when posterior-threshold catalogues are used as if they were definitive member lists. Retaining only stars above a chosen membership probability softens the boundary but still collapses graded uncertainty into a binary decision. The resulting population is again a constructed object whose composition depends on the threshold rule, and information carried by ambiguous stars is discarded rather than propagated into later inference (Sanders, 1971; Krone-Martins and Moitinho, 2014; Cantat-Gaudin et al., 2020). Catalogue quality filters—such as cuts on RUWE—can further reshape the admitted population in astrophysically non-neutral ways, since problematic astrometric solutions often trace unresolved companions rather than purely instrumental failures (Penoyre

et al., 2020; El-Badry, 2024).

3.2.2 The mixture framework and latent membership

The underlying reason deterministic membership fails is that membership is not directly observed in the catalogue. What is observed are the noisy measurements $\boldsymbol{y}_i$ together with their uncertainties $\boldsymbol{C}_i$. Membership is an inferential statement about whether a star is better explained as belonging to the cluster-associated population than to the surrounding field under some adopted generative construction.

The standard probabilistic framework treats the observed sample as a mixture of two or more populations. We write the marginal likelihood for source i as

$$p(\boldsymbol{y}_i \mid \boldsymbol{\Theta}) \;=\; \pi_c \, p_c(\boldsymbol{y}_i \mid \boldsymbol{\theta}_c) \;+\; (1-\pi_c)\, p_f(\boldsymbol{y}_i \mid \boldsymbol{\theta}_f)\,, \tag{3.3}$$

where π_c is the cluster prior fraction, p_c and p_f are the cluster and field component likelihoods conditional on their respective parameters $\boldsymbol{\theta}_c$ and $\boldsymbol{\theta}_f$, and $\boldsymbol{\Theta} = \{\pi_c, \boldsymbol{\theta}_c, \boldsymbol{\theta}_f\}$ collects all population-level parameters (Bovy, Hogg, and Roweis, 2011; Hogg, Myers, and Bovy, 2010). Each component likelihood should in principle be an error-convolved forward model that integrates through the measurement process of Eq. (3.1) rather than treating the observed $\boldsymbol{y}_i$ as exact latent states.

The posterior probability that source i belongs to the cluster component follows by Bayes' theorem:

$$P(z_i = c \mid \boldsymbol{y}_i, \boldsymbol{\Theta}) \;=\; \frac{\pi_c \, p_c(\boldsymbol{y}_i \mid \boldsymbol{\theta}_c)}{\pi_c \, p_c(\boldsymbol{y}_i \mid \boldsymbol{\theta}_c) \;+\; (1-\pi_c)\, p_f(\boldsymbol{y}_i \mid \boldsymbol{\theta}_f)}\,, \tag{3.4}$$

where $z_i \in \{c, f\}$ is the latent membership indicator (Bovy, Hogg, and Roweis, 2011; Hogg, Myers, and Bovy, 2010). This probability is not a property of the star itself. It is conditional on the model class, the variables included, the assumed

background structure, and the treatment of uncertainty (Bernardo and Smith, 2000; Jaynes, 2003). Different generative models, metric choices, or background specifications can assign materially different posterior membership probabilities to the same source. Membership is therefore latent and model-conditional: inferred through an analytical construction rather than observed as a primitive catalogue attribute.

Once membership is recognised in this way, ambiguous objects can no longer be treated as a nuisance to be forcibly resolved. Their ambiguity is part of the inferential state of the problem, and the appropriate response is to preserve that ambiguity as part of the evidential content carried forward into population-level reasoning.

3.2.3 The inferential object is an uncertain population

The central consequence of the mixture framework is that the thesis cannot treat the object of interest as a final list of definitively known members. Per-star membership probabilities are informative intermediate summaries, but they are not the endpoint. The scientific questions of interest—the age structure, spatial extent, kinematics, and internal organisation of NGC 6383—are population-level quantities that do not attach to a uniquely known set of members but to a stellar population whose composition remains uncertain because the evidence supporting inclusion or exclusion is itself graded and model-dependent.

Under this view, probabilistic membership replaces fixed-list reasoning with uncertain-population reasoning. Stars with high inferred association strongly shape the population-level picture, ambiguous stars remain part of the inferential problem rather than being artificially removed, and the decisive distinction is no longer between accepted and rejected stars, but between uncertainty attached to

individual latent stellar states and uncertainty attached to the population-level quantities used to describe the cluster as a whole (Hogg, Myers, and Bovy, 2010; Gelman et al., 2013). If those layers are collapsed—for instance, by treating a thresholded catalogue as a clean input—measurement noise is too easily mistaken for intrinsic astrophysical spread, and downstream physical interpretation inherits that distortion.

3.2.4 Recovery depends on representation, metric, and connectivity

Even when membership is treated probabilistically, the recovered population structure depends on choices that are not uniquely fixed by the data. Clustering procedures do not uncover an algorithm-independent population from the catalogue; they partition the feature space through a restricted hypothesis class whose geometrical and statistical commitments shape what can appear as a cluster at all (Cantat-Gaudin, 2022; Banfield and Raftery, 1993).

Three classes of assumptions govern the recovery. First, *similarity and connectivity*: density-based and neighbourhood-based methods presuppose that genuine members occupy connected regions that persist across a particular density scale, but astrophysical populations need not respect uniform-density expectations, and varying-density environments can induce chaining, fragmentation, or suppression of low-contrast substructure depending on operationalisation (Campello, Moulavi, and Sander, 2013; Hunt and Reffert, 2023). Second, *metric choice and scaling*: Gaia's heteroscedastic, non-isotropic feature space means that a fixed Euclidean geometry does not treat all sources comparably, and proximity depends in part on how angular coordinates are locally represented and on whether tangent-plane approximations are treated as innocuous (Cantat-Gaudin et al., 2020; Alfonso,

García-Varela, and Vieira, 2024). Third, *dimensionality and representation*: adding observational coordinates does not automatically reveal truer structure when extra dimensions carry heterogeneous precision or reduced discriminating power, and nearest-neighbour contrast can degrade rather than sharpen (Beyer et al., 1999; Aggarwal, Hinneburg, and Keim, 2001).

The consequence is that no single neutral cluster object exists independently of these choices. A misspecified metric can produce false merges, false splits, and boundary displacements that are generated by the recovery machinery rather than by the stellar system itself (Bovy, Hogg, and Roweis, 2011; Dasgupta and Raftery, 1998). Physical and dynamical interpretation therefore inherits the assumption load built into recovery: quantities such as extent, concentration, internal dispersion, and apparent coherence are summaries of a recovered population whose composition has already been conditioned by the representation used to construct it (Vasiliev and Baumgardt, 2021; Hunt and Reffert, 2023). This propagation is the subject of the next section.

3.3 Prior sensitivity and inferential fragility

3.3.1 Priors become consequential under weak identification

Section 3.2 established that membership recovery and population inference are conditioned by representation, metric, and generative assumptions. The present section addresses a sharper problem: once the target quantity is only weakly identified by the data, the posterior is no longer dominated by the likelihood, and the inferential role of the prior becomes active rather than merely formal.

The Bayesian posterior for any set of population-level parameters can be written

as

$$p(\boldsymbol{\Theta} \mid \{\boldsymbol{y}_i\}) \;\propto\; p(\boldsymbol{\Theta}) \prod_{i=1}^{N} p(\boldsymbol{y}_i \mid \boldsymbol{\Theta})\,. \tag{3.5}$$

When the likelihood surface is sharply peaked, $p(\boldsymbol{\Theta} \mid \{\boldsymbol{y}_i\})$ is insensitive to reasonable prior choices and the inference is said to be well identified. When the likelihood is flat or multimodal over a region of parameter space—because the signal is weak, the sample is sparse, the observational boundary truncates the informative regime, or the background and cluster overlap substantially—the posterior inherits a correspondingly larger fraction of its structure from $p(\boldsymbol{\Theta})$ (Gelman et al., 2013; Gustafson, 2005). Under those conditions, apparently modest stabilising choices cease to be innocent: regularisation, weak informativeness, flatness, and parameterisation all contribute substantive structure to the result.

This regime is not exceptional in Gaia-era cluster work. The distinction between a genuine halo, a dissolving tail, and structured field contamination is often only weakly constrained by the data, especially at the periphery of a candidate population where observational precision degrades and background contrast diminishes (Hunt and Reffert, 2023; Hunt et al., 2025). In that setting, prior structure governs how much coherence, persistence, and complexity the analysis is permitted to attribute to poorly constrained features of the observed phase space.

3.3.2 Distance inference as an anchor case

Distance inference provides a particularly clear illustration of prior sensitivity. For a star with observed parallax $\varpi_i \pm \sigma_{\varpi,i}$, the likelihood of the distance d_i is obtained by propagating the Gaussian parallax error through the nonlinear transformation $\varpi = 1/d$, which yields an asymmetric and potentially multimodal function in distance space (Luri et al., 2018; Bailer-Jones, 2015). The posterior distance therefore requires an explicit prior on the spatial distribution of stars.

A widely adopted family is the exponentially decreasing space-density prior,

$$p(d) \;=\; \begin{cases} \dfrac{1}{2\,L^3}\; d^2\; \exp\!\left(-\dfrac{d}{L}\right) & d > 0\,, \\[2ex] 0 & d \le 0\,, \end{cases} \tag{3.6}$$

where L is a length-scale parameter that encodes the assumed density fall-off (Bailer-Jones, 2015; Bailer-Jones et al., 2021). The d^2 factor reflects the volumetric prior that arises naturally from a uniform space density, while the exponential taper regularises the tail against implausibly large distances. The key inferential point is that, when $\sigma_{\varpi,i}/\varpi_i$ is non-negligible, the posterior $p(d_i \mid \varpi_i, \sigma_{\varpi,i})$ is materially shaped by L: different length-scales assign different posterior distances to the same observed parallax, especially in the regime where the fractional parallax error exceeds ~ 20 per cent (Luri et al., 2018).

This is not a merely formal nicety. A prior that appears flat when written on one distance-like quantity does not remain flat under reparameterisation, because the induced weighting over physically meaningful configurations changes with the coordinate (Bailer-Jones, 2015; Luri et al., 2018). In weakly identified regimes, astrophysical conclusions about distance-compatible membership, extinction, or location within a common structure are therefore prior-conditioned rather than purely data-determined.

3.3.3 Hyperpriors and complexity control

The same fragility deepens at higher inferential levels. Even when the immediate likelihood specification is held fixed, the analysis requires choices about how much heterogeneity to permit and how aggressively to pool weak structure toward simpler explanations. In a Bayesian vocabulary these are hyperprior and shrinkage questions; in the broader setting of astrometric clustering they appear

as regularisation choices that control whether the posterior landscape sustains multiple low-contrast components or is forced toward a smaller set of persistent structures (Campello et al., 2015; McInnes, Healy, and Astels, 2017).

This matters most when the latent astrophysical dispersion is only weakly identified relative to measurement error. Partial pooling is not a neutral convenience: it probabilistically borrows strength across the ensemble by shrinking weakly measured cases toward the population-level structure, and that borrowing can either stabilise weak structure or suppress it depending on how strongly the global scale is regularised (Gelman, 2006; Gelman et al., 2013). Restrictive variance priors can over-compress weak dispersion toward homogeneity, while heavier-tailed scale priors allow more heterogeneity to survive; neither choice is neutral when the data are only weakly informative about the population scale (Gelman, 2006; Polson and Scott, 2012; Simpson et al., 2017).

In density-based clustering, the analogous choices concern persistence and extraction rules. Methods that privilege stability across a wide density range—such as the λ-persistence hierarchy in HDBSCAN—effectively encode a preference for components that remain coherent under progressive simplification of the density landscape (Campello, Moulavi, and Sander, 2013; Campello et al., 2015). That preference can suppress tenuous substructure that may be scientifically plausible yet only weakly expressed in the observed sample. Under strong identification, this is a benign simplification. Under weak identification, it becomes part of the inferential content of the result.

3.3.4 Inferential fragility limits what can be claimed as stable

The preceding arguments motivate a precise doctrinal conclusion. Inferential fragility is present when scientifically relevant conclusions change materially

across plausible alternatives in prior structure, regularisation, parameterisation, or complexity control, especially in regions where the target is weakly identified by the observed data. This is not ordinary posterior uncertainty: the issue is not wide spread within a fixed inferential frame, but sensitivity to the frame itself.

The limitation extends beyond descriptive recovery. Once the existence, extent, or morphology of the target remains conditionally supported, downstream quantities such as intrinsic velocity dispersion, dynamical mass, virial state, or relaxation time cannot be treated as if they were detached from that dependence. They are model-conditioned summaries whose claim-status inherits the fragility of the structure on which they are built (Read and Steger, 2017; Read et al., 2020).

Where the scientific object is robust across admissible alternatives, stronger interpretation becomes defensible. Where it is not, the result must be carried forward as conditionally supported rather than stably established. That distinction leads directly to the question of what interpretive domain remains valid once inferential fragility has been recognised.

3.4 Domain of interpretive validity

3.4.1 The interpreted population is catalogue-conditioned

Any physical interpretation advanced after membership construction must be understood as referring to a recoverable population rather than to the cluster in an unrestricted ontological sense. In Gaia DR3, the available stellar sample is already delimited by the survey selection function, the observational footprint, pipeline products, and quality controls imposed before any inferential step (Gaia Collaboration et al., 2023; Everall and Boubert, 2021; Cantat-Gaudin et al., 2023). The empirical object carried forward into the analysis is therefore not the

full physical population of NGC 6383 but the subset that survives detectability, catalogue inclusion, and thesis-level filtering.

That restriction is interpretive, not merely operational. Magnitude limits, signal-to-noise thresholds, crowding-dependent failures, and quality cuts define which portions of the underlying population can enter the evidentiary base at all (Gaia Collaboration et al., 2023; Loredo, 2004). In Gaia, effective inclusion varies jointly with brightness, colour, sky position, crowding, scanning-law history, and the availability of specific data products such as radial velocities (Everall and Boubert, 2021; Boubert and Everall, 2021). The observed density field is therefore a selection-shaped realisation whose local representation depends on the interaction between astrophysical content and survey response (Rix et al., 2021; Boubert and Everall, 2021). Quality-based exclusions can further reshape the admitted population in astrophysically non-neutral ways—for example, when problematic astrometric solutions preferentially trace unresolved companions rather than instrumental failure (Penoyre, Belokurov, and Evans, 2022b; El-Badry, 2024). Later physical summaries therefore refer to the catalogue-conditioned population defined by the adopted data products and admissibility rules, not to the abstract totality of stars physically associated with the cluster.

3.4.2 Unresolved multiplicity and youth-driven complexity

Two additional validity boundaries restrict the passage from catalogue summary to physical interpretation.

First, *unresolved multiplicity* perturbs both kinematic and photometric observables in ways that can mimic intrinsic cluster physics. Astrometric photocentre motion from unresolved companions inflates the apparent proper-motion and parallax scatter, making an equilibrium population appear more dispersed or less bound

than it truly is (Penoyre, Belokurov, and Evans, 2022a; Kouwenhoven et al., 2009; Gieles et al., 2021). In the colour–magnitude diagram, unresolved binaries brighten sources, displace them from the single-star locus, and can imitate or amplify apparent age dispersion (Riello et al., 2021; El-Badry, 2024). A centrally enhanced binary fraction can also project as a false segregation signal if the detectability and radial incidence of binaries vary across the sample (Belokurov et al., 2020; Albrow, 2024). Apparent broadening in either kinematic or photometric space therefore cannot be read naively as a pure age, dispersion, or population-structure effect without accounting for the multiplicity channel.

Second, *youth-driven structural complexity* limits whole-cluster readings. Young stellar aggregates need not be dynamically mixed, morphologically regular, or sharply bounded; they may retain asymmetry, substructure, or locally coherent features whose observational appearance depends strongly on selection and measurement quality (Hunt and Reffert, 2023). In such systems, boundary language must remain disciplined. Observed extent, fitted scale radius, catalogue footprint, and physical boundary are not interchangeable: the first is fixed by observation, the second by a chosen summarising form, the third by selection, and the fourth is a stronger astrophysical claim not automatically licensed by the preceding three (Everall and Boubert, 2021; Hunt et al., 2025). The same caution extends to dynamical interpretation: in a system that may still retain expansion signatures, incomplete phase mixing, or formation-driven substructure, quantities such as virial state or boundedness fraction cannot be treated as directly read off from observed kinematics without strong additional assumptions about equilibrium and geometry (Kuhn et al., 2019; Wright et al., 2024; Read et al., 2020).

3.4.3 Model serviceability does not license global physical realism

The preceding restrictions acquire their full force when translated into claim-status discipline. A model may remain scientifically serviceable even when it is not globally adequate as a literal description of the system. Under misspecification, fitted summaries can converge to pseudo-true values that best approximate reality within a restricted model family rather than to uniquely correct physical parameters (White, 1982; Brynjarsdóttir and O'Hagan, 2014). This distinction is decisive because several of the summaries later used in this thesis—the King profile parameters, the half-mass relaxation time, the virial ratio—are valuable precisely as controlled approximations, not as guarantees of literal global realism.

The correct response is not to discard serviceable models but to discipline what they are allowed to mean. The relevant distinction is between a model that yields useful local or conditional summaries of the recoverable population and one that would justify unqualified claims about the full physical system. The former is attainable under the analysis described in this thesis. The latter is not secured simply by the internal coherence of fitted outputs.

3.5 Claim-status rule for later chapters

The preceding sections establish a graded doctrine of interpretive authority for NGC 6383. The analysis can support local and conditional claims about the recoverable, filtered, and partially purified population represented in the adopted catalogue. It can also support careful summaries of how that population is organised in the admitted observables. What it cannot automatically support are global and unqualified statements about the full physical cluster, as though

catalogue conditioning, residual contamination, unresolved multiplicity, and the Gaia uncertainty architecture of Section 3.1.2 were negligible.

This distinction applies across the main scientific targets of the thesis. Outer structure may be conditionally informative, but not equivalent to a definitive physical boundary. Mass segregation or central concentration may be descriptively real within the examined sample, while remaining vulnerable to multiplicity, incompleteness, and truncation of the low-density outskirts. CMD width may encode genuine astrophysical diversity, yet it does not by itself isolate age spread from binaries, extinction structure, or variability. Internal kinematic organisation may summarise the measured population, but it is bounded by the systematic covariance floor and by perspective effects that can imitate physical structure (Vasiliev and Baumgardt, 2021; Hunt and Reffert, 2023; G. van de Ven et al., 2006).

The operative rule for the remainder of the dissertation is therefore:

> *Later chapters may interpret derived summaries physically only within the validity domain established here. When a result concerns the filtered and recoverable population directly represented in the data, and when the relevant confounding mechanisms have not been silently ignored, a local or conditional physical interpretation is admissible. When a result would require the sample to stand transparently for the full cluster, or would require the fitted description to be globally realistic, that stronger reading is not licensed by Chapter 3.*

We accordingly distinguish three claim-status levels that will be referenced throughout Chapters 5–8[1]:

[1]The distinction between data-driven, model-conditioned, and interpretive claims is related to, but not identical with, the taxonomy of model-ladenness proposed by Bokulich (2020), who classified the *roles* that models play in data processing rather than the *epistemic status* of

1. **Data-driven:** The claim follows directly from the observations under the adopted measurement model and membership construction, with uncertainty propagated through the generative framework. Examples: the posterior parallax distribution, the fitted King-profile parameters, the pseudoprobability ranking.

2. **Model-conditioned:** The claim requires an explicit generative model, prior specification, or structural assumption beyond the measurement model itself. Examples: the cluster age from isochronal fitting, the intrinsic velocity dispersion after deconvolution, the half-mass relaxation time.

3. **Interpretive:** The claim involves physical or causal reasoning that extends beyond what the generative model can directly produce. Examples: the dynamical-state interpretation, the primordial-vs-dynamical attribution of segregation, the gas-expulsion history.

Each level inherits the fragility of the levels below it, and claims at any level must respect the catalogue conditioning, measurement structure, and inferential assumptions documented in this chapter. Results may be read physically only under explicit domain control. Where the argument reaches beyond that control, the appropriate language is conditionality, locality, and provisional scope rather than certainty. The formal tools required to operate within these constraints are developed in Chapter 4.

the resulting claims. The classical notion that data are always "theory-laden" (Suppes, 1969) motivates our tripartite approach but does not provide the operational labelling scheme we require for a multi-method cluster analysis.

Chapter 4

Physical and Statistical Formalism

Chapter 3 established that every physical claim about NGC 6383 is bounded by the uncertainty structure of the inference from which it derives, the prior assumptions on which that inference depends, and the adequacy of the generative model within which the data are interpreted. The present chapter builds the formal tools required to operate within those constraints. We develop the probabilistic inference architecture that governs all subsequent analysis (Section 4.1), formalise membership as a latent population-level quantity rather than a fixed preprocessing label (Section 4.2), construct the parallax-to-distance inference framework (Section 4.3), establish the density-based clustering formalism (Section 4.4), and define the photometric and isochronal inference that yields the cluster's age, metallicity, and extinction (Section 4.5). Each section inherits the general architecture of Section 4.1 and operates under the claim-status discipline of Section 3.5. The application of this formal machinery to the specific observational dataset of NGC 6383 begins in Chapter 5.

4.1 Probabilistic inference architecture

The physical characterisation of NGC 6383 from Gaia-era astrometry and photometry requires a formal inferential architecture that separates what is observed from what is inferred, propagates realistic uncertainty through every stage of the analysis, and diagnoses the adequacy of the assumptions on which physical claims rest. This section establishes that architecture in general terms. The formal objects defined here—the measurement model, the likelihood, the posterior, the hierarchical structure, the prior, the sampling machinery, and the model-criticism framework—constitute the inferential grammar that Sections 4.2–4.5 instantiate for each physical domain.

4.1.1 The measurement model: from observed to latent quantities

We begin from a sharp architectural separation between the *process model*, which encodes the true underlying astrophysical state, and the *data model*, which encodes the observational processes through which that state becomes a catalogued number (Andreon and Hurn, 2013). Within this architecture, the true physical quantities—the noise-free parallax of a star, its intrinsic proper motion, its bolometric luminosity—are latent variables that are simultaneously constrained by the data from below and by population-level priors from above (Bovy, Hogg, and Roweis, 2011). Parameters that are necessary for correct likelihood specification but carry no direct scientific interest—binary mass ratios, individual extinction values, astrometric zero-point offsets—enter as nuisance parameters, handled through marginalisation (Section 4.1.3).

The generative perspective imposes a strict directionality: the model specifies how physical parameters produce observations; the data constrain which parameters

are consistent with what was observed. This forward direction is the only direction in which the model is well-posed. Inverting the mapping directly—as in naive parallax inversion—violates this directionality and produces estimators whose statistical properties are pathological under realistic noise (Loredo, 2004). The Bayesian framework respects this by evaluating the likelihood in the space of observables while defining the prior in the space of physical quantities (Gelman et al., 2013).

4.1.2 Likelihood construction under heteroscedastic and correlated errors

The likelihood function quantifies the probability of the observed data given a proposed set of model parameters. It carries the entirety of the information that the data contribute to the inference and simultaneously represents the most consequential source of inferential failure: if the likelihood misrepresents the data-generating process, the posterior converges confidently to a biased region of the parameter space (Moran, Cunningham, and Blei, 2022).

For a set of N stars with observed data vectors $\boldsymbol{y}_i$ and individual covariance matrices $\boldsymbol{C}_i$, the likelihood under a generative model $\boldsymbol{f}(\boldsymbol{\theta})$ takes the general form

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^{N} p(\boldsymbol{y}_i \mid \boldsymbol{\theta}, \boldsymbol{C}_i), \tag{4.1}$$

where $\boldsymbol{\theta}$ collects the model parameters and each $p(\boldsymbol{y}_i \mid \boldsymbol{\theta}, \boldsymbol{C}_i)$ is typically a multivariate Gaussian centred on the forward-model prediction with covariance $\boldsymbol{C}_i$. The point-dependent covariance matrices are mandatory: Gaia astrometric uncertainties are strongly heteroscedastic and the five-parameter solution carries non-trivial off-diagonal terms that encode geometric constraints from the scanning law (Kelly, 2007; Luri et al., 2018).

When the noise is not well described by a Gaussian, robustification through heavy-tailed distributions provides protection against outliers. Replacing the Gaussian with a Student-t distribution naturally downweights extreme residuals—from unresolved binary motion or instrumental artefacts—without requiring arbitrary sigma-clipping thresholds (Martin and Mortlock, 2025). Similarly, when the catalogue is magnitude-limited and spatially incomplete, the likelihood must be truncated by dividing the base probability by the survey selection function integrated over the observable domain, preventing catastrophic Malmquist biases (Rix et al., 2021; Boubert and Everall, 2020).

4.1.3 Latent variable marginalisation and the posterior distribution

The posterior distribution is the central target of Bayesian inference. For model parameters $\boldsymbol{\theta}$ and observed data $\boldsymbol{Y}$, Bayes' theorem gives

$$p(\boldsymbol{\theta} \mid \boldsymbol{Y}) = \frac{p(\boldsymbol{Y} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})}{p(\boldsymbol{Y})}\,, \tag{4.2}$$

where $p(\boldsymbol{Y} \mid \boldsymbol{\theta})$ is the likelihood, $p(\boldsymbol{\theta})$ is the prior, and $p(\boldsymbol{Y}) = \int p(\boldsymbol{Y} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta}$ is the marginal likelihood or evidence. The evidence is typically intractable in high dimensions, and standard Bayesian computation bypasses it entirely through sampling algorithms that generate draws from $p(\boldsymbol{\theta} \mid \boldsymbol{Y})$ without requiring explicit normalisation (Gelman et al., 2013).

Marginalisation is the operation by which the posterior for the parameters of interest $\boldsymbol{\theta}_\mathrm{s}$ is obtained from the joint posterior by integrating over the nuisance parameters $\boldsymbol{\theta}_\mathrm{n}$:

$$p(\boldsymbol{\theta}_\mathrm{s} \mid \boldsymbol{Y}) = \int p(\boldsymbol{\theta}_\mathrm{s},\, \boldsymbol{\theta}_\mathrm{n} \mid \boldsymbol{Y})\, \mathrm{d}\boldsymbol{\theta}_\mathrm{n}\,. \tag{4.3}$$

This operation propagates the full uncertainty contributed by the nuisance parameters into the marginal posteriors, producing wider and more honest uncertainty estimates than conditioning on fixed nuisance values (Trotta, 2008).

A critical computational constraint arises with discrete latent variables, such as the indicator $z_i \in \{1, \ldots, K\}$ that assigns star i to one of K mixture components. Gradient-based samplers require a continuous target distribution, so we analytically marginalise the discrete indicators out of the likelihood before sampling. For a K-component mixture with component weights $\boldsymbol{\pi}$ and component-specific densities p_k, the marginalised likelihood for star i is

$$p(\boldsymbol{y}_i \mid \boldsymbol{\theta}, \boldsymbol{\pi}) = \sum_{k=1}^{K} \pi_k \, p_k(\boldsymbol{y}_i \mid \boldsymbol{\theta}_k) \,. \tag{4.4}$$

This operation is architecturally fundamental for the membership inference of Section 4.2, where the cluster–field assignment of each star is handled as a latent quantity rather than a fixed label (Betancourt, 2018).

4.1.4 Hierarchical model architecture and reparameterisation

When the inferential target is a stellar population rather than an individual star, we adopt a hierarchical model in which the parameters θ_i describing individual stars are draws from a shared population-level distribution governed by hyperparameters $\boldsymbol{\phi}$:

$$\theta_i \sim p(\theta \mid \boldsymbol{\phi}), \qquad y_i \mid \theta_i \sim p(y \mid \theta_i, \sigma_i) \,. \tag{4.5}$$

The hyperparameters $\boldsymbol{\phi}$—encoding the mean cluster distance, the intrinsic velocity dispersion, the characteristic age—are the population-level quantities of scientific interest. This architecture produces partial pooling: well-measured stars anchor

the hyperparameters, while noisy stars are shrunk toward the population mean, minimising the overall mean squared error (Gelman et al., 2013).

The hierarchical structure introduces characteristic geometric pathologies. The most consequential is the *funnel geometry*, which arises when a hierarchical variance hyperparameter approaches zero: the individual-level parameters are forced into an infinitely narrow corridor, creating extreme curvature that gradient-based samplers cannot traverse. The standard remedy is non-centred reparameterisation, in which the individual-level parameters are expressed as deterministic transformations of auxiliary standard-normal variables, homogenising the posterior curvature (Betancourt and Girolami, 2013).

Label switching is a second geometric pathology specific to mixture models. Because the likelihood of Eq. (4.4) is invariant under permutation of the component indices, the posterior contains $K!$ symmetric modes. Without symmetry-breaking constraints—such as ordering the components by velocity dispersion—the sampler transitions between modes and corrupts marginal summaries (Baribault and Collins, 2025). Intrinsic parameter degeneracies, such as the age–extinction ridge in colour–magnitude fitting, produce near-singular Hessians that further impede sampling. These pathologies are addressed domain-specifically in Sections 4.3–4.5.

4.1.5 Prior specification as formal scientific commitment

We adopt weakly informative priors that suppress physically implausible regions while remaining broad enough that the likelihood dominates where the data carry genuine constraining power (Gelman, Simpson, and Betancourt, 2017; Gelman et al., 2020).

Prior-data conflict arises when the prior and likelihood concentrate on disjoint regions of the parameter space. In high-dimensional models, this conflict can pull

the joint posterior into low-probability valleys between the two modes, producing parameter estimates supported by neither the data nor the prior (Evans and Moshonov, 2006). The diagnostic is the prior predictive check: synthetic data generated from the model using prior draws alone are compared to the observations. If the prior predictive distribution assigns negligible probability to data resembling the actual observations, the prior must be revised (Gabry et al., 2019).

In the data-poor regime—characteristic of faint NGC 6383 members with $\sigma_{\varpi}/\varpi > 0.2$—the likelihood flattens and the posterior becomes prior-dominated. This dominance is the correct formal behaviour when the data lack constraining power. The scientific obligation is to demonstrate it explicitly through prior sensitivity analysis: repeating the inference under alternative priors and assessing posterior stability (Gelman et al., 2020). Prior-dominated parameters must be stated as conditional on the adopted prior rather than as data-driven results, a requirement operationalised in Section 4.2.5.

4.1.6 Posterior sampling, approximation, and convergence

The posteriors arising from the models in this thesis are high-dimensional, correlated, and analytically intractable. We characterise them through Markov Chain Monte Carlo sampling, generating correlated draws whose stationary distribution matches the target posterior under ergodicity and detailed balance (Robert and Casella, 2005).

Hamiltonian Monte Carlo (HMC) overcomes the diffusive inefficiency of random-walk algorithms by introducing auxiliary momentum variables and simulating Hamiltonian dynamics on the posterior surface. The deterministic trajectory glides along the contours of the target density, proposing distant states with near-unit acceptance probability (Betancourt, 2017; Neal, 2012). The No-U-Turn

Sampler (NUTS) extends HMC by adaptively determining the trajectory length, eliminating manual tuning of the integration time (Hoffman and Gelman, 2011).

Despite its efficiency, HMC is acutely sensitive to pathological posterior geometries. When the leapfrog integrator encounters extreme curvature, the simulated Hamiltonian energy diverges, producing a *divergent transition*—a rigorous diagnostic that the sampler has failed to represent the target geometry in that region (Betancourt, 2016). Any inference proceeding despite divergences is formally compromised.

We assess convergence and validity through a hierarchy of diagnostics. The rank-normalised split-$\hat{R}$ statistic compares within-chain and between-chain variance; convergence requires $\hat{R} < 1.01$ for all parameters (Vehtari et al., 2021). The bulk and tail effective sample sizes (Bulk-ESS and Tail-ESS) must both exceed 400 to ensure negligible Monte Carlo error. The Energy Bayesian Fraction of Missing Information (E-BFMI) quantifies momentum-resampling efficiency; values below 0.3 indicate insufficient kinetic energy to traverse the posterior breadth (Betancourt, 2016). In the COSMIC framework, these diagnostics govern all posterior claims.

4.1.7 Model criticism and posterior predictive checking

A valid posterior sample does not guarantee that the model is adequate. A model can satisfy all convergence diagnostics while remaining fundamentally misspecified with respect to the data-generating process. The final component of the inferential architecture is therefore a framework for assessing model adequacy, distinct from sampling validity.

The posterior predictive distribution provides the primary diagnostic. Replicated datasets are simulated by drawing parameter values from the fitted posterior and

generating synthetic observations through the full generative model. Systematic discrepancies between the observations and the posterior predictive replicates expose specific likelihood failures: an inability to reproduce the CMD morphology, the proper-motion structure, or the spatial residual pattern signals a missing physical or statistical ingredient (Schad, Betancourt, and Vasishth, 2021; Gelman et al., 2020). The Bayesian workflow treats these discrepancies as diagnostic signals triggering targeted model expansion—the addition of a nuisance parameter, the relaxation of a distributional assumption, or the introduction of a previously neglected physical process (Moran, Cunningham, and Blei, 2022).

As Box (1976) emphasised, all models are approximations, and the objective of inference is not to discover a true model but to construct a representation whose misspecifications are orthogonal to the scientific questions being asked. A model may be adequate for constraining the mean cluster distance while being inadequate for the intrinsic velocity dispersion. Model adequacy is therefore question-dependent, and the posterior predictive framework must target the specific physical summaries that we intend to report.

For NGC 6383, the COSMIC framework must demonstrate that its generative models are adequate for the specific claims the thesis makes: membership-conditioned physical inference, structural characterisation, and age determination. The formal machinery for these inferences is developed in the remainder of this chapter, with each domain inheriting the architecture established here and submitting to the model-criticism framework defined in this subsection.

4.2 Latent population inference and membership formalism

Every physical quantity reported for NGC 6383 is conditioned on which stars are treated as members of the system. Chapter 3 established that treating membership as a fixed, binary preprocessing label is inferentially insufficient: the recovered population is not a property of the cluster alone but of the interaction between the cluster, the field, the survey, and the method used to separate them (Section 3.2.1). This section formalises that insight. We define the cluster–field decomposition as a probabilistic mixture, establish membership probability as a posterior marginal quantity, identify the uncertain stellar population as the formal inferential object, introduce hierarchical shrinkage at the population level, address the sensitivity of the recovered population to prior specification and model misspecification, and close by defining physical summary statistics as expectations over the uncertain population.

4.2.1 The cluster–field decomposition as a probabilistic mixture

A stellar field observed toward NGC 6383 contains stars drawn from at least two physically distinct generative processes: the cluster population, which shares common distance, kinematics, age, and chemical composition, and the Galactic field, which is kinematically hot and chemically heterogeneous. We decompose the observed sample through a finite mixture model. For a star i with observed data vector $\boldsymbol{y}_i$, the mixture likelihood takes the form

$$p(\boldsymbol{y}_i \mid \boldsymbol{\theta}, \pi) = \pi \, p_{\mathrm{cl}}(\boldsymbol{y}_i \mid \boldsymbol{\theta}_{\mathrm{cl}}) + (1 - \pi) \, p_{\mathrm{fld}}(\boldsymbol{y}_i \mid \boldsymbol{\theta}_{\mathrm{fld}}), \tag{4.6}$$

where π is the mixing weight representing the prior probability of cluster membership, p_{cl} and p_{fld} are the component-specific densities, and $\boldsymbol{\theta}_{\mathrm{cl}}$, $\boldsymbol{\theta}_{\mathrm{fld}}$ collect the component parameters. This formulation replaces the binary question of membership with a continuous probability that reflects the degree to which each star's observables are consistent with each generative component (Guglielmetti, Fischer, and Dose, 2009).

The mixture can be extended to accommodate additional components—a second kinematic population, a contamination component with explicitly broad variance designed to absorb catastrophic outliers—without altering the fundamental architecture (Stenning et al., 2016). Li, Li, and Shao (2025) demonstrated that integrating the full Gaia astrometric covariance matrix directly into the mixture likelihood ensures that the posterior membership probabilities account for structured, heteroscedastic individual errors. In the photometric domain, the mixture evaluates each star's probability under the cluster isochrone model against an empirical field-star distribution in colour–magnitude space (Perren, Vázquez, and Piatti, 2015).

The dominant operational paradigm in Gaia-era cluster science synthesises density-based and mixture-model approaches in a two-stage architecture. HDBSCAN identifies high-confidence topological cores without shape assumptions (Section 4.4), and its labels initialise a subsequent Gaussian mixture model that refines the membership probabilities (Noormohammadi, Khakian Ghomi, and Javadi, 2024; Cantat-Gaudin et al., 2018). This hybrid strategy reflects a formal recognition that the two methods produce fundamentally different kinds of membership outputs, a distinction developed in the next subsection.

4.2.2 Membership probability as a posterior marginal quantity

The membership indicator z_i for each star is a discrete latent variable that is analytically marginalised out of the likelihood, as described in Section 4.1.3. After marginalisation and posterior sampling, the membership probability for star i is recovered as a posterior marginal:

$$p(z_i = \mathrm{cl} \mid \boldsymbol{y}_i, \boldsymbol{\theta}) = \frac{\pi \; p_{\mathrm{cl}}(\boldsymbol{y}_i \mid \boldsymbol{\theta}_{\mathrm{cl}})}{\pi \; p_{\mathrm{cl}}(\boldsymbol{y}_i \mid \boldsymbol{\theta}_{\mathrm{cl}}) + (1 - \pi) \; p_{\mathrm{fld}}(\boldsymbol{y}_i \mid \boldsymbol{\theta}_{\mathrm{fld}})} \, . \tag{4.7}$$

This quantity is calibrated to the generative model: a membership probability of 0.8 means that, under the assumed model, the data are four times more likely under the cluster component than under the field component. It integrates over parameter uncertainty through the posterior: as the parameters are sampled, the membership probabilities fluctuate, and the reported values represent expectations over the full parameter posterior (Betancourt, 2018).

A distinction that is doctrinally critical for this thesis is the difference between these posterior marginal probabilities and the soft outputs of HDBSCAN. As we establish in Section 4.4.2, HDBSCAN's membership values are geometric persistence metrics—λ-persistence ratios that normalise each point's density-parameter fallout between the cluster's topological birth and death thresholds (Malzer and Baum, 2020). They measure relative spatial centrality within the density manifold, not the conditional probability of generative origin. A star with an HDBSCAN value of 0.9 resides deep within the persistent density core; it does not follow that it has a ninety per cent probability of cluster origin. The persistence metric carries no information about the field distribution, no calibration against a generative likelihood, and no propagation of parameter uncertainty.

The practical consequence is that the COSMIC framework must convert the

topological outputs of HDBSCAN into probabilistically calibrated quantities through the mixture-model refinement stage before using them in downstream physical inference.

4.2.3 The uncertain stellar population as the formal inferential object

The output of the membership inference is not a list of confirmed members but a probability distribution over the composition of the stellar population. The formal inferential object of this thesis is the uncertain stellar population defined by the full set of membership posteriors—not a fixed catalogue.

A hard threshold—selecting all stars with $p(z_i = \mathrm{cl}) > p_{\mathrm{thr}}$—converts the uncertain population back into a fixed list and discards the information encoded in the membership probabilities of excluded stars. Stars near the boundary, which carry the greatest uncertainty, are precisely those whose inclusion or exclusion most affects the cluster periphery, the low-mass end of the mass function, and the extended spatial structure. Pulgar-Escobar et al. (2024) demonstrated this threshold-dependence for NGC 6383, recovering membership counts of 321, 254, 202, and 161 at probability thresholds of 0.5, 0.6, 0.7, and 0.8 respectively, with each threshold producing a qualitatively different picture of the cluster's spatial extent and mass distribution.

The uncertain population is further conditioned by the catalogue from which it is drawn. The Gaia survey is magnitude-limited, spatially incomplete in crowded regions, and subject to quality filters that preferentially exclude faint, red, or astrometrically complex sources. Any physical claim derived from this population must be understood as a claim about the recoverable population under the adopted survey, quality controls, and membership construction—not about an abstract

complete realisation of the cluster (Section 3.4.1).

4.2.4 Hierarchical population parameters: shrinkage and pooling

The population-level parameters of the cluster—its mean proper motion, mean parallax, intrinsic velocity dispersion, and spatial extent—are inferred through the hierarchical architecture of Eq. (4.5). Stars with high measurement precision and high membership probability anchor the hyperparameters, while uncertain or ambiguous stars are shrunk toward the population mean (Morris and Lysy, 2012). This shrinkage is statistically optimal but has a physical consequence: it compresses the inferred extent toward the well-measured core, potentially underrepresenting the true spatial and kinematic breadth of the system.

For NGC 6383, where the scientific questions specifically concern the outer structure, mass segregation, and dynamical state, this compression is consequential. The faint, peripheral members that define the halo and tidal features are precisely the stars with the largest measurement uncertainties. Under hierarchical shrinkage, they are systematically pulled toward the cluster centre in the inferred parameter space. The sensitivity of the recovered spatial extent to the strength of the shrinkage—governed by the inferred population dispersion and its prior—must therefore be assessed explicitly, and any structural claim must acknowledge the degree to which the hierarchical architecture shapes the recovered population.

4.2.5 Prior sensitivity and misspecification effects on the recovered population

The population recovered by the membership inference depends not only on the data but also on the structural assumptions embedded in the mixture model.

Misspecification of the number of components, the parametric form of each component, or the priors on the mixing weights and component parameters can alter the recovered population in ways that propagate silently into downstream summaries—the inferential fragility established in Section 3.3.4.

The most immediate risk is the assumed number of components K. A model with $K = 2$ cannot accommodate a kinematically distinct sub-population and will assign its members to the closer available component. An overspecified model may fragment a coherent population into artificial sub-groups (McLachlan and Peel, 2000). Gaussian components assume ellipsoidal structure in the feature space, which is reasonable for a relaxed core but poor for a fractal star-forming complex. When covariance matrices are poorly conditioned, the maximum likelihood solution can drive a component's variance toward zero, producing a degenerate likelihood that must be regularised through explicit penalisation (Hathaway, 1985). Label switching, discussed in Section 4.1.4, has specific consequences for membership: the natural symmetry-breaking constraint is to require that the cluster component have a smaller velocity dispersion or a more concentrated spatial distribution than the field (Betancourt, 2018).

The prior on the mixing weight π directly affects the recovered population. A prior concentrating π near zero produces a conservative, high-purity but potentially incomplete cluster. A prior permitting large π recovers peripheral members at the cost of increased contamination. We assess this sensitivity explicitly for NGC 6383, reporting how the recovered population varies across the plausible range of mixing-weight priors.

4.2.6 Physical summary statistics as population-integrated quantities

The physical properties of a cluster are population-level summaries that must be computed as expectations over the uncertain membership and the uncertain individual parameters simultaneously (Gelman et al., 2013). For a population-level quantity Q that depends on individual stellar parameters $\{\theta_i\}$ and membership indicators $\{z_i\}$, we write

$$\langle Q \rangle = \int Q(\{\theta_i\}, \{z_i\})\, p(\{\theta_i\}, \{z_i\} \mid \boldsymbol{Y})\, \mathrm{d}\{\theta_i\} \tag{4.8}$$

where the integral runs over the individual latent parameters and the sum over discrete membership indicators is implicit. The mean cluster proper motion, for instance, is the expectation of the proper motion over the joint posterior of the membership indicators and the individual kinematic parameters—not the arithmetic average of stars above a threshold.

The velocity dispersion is a hyperparameter of the hierarchical model, inferred jointly with the individual velocities and membership probabilities; its posterior reflects the full uncertainty structure of the population inference (Baumgardt and Hilker, 2018). Unresolved binaries inflate the observed dispersion because their orbital motions contribute scatter unrelated to the cluster's dynamical state; if the binary fraction is not modelled explicitly, the inferred dispersion will be systematically overestimated (Randich et al., 2018). Similarly, the systemic velocity must account for spatial projection effects and the measurement covariance rather than simple averaging of radial velocity measurements (Watkins et al., 2015).

The photometric properties—age, metallicity, extinction—are equally conditioned on membership. The colour–magnitude diagram entering the isochronal likelihood

of Section 4.5 is a probabilistically weighted ensemble in which each star contributes in proportion to its membership probability. The resulting age posterior reflects not only the photometric uncertainty and model systematics, but also the membership uncertainty that governs which stars are contributing and with what weight.

No physical summary reported for NGC 6383 is a fixed-sample property. Every quantity is an expectation over the uncertain population, conditioned on the survey, the model, and the priors. The interpretive validity of these summaries is bounded by the claim-status framework of Chapter 3: they are defensible within the domain defined by the adopted catalogue, quality controls, membership construction, and generative model, and they lose their warrant when generalised beyond that domain.

4.3 Parallax-to-distance inference

The conversion of observed trigonometric parallax into physical stellar distance is among the most fundamental inferential operations in Galactic astrophysics. The geometric relationship $d = 1/\varpi$ is non-linear, and this non-linearity transforms symmetric, well-behaved measurement noise in parallax space into asymmetric, heavy-tailed distributions in distance space. For NGC 6383, where the membership encompasses both bright core stars with sub-milliarcsecond Gaia precision and faint pre-main-sequence candidates whose fractional parallax errors σ_ϖ/ϖ approach or exceed unity, the inferential consequences are severe: naive parallax inversion produces biased or undefined distance estimates precisely for the stars that govern the inferred spatial extent, mass function, and dynamical state of the cluster. This section establishes the formal framework that avoids these pathologies, following the general probabilistic architecture of Section 4.1 and operating strictly within the validity domain established in Chapter 3.

4.3.1 The non-linearity of the parallax-to-distance transformation

The foundation of Bayesian astrometric distance inference is the construction of the likelihood in the observable domain rather than the physical domain. For a star i with observed parallax $\varpi_{\mathrm{obs},i}$, reported uncertainty $\sigma_{\varpi,i}$, and true distance $d_i > 0$, we write the parallax likelihood as

$$p(\varpi_{\mathrm{obs},i} \mid d_i,\, \sigma_{\varpi,i}) = \mathcal{N}\left(\varpi_{\mathrm{obs},i} \;\middle|\; \frac{1}{d_i},\; \sigma^2_{\varpi,i}\right), \tag{4.9}$$

where $\mathcal{N}(x \mid \mu, \sigma^2)$ denotes the Gaussian density evaluated at x with mean μ and variance σ^2. This formulation preserves the strict positivity of d_i as a model parameter while naturally accommodating negative observed parallaxes, which arise routinely when the true parallax is small relative to the measurement uncertainty (Astraatmadja and Bailer-Jones, 2016).

Bailer-Jones (2015) demonstrated that as σ_{ϖ}/ϖ approaches and exceeds ~ 0.2, the probability density of the implied distance develops a pronounced heavy tail toward infinity, and the expectation value $\mathrm{E}[d]$ becomes formally undefined. Under these conditions, reporting distances as simple reciprocals with symmetric error bars is not merely imprecise but statistically ill-defined. The correct response is to work entirely within the forward-model framework of Eq. (4.9), treating the true distance as a latent parameter to be inferred rather than a quantity to be computed by inversion.

A direct consequence is that negative observed parallaxes carry genuine constraining power and must not be discarded. Within Eq. (4.9), a negative $\varpi_{\mathrm{obs},i}$ shifts the likelihood toward large d_i, providing an informative upper constraint on proximity. Discarding such sources truncates the noise distribution at zero, inducing an irreversible selection bias that systematically compresses the distance scale of the

remaining sample (Luri et al., 2018).

When proper motions are incorporated to form a kinegeometric distance estimate, the likelihood generalises to a multivariate normal over the vector $(\varpi_{\rm obs},\, \mu_{\alpha*,{\rm obs}},\, \mu_{\delta,{\rm obs}})$ with the full 3×3 covariance sub-matrix linking parallax and proper motion errors. Because these errors are structurally correlated by the scanning geometry, the multivariate formulation allows a well-measured proper motion component to restrict the permissible parameter space of a poorly measured parallax, providing geometric leverage that is entirely lost in a univariate treatment (Bailer-Jones, 2023).

4.3.2 Prior specification for stellar distances

Because the parallax likelihood becomes broad and uninformative when σ_ϖ/ϖ is large, the prior distribution over distance d plays a consequential—and in the low signal-to-noise regime, dominant—role in shaping the posterior. As Gelman, Simpson, and Betancourt (2017) emphasised, a uniform prior in distance is particularly hazardous: because the volume of a spherical shell grows as d^2, a flat distance prior assigns overwhelming weight to large distances, dragging noisy posteriors toward unphysical extremes.

The Exponentially Decreasing Space Density (EDSD) prior, introduced by Bailer-Jones (2015), provides the minimal viable complexity for regularising the distance posterior. We write it as

$$p(d \mid L) = \begin{cases} \dfrac{1}{2\,L^3}\, d^2\, \exp\!\left(-\dfrac{d}{L}\right), & d > 0, \\ 0, & d \leq 0, \end{cases} \tag{4.10}$$

where L is a length-scale parameter that sets the mode of the distribution at $d_{\rm mode} = 2L$. The d^2 factor encodes the geometric volume element, while the

exponential decay prevents divergence toward infinity. The choice of L must reflect the characteristic distance regime of the population under study; for NGC 6383 at $d \approx 1.1$ kpc, we adopt L in the range 1.0–1.5 kpc in the COSMIC pipeline, with explicit sensitivity analysis across this interval.

Direction-dependent priors provide more realistic regularisation by replacing the isotropic EDSD with a Generalised Gamma Distribution whose shape parameters and length scale vary across the sky. Bailer-Jones et al. (2021) calibrated these parameters via mock stellar catalogues derived from the Besançon Galaxy model, producing the geometric and photogeometric distance catalogues for Gaia DR3. A third class of prior augments the geometric spatial information with photometric constraints, mapping the apparent colour of a star to a theoretical absolute magnitude via isochrones to form a distance modulus prior (Anderson et al., 2018). While powerful for breaking astrometric degeneracies, photogeometric priors introduce sensitivity to the adopted extinction law, metallicity, and evolutionary tracks—assumptions that carry substantial systematic risk for young clusters with poorly calibrated pre-main-sequence isochrones (Section 4.5).

The regime-dependence of the prior's influence is the central message of this subsection. For bright sources with $\sigma_\varpi/\varpi \ll 0.1$, the likelihood is sharp and the posterior is effectively prior-independent. For faint sources with $\sigma_\varpi/\varpi \gtrsim 0.2$, the prior dominates. Any distance estimate in this regime must be explicitly stated as prior-dependent rather than data-driven, a requirement already established in Chapter 3.

4.3.3 Individual stellar distance posteriors

Combining the parallax likelihood of Eq. (4.9) with the EDSD prior of Eq. (4.10), the posterior for the distance to an individual star takes the form

$$p(d_i \mid \varpi_{\mathrm{obs},i},\, \sigma_{\varpi,i},\, L) \;\propto\; \mathcal{N}\!\left(\varpi_{\mathrm{obs},i} \;\middle|\; \frac{1}{d_i},\, \sigma^2_{\varpi,i}\right) \frac{d_i^2}{2\,L^3}\, \exp\!\left(-\frac{d_i}{L}\right). \tag{4.11}$$

This posterior is generally unimodal but asymmetric, with a sharper decline toward short distances and a heavier tail toward large distances inherited from the d^2 volume factor. The appropriate summary is the maximum a posteriori (MAP) estimate accompanied by asymmetric highest posterior density (HPD) intervals (Gelman et al., 2013). Reporting the reciprocal of the observed parallax with symmetric error bars is not warranted under any signal-to-noise regime.

The classical Lutz–Kelker bias is formally absorbed within this framework. Because the spatial volume increases with distance, a magnitude-limited sample contains more true background stars scattering inward than foreground stars scattering outward, producing a net inward bias in the observed parallax distribution. As Stepanishchev and Bobylev (2013) showed, the d^2 term in the spatial prior natively corrects this geometric asymmetry, eliminating the need for post hoc analytical corrections provided the prior accurately reflects the underlying source density.

Prior sensitivity constitutes the principal diagnostic requirement. We assess this by repeating the inference under alternative priors—varying L by a factor of two, or comparing geometric versus photogeometric outputs. If the HPD interval shifts substantially between tests, the source is classified as prior-dominated and its distance enters downstream summaries as explicitly conditional on the adopted prior (Piironen and Vehtari, 2017). For NGC 6383, faint pre-main-sequence candidates at the cluster periphery have the largest σ_ϖ/ϖ and therefore the strongest prior dependence; this diagnostic governs the interpretability of the

recovered spatial structure.

4.3.4 Cluster-level hierarchical distance inference

Individual distance posteriors, however carefully constructed, do not exploit the physical constraint that the stars of an open cluster share a common spatial origin. We formalise this constraint through a hierarchical model in which the true parallax of each member i is drawn from a shared cluster-level distribution:

$$\varpi_{\mathrm{true},i} \sim \mathcal{N}\big(\bar{\varpi}_{\mathrm{cl}},\ \sigma^2_{\varpi,\mathrm{int}}\big)\,, \qquad \varpi_{\mathrm{obs},i} \mid \varpi_{\mathrm{true},i} \sim \mathcal{N}\big(\varpi_{\mathrm{true},i},\ \sigma^2_{\varpi,i}\big)\,, \tag{4.12}$$

where $\bar{\varpi}_{\mathrm{cl}}$ is the cluster mean parallax (corresponding to $d_{\mathrm{cl}} = 1/\bar{\varpi}_{\mathrm{cl}}$) and $\sigma_{\varpi,\mathrm{int}}$ encodes the intrinsic parallax dispersion arising from the physical line-of-sight depth of the system. This architecture enacts partial pooling: bright, high signal-to-noise members anchor the cluster-level hyperparameters, while faint, noisy members are shrunk toward the cluster mean, reducing the overall mean squared error relative to independent star-by-star inference (Morris and Lysy, 2012).

The explicit introduction of the intrinsic depth $\sigma_{\varpi,\mathrm{int}}$ as a hyperparameter is critical. Without it, the model enforces complete pooling—the assumption that all members reside at exactly the same distance. Under complete pooling, the true physical extent of the cluster, which may span several parsecs for a young system such as NGC 6383, is absorbed into the measurement error budget. This absorption violates the Gaussian error specification, inflates the effective noise variance, and produces artificially narrow credible intervals for $\bar{\varpi}_{\mathrm{cl}}$. Gelman (2006) showed that the zero-depth failure mode is formally identifiable: if the posterior for $\sigma_{\varpi,\mathrm{int}}$ concentrates against zero, the model cannot distinguish physical extent from noise, and the hierarchical inference is compromised.

The hierarchical framework finds its operational realisation in codes such as

Kalkayotl, which samples the joint posterior over individual distances and cluster-level hyperparameters simultaneously (Olivares et al., 2020). The COSMIC framework developed in this thesis approximates this architecture for NGC 6383, inheriting the partial-pooling structure while adapting the prior specification to the cluster's distance regime ($d \approx 1.1$ kpc) and the quality characteristics of the Gaia DR3 data. The most powerful extensions integrate astrometric and photometric constraints jointly, requiring the latent distances to be consistent with both the parallax data and the cluster isochrone (Stein, 2012). The formal machinery for the photometric side of this joint inference is developed in Section 4.5.

4.3.5 Astrometric covariance, systematic offsets, and perspective effects

The formal validity of the parallax-to-distance inference depends on the fidelity of the astrometric inputs entering Eq. (4.9). Gaia astrometry carries structured systematic effects that must be either corrected prior to inference or modelled as explicit components of the likelihood.

The parallax zero-point offset constitutes the most consequential systematic. Gaia parallaxes exhibit a spatially correlated bias relative to the quasar reference frame that varies with apparent magnitude, effective colour, and ecliptic latitude (Lindegren et al., 2021a). We absorb this offset into the likelihood by replacing the raw parallax with the corrected quantity,

$$\varpi_{\mathrm{corr},i} = \varpi_{\mathrm{obs},i} - \varpi_{\mathrm{zp}}(G_i,\, \nu_{\mathrm{eff},i},\, \beta_i), \tag{4.13}$$

where ϖ_{zp} is the zero-point correction evaluated as a function of the source's apparent G-band magnitude G_i, effective wavenumber $\nu_{\mathrm{eff},i}$, and ecliptic latitude β_i. For NGC 6383, whose members span a wide range in G and colour, this

correction is source-dependent rather than a single scalar. A systematic offset of ~ 20 μas, while negligible for nearby bright stars, produces catastrophic distance inflation for populations whose true parallax is comparable in magnitude to the bias. We explicitly differentiate between sources receiving five-parameter astrometric solutions—which utilise true colour information—and those receiving six-parameter solutions with pseudo-colour, as the residual systematics differ qualitatively between the two solution types (Fabricius et al., 2021).

The astrometric covariance structure provides a second layer of complexity. The formal parallax error varies with apparent magnitude, colour, and the number of transits accumulated under the scanning law at the source's ecliptic coordinates. The five astrometric parameters are not independently determined; the covariance between parallax and proper motion reflects the geometric coupling imposed by the scanning geometry (Lindegren et al., 2018). A likelihood that ignores this covariance discards geometric information that can sharpen the distance posterior, as discussed in Section 4.3.1.

For a spatially extended cluster at $d \approx 1$ kpc, the projection of the three-dimensional space velocity onto the celestial sphere introduces a systematic proper-motion gradient across the cluster's angular extent. This perspective effect arises purely from viewing geometry and produces apparent kinematic structure that has no dynamical origin (G. van de Ven et al., 2006). If the hierarchical model of Eq. (4.12) does not forward-model this projection, the perspective-induced gradients will be misinterpreted as intrinsic velocity dispersion or line-of-sight depth.

Unresolved binary companions inject a further source of astrometric contamination. Orbital motion displaces the photocentre from the primary, inflating the single-star astrometric residuals and manifesting as elevated RUWE values in the Gaia catalogue. If the hierarchical likelihood assumes purely Gaussian measurement

errors, these non-Gaussian residuals exert disproportionate leverage on the cluster posterior. Robust treatment requires either replacing the Gaussian parallax error in Eq. (4.9) with a heavy-tailed distribution or implementing explicit variance-inflation factors (Penoyre, Belokurov, and Evans, 2022a). Spatial correlations in the parallax residuals across the cluster field, arising from unmodelled attitude or calibration effects, represent an additional covariance component that may require explicit modelling for nearby, angularly extended systems (Vasiliev and Baumgardt, 2021).

4.3.6 Distance uncertainty propagation into physical quantities

The distance posterior does not terminate the inferential chain; it initiates it. Absolute magnitudes scale logarithmically with distance, so a modest fractional distance uncertainty translates into substantial uncertainty in luminosity and, through the mass–luminosity relation, in inferred stellar mass (Luri et al., 2018). Transverse velocities scale linearly with distance but inherit the full asymmetric structure of the distance posterior. Structural parameters—half-light radii, tidal radii, core radii—are physical lengths computed from angular scales and carry the distance uncertainty directly.

The interaction between distance uncertainty and survey selection compounds this propagation. The population recoverable at a given apparent magnitude limit depends on the true distance: a cluster farther than assumed will have lost its faintest members below the detection threshold. This coupling means that the inferred mass function and spatial density profile are conditioned not only on the distance posterior but also on the selection function evaluated at that distance (Everall and Boubert, 2021). Ignoring this coupling produces systematically biased

physical summaries.

Hard parallax-based membership cuts further bias the recovered population. Selecting stars within a fixed number of σ_{ϖ} of the cluster mean parallax preferentially retains bright, well-measured core members and rejects faint, peripheral candidates. This filtering does not merely reduce sample size; it systematically suppresses the extended halo and tidal features whose characterisation is a primary objective of this thesis. The formal treatment of membership as a probabilistic, uncertain quantity—developed in Section 4.2—is therefore deeply coupled to the distance inference.

The formal claim that closes this section is that every physical quantity reported for NGC 6383 inherits the full distance uncertainty structure. Cluster-level quantities—the mean distance, velocity dispersion, age, mass—must be reported as expectations over the joint posterior of the distance and population parameters, not as values computed at a fixed-distance point estimate. Where the posterior is prior-dominated for a substantial fraction of the membership, the resulting summaries must be stated as conditional on the adopted distance prior. The interpretive domain of any physical result is therefore bounded by the uncertainty structure of the distance inference on which it depends.

4.4 Density structure and clustering formalism

The identification of a young open cluster within a dense Galactic field requires a formal framework for separating persistent overdensities from stochastic background fluctuations. This section establishes the density-based clustering formalism that serves as the first stage of the COSMIC membership pipeline. We frame clustering as a non-parametric density estimation problem, develop the HDBSCAN algorithm as a hierarchical topological approach that avoids the rigid

global thresholds of earlier methods, contrast its outputs with those of parametric Gaussian mixture models, characterise the types of structure that the formalism can and cannot recover, and identify the failure modes that must be controlled.

4.4.1 Clustering as non-parametric density level-set estimation

We frame the cluster detection problem as the recovery of topologically connected components of the super-level sets of an unknown probability density $f(\boldsymbol{x})$. For a density threshold λ, a cluster is defined as a maximal connected subset of the level set $\{ \boldsymbol{x} : f(\boldsymbol{x}) \geq \lambda \}$. This formulation casts the historically heuristic task of finding groups into a mathematically well-posed problem: mapping the topological structure of a density manifold (Hartigan, 1975).

Topological persistence theory provides the formal guarantee that hierarchical density-level-set algorithms recover the true homology of the underlying manifold as the sample size grows. Chazal et al. (2013) demonstrated that the persistence of topological features across varying density thresholds quantifies cluster significance, providing a rigorous framework for distinguishing genuine stellar populations from stochastic field fluctuations. In practice, the continuous density f is approximated from discrete data through k-nearest-neighbour density estimates, whose adaptive bandwidth naturally accommodates the varying local densities characteristic of young cluster environments (Wasserman, 2004).

4.4.2 HDBSCAN: mutual reachability, condensation, and persistence

HDBSCAN unites the level-set formalism of Hartigan (1975) with the practical execution of density-based clustering by systematically exploring all possible

density thresholds rather than requiring a single global value. The algorithm proceeds in three stages: space transformation, hierarchical construction, and cluster extraction (Campello, Moulavi, and Sander, 2013; Campello et al., 2015).

The space transformation replaces the standard pairwise distance $d(\boldsymbol{x}_i, \boldsymbol{x}_j)$ with the mutual reachability distance,

$$d_{\text{mreach}}(\boldsymbol{x}_i,\, \boldsymbol{x}_j) = \max\big\{\, d_{\text{core}}(\boldsymbol{x}_i),\; d_{\text{core}}(\boldsymbol{x}_j),\; d(\boldsymbol{x}_i, \boldsymbol{x}_j) \,\big\}, \tag{4.14}$$

where $d_{\text{core}}(\boldsymbol{x}_i)$ is the distance from point i to its kth nearest neighbour and $k = m_{\text{pts}}$ is the smoothing parameter. This transformation lowers the effective density of sparse regions while preserving the relative geometry of dense cores, ensuring that the subsequent minimum spanning tree does not spuriously chain across low-density noise bridges (McInnes, Healy, and Astels, 2017).

From the minimum spanning tree constructed in mutual-reachability space, HDBSCAN builds a hierarchical cluster dendrogram by progressively removing edges in order of decreasing weight, equivalent to sweeping the density parameter $\lambda = 1/d_{\text{mreach}}$ from low to high. The condensed tree is then extracted by enforcing a minimum cluster size m_{cl}: only components containing at least m_{cl} points persist as candidate clusters, while smaller fragments are classified as noise. Cluster extraction proceeds by maximising the total excess of mass across λ, selecting the components whose members persist over the widest density range (Campello et al., 2015).

The soft membership values produced by HDBSCAN are λ-persistence ratios that normalise each point's fallout density between the cluster's birth and death thresholds:

$$p_{\text{HDBSCAN},i} = \frac{\lambda_i - \lambda_{\text{birth}}}{\lambda_{\text{death}} - \lambda_{\text{birth}}}, \tag{4.15}$$

where λ_i is the density parameter at which point i separates from its parent

cluster, λ_{birth} is the threshold at which the cluster first appears, and λ_{death} is the threshold at which it merges or dissolves. These values measure relative spatial centrality within the density structure. They are *not* posterior probabilities of generative origin—a doctrinal distinction developed in Section 4.2.2.

In the COSMIC framework, we construct a composite ranking metric that combines the iterative stability of each source across HDBSCAN hyperparameter sweeps with the persistence metric of Eq. (4.15). We define the *pseudoprobability* for source i as

$$\tilde{p}_i \;=\; f_i \;\times\; p_{\text{HDBSCAN},i}\,, \tag{4.16}$$

where $f_i \in [0,1]$ is the fraction of HDBSCAN runs (across a grid of minimum cluster sizes) in which source i was assigned to the cluster component. The pseudoprobability $\tilde{p}_i$ rewards both spatial centrality (through persistence) and recovery robustness (through the iteration fraction). It is *not* a calibrated posterior probability: it carries no information about the field distribution, no generative likelihood, and no propagation of parameter uncertainty. The calibrated posterior membership of Eq. (4.7) and the pseudoprobability of Eq. (4.16) are therefore formally distinct objects that must never be conflated. Throughout this thesis, we use $\tilde{p}_i$ as an operational ranking for threshold-based sample definition, while reserving the posterior marginal $p(z_i = \text{cl} \mid \boldsymbol{y}_i, \boldsymbol{\theta})$ for any inference that requires probabilistic calibration.

4.4.3 HDBSCAN vs. Gaussian mixture models: complementary partitioning logics

HDBSCAN and Gaussian mixture models (GMMs) encode fundamentally different definitions of what constitutes a cluster. HDBSCAN identifies persistent topological components of the density manifold without assuming a parametric

form for the component distributions. GMMs assume that the data are generated by a finite number of multivariate Gaussian components and recover clusters as the posterior-weighted domains of each component (McLachlan and Peel, 2000).

These differences are not merely algorithmic; they produce materially different outputs. HDBSCAN recovers arbitrarily shaped structures tied to the density landscape but produces membership values that lack probabilistic calibration. GMMs produce genuine posterior marginal probabilities through Eq. (4.7) but can only represent ellipsoidal structures, struggling with elongated tidal features or fractal sub-populations. In the young open cluster regime, where the cluster core may be compact and roughly Gaussian while the extended halo follows an irregular, substructured morphology, neither method alone captures the full structure.

The hybrid HDBSCAN $\rightarrow$ GMM paradigm, now standard in Gaia-era cluster science, exploits the complementarity. HDBSCAN identifies high-confidence cores without shape assumptions; its labels initialise and constrain a subsequent GMM that refines the membership probabilities into calibrated posteriors (Noormohammadi, Khakian Ghomi, and Javadi, 2024; Cantat-Gaudin et al., 2018). The COSMIC framework adopts this architecture for NGC 6383, using HDBSCAN as a non-parametric first pass and a covariance-aware GMM for probabilistic refinement.

4.4.4 Recoverable structure: dimensionality, metric, and resolution limits

The structure that any density-based method can recover is bounded by the dimensionality and metric of the feature space, the survey selection function, and the smoothing parameters. In Gaia astrometric space, the relevant features—sky position, parallax, and proper motion—live on different units and geometric

conventions. Without a representation that reflects unequal uncertainty, faint sources with large error ellipsoids are artificially displaced from compact cores, and apparent separations between subpopulations may be generated by the metric rather than by intrinsic structure (Hunt and Reffert, 2023).

Hunt and Reffert (2020) demonstrated that the choice of feature scaling, Mahalanobis versus Euclidean distance, and the inclusion or exclusion of photometric dimensions materially alter which clusters are recovered and at what significance level. Covariance-aware distances provide a partial correction by reshaping the geometry to respect Gaia measurement uncertainties, but they transfer the assumption load from the clustering algorithm to the covariance model. The resolution limits imposed by m_{pts} and m_{cl} further restrict recoverability: low-mass, sparse clusters whose membership falls below m_{cl} are structurally invisible to HDBSCAN regardless of their physical reality (Kounkel and Covey, 2019).

For NGC 6383, these limits are especially relevant in the cluster outskirts, where membership density declines, measurement uncertainties increase, and the transition between cluster and field becomes diffuse. The formal machinery cannot extract structure that the data do not support at the adopted smoothing scale.

4.4.5 Failure modes, diagnostics, and boundary of applicability

The principal failure modes of density-based clustering in the open cluster regime are chaining, fragmentation, and noise-bridge contamination. Chaining occurs when a narrow filament of background sources connects two genuinely distinct structures, producing a single spurious merged cluster. Fragmentation occurs when a genuine cluster with internal density variation is split into multiple components because the local minima exceed the extraction threshold. Noise bridges arise

when the mutual-reachability transformation is insufficient to suppress low-density connections between a cluster and a nearby field overdensity.

Each failure mode has a diagnostic. Chaining is detectable through visual inspection of the condensed tree and through the persistence spectrum: a chained component typically shows a narrow persistence interval relative to its constituent fragments. Fragmentation is identifiable when the condensed tree splits a component at a λ value that is inconsistent with independent evidence—such as a coherent CMD locus or a continuous proper-motion distribution. Noise bridges produce members with low λ-persistence values concentrated at the boundary between components.

For the specific case of NGC 6383, the most consequential failure mode is the inability of HDBSCAN to separate kinematically superimposed populations that share similar astrometric properties but differ in age, chemical composition, or evolutionary state. Such populations appear as a single density peak in astrometric space, and their decomposition requires the generative mixture-model machinery of Section 4.2 operating in the photometric domain rather than density-based partitioning alone. This limitation reinforces the hybrid architecture adopted in the COSMIC framework: HDBSCAN provides the topological scaffolding, but the physical decomposition is completed by the probabilistic mixture. The specific hyperparameter choices, iteration strategy, and diagnostic runs applied to NGC 6383 are documented in Section 6.2.

4.5 Photometric and isochronal inference

The extraction of fundamental cluster parameters—age τ, metallicity Z, and extinction A_V—proceeds through the comparison of observed photometry with theoretical predictions generated by stellar evolution models. For a young system

such as NGC 6383, this comparison is mediated by a highly non-linear forward model subject to severe internal degeneracies and astrophysical complications that are especially acute in the pre-main-sequence regime. This section establishes the formal framework for photometric and isochronal inference, beginning with the definition of isochrones as parametric generative families, constructing the colour–magnitude diagram (CMD) likelihood under realistic systematics, mapping the degeneracy structure, cataloguing the PMS complications, addressing multiplicity and differential extinction, establishing formal claim bounds through model-discrepancy analysis, and closing the chapter with the inferential interface to downstream application.

4.5.1 Isochrones as parametric families of stellar evolution tracks

The classical methodology for determining open cluster properties relied on visual superposition of theoretical isochrones onto observed CMDs, adjusting distance modulus, reddening, and age by hand. This approach is fundamentally ill-posed as a statistical procedure: it cannot capture parameter covariance, ignores non-linear uncertainty propagation, and provides no mechanism for marginalising over latent variables such as the binary fraction or contamination (Hippel et al., 2006). We replace this heuristic matching with Bayesian forward modelling, in which individual stellar masses are drawn from a specified initial mass function, binary companions are assigned according to an empirical mass-ratio distribution, bandpass-specific extinction is applied through a reddening law, and heteroscedastic photometric noise is injected. The resulting synthetic CMD is evaluated against the data through a formal likelihood, yielding a joint posterior that maps all parameter degeneracies.

The isochrone grids underpinning this forward model—MIST, PARSEC, the Baraffe models, the spotted SPOTS grids—employ distinct treatments of convective boundary mixing, atmospheric boundary conditions, and opacity tables (Choi et al., 2016; Baraffe et al., 2015; Somers, Cao, and Pinsonneault, 2020). In the pre-main-sequence regime, fitting the same population with different grids produces systematic age offsets exceeding the statistical precision of any single fit. The grid choice is itself a modelling assumption whose consequences must propagate into the uncertainty budget, either by treating the grid as a discrete model-selection variable or by reporting the grid-to-grid variance as a systematic floor (Section 4.5.6).

Automated engines such as BASE-9 and ASteCA implement this forward-modelling architecture computationally (Perren, Vázquez, and Piatti, 2015). The COSMIC framework builds on ASteCA for NGC 6383, inheriting the forward-modelling logic while adapting the likelihood and prior specification to Gaia-era photometric characteristics. Modern implementations increasingly employ neural network emulators to interpolate across isochrone grids, avoiding the catastrophic numerical noise generated by linear interpolation across sparse grid nodes and preserving the non-linear topological features of stellar evolution (Hon, Li, and Ong, 2024).

4.5.2 CMD likelihood under photometric uncertainty and systematic effects

For a star i with observed multi-band photometry $\boldsymbol{m}_{\mathrm{obs},i}$ and photometric covariance $\boldsymbol{C}_{\mathrm{phot},i}$, we write the CMD likelihood under the cluster model as

$$p(\boldsymbol{m}_{\mathrm{obs},i} \mid M_i,\, \tau,\, Z,\, A_V,\, d_i) = \mathcal{N}(\boldsymbol{m}_{\mathrm{obs},i} \mid \boldsymbol{m}_{\mathrm{iso}}(M_i,\, \tau,\, Z) + \boldsymbol{A}(A_V) + \mu(d_i), \boldsymbol{C}_{\mathrm{phot},i})\,, \tag{4.17}$$

where $\boldsymbol{m}_{\mathrm{iso}}(M_i, \tau, Z)$ is the absolute magnitude vector predicted by the isochrone for initial mass M_i, age τ, and metallicity Z; $\boldsymbol{A}(A_V)$ is the bandpass-dependent extinction vector; and $\mu(d_i) = 5\log_{10}(d_i/10\,\mathrm{pc})$ is the distance modulus, linked to the parallax-to-distance inference of Section 4.3.

The most consequential photometric systematic in the Gaia bandpass system is the Forbes effect: the broad G band ($\sim$330–1050 nm) has an effective wavelength that shifts with the source's spectral energy distribution, causing the extinction coefficient to vary as a function of intrinsic colour and total extinction (Riello et al., 2021). We account for this by integrating synthetic stellar spectra over the specific Gaia passband response curves, producing temperature-dependent and extinction-dependent bolometric corrections (Fitzpatrick et al., 2019). The total-to-selective extinction ratio R_V must either be treated as a free parameter or informed by external three-dimensional dust maps, since the classical assumption of a universal value is inappropriate for young clusters in heterogeneous interstellar environments(Fitzpatrick et al., 2019).

Differential reddening—spatially variable extinction across the cluster field—broadens the main sequence perpendicular to the isochrone ridge line. If un-

modelled, the likelihood absorbs this broadening by inflating the binary fraction or interpreting the dispersion as an extended star-formation history (Pancino et al., 2024). Spatial extinction mapping through Gaussian process models can recover the intrinsic isochronal sequence (Edenhofer et al., 2024). Photometric zero-point biases, which vary with magnitude, colour, and celestial position, must be corrected in concert with the parallax zero-point of Eq. (4.13) to prevent the sampler from compensating for an astrometric systematic by adjusting the cluster age or extinction (Lindegren et al., 2021a). Photometric variability from rotational modulation and accretion bursting must be modelled as additive jitter in the likelihood rather than interpreted as a physical age spread (Naylor and Jeffries, 2006).

The mixture formulation of Eq. (4.6) applies equally in the photometric domain, allowing the CMD likelihood to marginalise over field-star contamination without hard membership cuts (Perren, Vázquez, and Piatti, 2015).

4.5.3 The age–metallicity–extinction degeneracy structure

The fundamental challenge of broad-band CMD fitting is the near-alignment of the perturbation vectors associated with age, metallicity, and extinction. Increasing age cools and dims the turn-off; increasing metallicity increases line blanketing; increasing extinction dims and reddens the photometry. Without wide wavelength coverage, these transformations constitute a nearly degenerate affine mapping that produces strongly correlated posteriors (Naylor and Jeffries, 2006).

The information-geometric interpretation is that the Fisher Information Matrix (FIM) of the CMD likelihood becomes near-singular along specific directions. Wieland et al. (2021) showed that eigenvalues approaching zero imply divergent Cramér–Rao bounds along the corresponding eigenvectors, defining degeneracy

manifolds where the likelihood is effectively flat. Standard Monte Carlo approximations of the FIM are susceptible to numerical noise that artificially inflates eigenvalues and produces overconfident assessments of constraining power.

Breaking these degeneracies requires orthogonal data that steepen the likelihood along the previously flat directions. The hierarchical distance posterior of Section 4.3 constrains the distance modulus directly, collapsing a major dimension of the covariant error structure. Near-infrared photometry extends the wavelength baseline, separating the reddening vector from the temperature and metallicity vectors (Cardelli, Clayton, and Mathis, 1989). Gaia XP spectrophotometry provides metallicity estimates whose injection as informative priors collapses the age–metallicity covariance, enabling precise age determinations even for sparse clusters (Huang et al., 2025; Nizovkina et al., 2025). The morphology of distinct evolutionary phases can also break degeneracies: the nearly vertical Hayashi tracks of PMS stars cross the reddening vector at steep angles, aiding age–extinction separation (Castro-Ginard et al., 2022).

4.5.4 Pre-main-sequence complications and formal inference limits

For a cluster with $\tau \sim$ few Myr, the majority of the membership populates the pre-main-sequence. Isochronal inference in this regime faces physical complications that impose hard limits on the achievable accuracy.

The most fundamental is the invalidity of the classical stellar birthline. Krumholz, McKee, and Bland-Hawthorn (2019) demonstrated that the initial entropy of a PMS star depends on whether early accretion was predominantly hot or cold, introducing a stochastic, accretion-history-dependent starting condition. Episodic accretion bursts of the FU Orionis type alter the thermal structure, and subsequent

contraction deviates strongly from standard tracks. A strictly coeval population with heterogeneous accretion histories can produce an apparent age spread of several Myr without genuine temporal extension of star formation (Baraffe, Chabrier, and Gallardo, 2009).

Magnetic activity compounds the problem. PMS dynamos suppress convective energy transport, forcing structural expansion and shifting evolutionary tracks to cooler $T_{\rm eff}$ and redder colours (Feiden, 2016). Heavy starspot coverage, parameterised as a filling factor in the SPOTS grids, further amplifies this shift; fitting spotted grids resolves the spurious mass-dependent age trends that appear with standard unspotted models (Somers, Cao, and Pinsonneault, 2020).

Circumstellar disks introduce inclination-dependent extinction, near-infrared excess from warm dust emission, and blue veiling from accretion shocks (Vorobyov and Basu, 2015; Aerts, Mathis, and Rogers, 2019). These effects are stochastic, orientation-dependent, and time-variable, making them extremely difficult to model deterministically.

Cross-validation against independent indicators provides the most robust external check. The lithium depletion boundary yields a nearly model-independent age calibration for clusters in the 20–200 Myr range; direct comparisons show that standard non-magnetic models yield ages significantly younger than LDB ages, and that achieving concordance requires magnetically active stellar models (Binks et al., 2021; Jeffries et al., 2023). For NGC 6383, whose $\tau \sim$ few Myr places it below the LDB range, the pattern—that standard models underestimate ages for active populations—remains directly relevant.

4.5.5 Unresolved multiplicity and differential extinction

The CMD of a young cluster is broadened by at least three physically distinct mechanisms: genuine age spread, unresolved multiplicity, and differential extinction. Disentangling these is essential for recovering intrinsic cluster properties.

An unresolved equal-mass binary appears ~ 0.75 mag brighter than a single star of the same primary mass, placing it in the CMD region occupied by genuinely younger single stars. The binary locus is bounded: unequal mass ratios produce smaller vertical displacements. We forward-model the binary mass-ratio distribution and binary fraction as explicit parameters, allowing marginalisation over the multiplicity (Bodensteiner, Shenar, and Sana, 2020).

Differential reddening broadens the sequence along the reddening vector, which is geometrically distinct from the vertical binary displacement in well-calibrated multi-band photometry. Spatial correlation provides a diagnostic: differential reddening is coherent across angular scales set by the cloud structure, while binary scatter is spatially uncorrelated (Naylor and Jeffries, 2006).

Kinematic interlopers from older foreground and background associations contaminate the CMD above the zero-age main sequence, mimicking a PMS population (Hunt and Reffert, 2024). The mixture structure of Eq. (4.6) provides the formal mechanism for handling this contamination.

The three-way confusion imposes a practical identifiability limit. The forward model must treat binary fraction, differential reddening amplitude, and age dispersion as distinct but potentially degenerate parameters, reporting their joint posterior. Where strong covariance appears, the resulting age estimate must be stated as conditional on the treatment of multiplicity and extinction.

4.5.6 Formal claim bounds on derived astrophysical parameters

The complications catalogued above establish that isochronal precision is limited by systematic uncertainties in stellar physics, the extinction environment, and population composition. We quantify these through model-discrepancy analysis. Following Loredo and Wolpert (2024), we augment the generative equation with a non-parametric discrepancy function $\delta(\boldsymbol{\theta})$:

$$\boldsymbol{m}_{\mathrm{pred},i} = \boldsymbol{m}_{\mathrm{iso}}(M_i,\, \tau,\, Z) + \delta(\boldsymbol{\theta}) + \boldsymbol{\varepsilon}_i\,, \tag{4.18}$$

where $\delta(\boldsymbol{\theta})$ absorbs the epistemic theoretical uncertainties—missing physics, numerical instabilities—that persist irrespective of data quality. Marginalising over δ explicitly widens the posterior bounds, ensuring that inferred ages reflect the structural limitations of the underlying physics rather than collapsing to artificially precise estimates. The framework of Gaussian Process regression over the discrepancy function was introduced by Kennedy and O'Hagan (2001) for Bayesian calibration of computer models, and has since been adopted in astrophysics to quantify the epistemic floor imposed by incomplete physics in stellar evolutionary tracks.

The parameters governing internal mixing—convective core overshooting, rotational mixing, microscopic diffusion—are among the most consequential sources of systematic uncertainty. We treat them as nuisance hyperparameters with broad, physically motivated priors, marginalising over them during sampling (Aerts et al., 2018; Cummings et al., 2018). The grid-to-grid variance provides a complementary measure: running the full pipeline with MIST, PARSEC, and SPOTS independently and reporting the variance among the MAP estimates exposes the epistemic systematic floor (Schneider et al., 2013). This spread is

not a statistical uncertainty; it defines the hard floor below which additional photometric data cannot improve the age determination without improving the underlying stellar models.

Reporting a single maximum-likelihood age for a young cluster is therefore scientifically inadequate. We report the full marginalised posterior, using the median and asymmetric HPD intervals. The systematic contributions—from model discrepancy, grid dependence, and magnetic activity—are reported alongside or integrated into the statistical uncertainty (Schneider et al., 2013).

4.5.7 Formal specification of the inferential interface with downstream application

The interface between the formal machinery of this chapter and its downstream application is governed by three constraints. First, every physical quantity reported in subsequent chapters is a posterior expectation over the joint distribution of model parameters and uncertain membership, as formalised in Eq. (4.8). Second, where the posterior for a given parameter is prior-dominated, the resulting claims must be stated as conditional on the adopted prior, and their sensitivity to alternatives must be documented. Third, where the model is inadequate—as indicated by posterior predictive discrepancies, grid-to-grid variance exceeding statistical precision, or unresolved degeneracies—the resulting claims must be reported as local descriptive summaries rather than global physical truths.

These constraints instantiate the claim-status framework of Chapter 3. The formal machinery of this chapter provides the tools for conducting inference rigorously; the claim-status discipline governs what may be concluded from its outputs. The application of these tools to NGC 6383 begins in Chapter 5 with the data assembly, continues in Chapter 6 with the COSMIC implementation, and culminates in

Chapter 7 with the physical characterisation of the cluster.

Chapter 5

Observational Data and Systematics

The formal machinery of Chapter 4 operates on observational inputs whose quality, completeness, and systematic structure bound the validity of every downstream physical claim. As established in Section 4.1.1, the inference begins from a generative measurement model that separates the true astrophysical state from its noisy observational trace; the dataset documented in this chapter constitutes that observational trace for the analysis of NGC 6383. We describe the Gaia DR3 astrometric and photometric data and the rationale for the query strategy (Section 5.1), the systematic corrections applied to the raw catalogue quantities (Section 5.2), the crossmatch with the Two Micron All Sky Survey that extends the photometric baseline into the near-infrared (Section 5.3), the cumulative quality filters that define the working sample entering the COSMIC pipeline (Section 5.4), and the residual observational biases that condition the domain of interpretive validity for all physical claims reported in Chapter 7 (Section 5.5).

5.1 Gaia DR3 astrometric and photometric data

5.1.1 The Gaia DR3 data model

The third Gaia Data Release (Gaia Collaboration et al., 2023) provides the most comprehensive astrometric and photometric survey available for the analysis of Galactic stellar populations. For each source, the Gaia astrometric pipeline delivers a five-parameter solution: two sky coordinates (α, δ) at the reference epoch J2016.0, the trigonometric parallax ϖ, and the two components of proper motion $(\mu_{\alpha*}, \mu_\delta)$, where $\mu_{\alpha*} \equiv \mu_\alpha \cos\delta$ absorbs the cosδ projection factor. The five parameters are not independent; the iterative least-squares solution that fits the along-scan observations to the astrometric model produces a full 5×5 covariance matrix $\boldsymbol{C}_i$ for each source, encoding the correlations imposed by the satellite's scanning law and the celestial position of the source (Lindegren et al., 2021b). These off-diagonal terms are not negligible and must enter any likelihood that claims to use the Gaia astrometric information correctly, as established in Section 4.1.2.

The formal uncertainties on ϖ and μ are strongly heteroscedastic: they depend on apparent magnitude, colour, ecliptic latitude, and the number of field-of-view transits accumulated over the 34-month observing baseline of DR3 (Gaia Collaboration et al., 2023). At the bright end ($G \lesssim 14$ mag), parallax uncertainties are $\sigma_\varpi \sim 0.01$–0.03 mas, yielding fractional errors below 5% at the distance of NGC 6383 ($\varpi \approx 0.9$ mas). At the faint end ($G \sim 20$ mag), uncertainties rise to $\sigma_\varpi \sim 0.5$–1.0 mas, placing faint members firmly in the regime where the prior dominates the distance posterior (Section 4.1.5).

Gaia DR3 also provides integrated photometry in three broadband filters: the white-light G band ($\sim$ 330–1050 nm), the blue photometer $G_{\rm BP}$ ($\sim$ 330–680 nm), and the red photometer $G_{\rm RP}$ ($\sim$ 630–1050 nm). The broad wavelength range of the

G band introduces a colour-dependent effective wavelength that must be accounted for in the extinction treatment (Section 4.5.2). Photometric uncertainties are typically $\lesssim 1$ mmag for $G < 13$ and degrade to ~ 10 mmag at $G \approx 20$ (Riello et al., 2021). Radial velocities from the Radial Velocity Spectrometer (RVS) are available for a subset of brighter sources ($G_{\mathrm{RVS}} \lesssim 14$ mag), with precisions of ~ 1 km s^{-1} for FGK-type stars; coverage is sparse for the fainter, lower-mass membership and is discussed quantitatively in Section 5.5.

5.1.2 Query strategy and field definition

We queried the Gaia DR3 archive[1] using a cone search of 40.0 arcmin radius centred on the nominal coordinates of NGC 6383 ($\ell = 355.68°$, $b = 0.05°$; $\alpha = 263.68°$, $\delta = -32.58°$). The search radius was chosen to exceed the expected tidal radius of the cluster by a factor of ~ 1.5, based on the structural estimates of Piskunov et al. (2008) ($R_t = 29.0 \pm 6.6$ arcmin) and Hunt and Reffert (2024) ($R_t = 22.7$ arcmin). This generous angular aperture ensures the inclusion of potential escapers in the halo and tidal tails, whose dynamical significance for the assessment of the cluster's bound fraction and expansion state is discussed in Sections 2.4.2 and 8.3. The query returned 23 740 sources with at least a partial astrometric solution.[2]

5.1.3 Initial parallax pre-filter

We restricted the initial selection to sources satisfying the parallax window

$$0.750 \leq \varpi_{\mathrm{obs},i} \leq 1.100 \text{ mas}\,, \tag{5.1}$$

[1] `https://gea.esac.esa.int/archive/`

[2] The full ADQL query is included in the COSMIC repository (Section 9.5).

corresponding to a heliocentric distance interval of approximately 0.91–1.33 kpc under naive parallax inversion. This interval was informed by the preliminary analysis of Pulgar-Escobar et al. (2024), who established it as the plausible domain for cluster membership on the basis of the known distance literature and the DR3 parallax distribution of kinematically selected candidates. The window extends $\sim \pm 20\%$ beyond the expected cluster parallax of ~ 0.9 mas, accommodating both the intrinsic parallax dispersion of cluster members and the parallax zero-point offset of Section 5.2.1.

The pre-filter serves two purposes. First, it removes foreground and background contaminants whose parallaxes are incompatible with membership at any plausible cluster distance, reducing the computational cost of the subsequent HDBSCAN clustering step. Second, it prevents distant, highly uncertain sources with $\sigma_{\varpi}/\varpi \gg 1$ from entering the analysis, where their broad parallax posteriors would contribute negligible constraining power while introducing noise into the density structure of the proper-motion space. The formal membership inference operates within this pre-filtered sample and applies its own probabilistic selection criteria (Chapter 6).

5.2 Astrometric systematic corrections

The raw Gaia DR3 astrometric parameters carry systematic biases that, if uncorrected, propagate directly into the likelihood evaluations of Sections 4.3–4.5. Three corrections were applied before any inferential processing: a parallax zero-point correction, a proper-motion frame correction for bright sources, and an astrometric quality filter based on a machine-learned fidelity parameter. Each correction is documented below together with its formal justification and magnitude.

5.2.1 Parallax zero-point correction

The Gaia DR3 parallaxes exhibit a global zero-point offset $\varpi_{\rm zp}$ that is a complex function of source magnitude, colour, and ecliptic latitude. The offset arises from the coupling between the basic-angle variations of the satellite and the astrometric calibration model, and its structure was characterised by Lindegren et al. (2021a) using quasar observations. For the magnitude range and sky position of NGC 6383, the offset is typically $\varpi_{\rm zp} \sim -20$ to -30 μas, with the sign convention that the catalogue parallaxes are systematically too small. While small in absolute terms, this offset is non-negligible at the level of precision targeted by the hierarchical distance model of Section 4.3.4: an uncorrected bias of -25 μas shifts the inferred distance of a cluster at $\varpi \approx 0.9$ mas by $\sim$30 pc, comparable to the statistical uncertainty on the distance posterior.

We corrected the published parallaxes using the GAIADR3_ZEROPOINT Python package[3], which implements the five-parameter correction recipe of Lindegren et al. (2021a) as a function of ecliptic latitude β, effective wavenumber $\nu_{\rm eff}$, magnitude G, and solution type (five- or six-parameter). The corrected parallax for each source is

$$\varpi_{{\rm corr},i} = \varpi_{{\rm obs},i} - \varpi_{\rm zp}(G_i,\, \nu_{{\rm eff},i},\, \beta_i), \tag{5.2}$$

which enters the parallax likelihood of Eq. (4.9) in place of the raw catalogue value. The correction is applied source-by-source; its magnitude varies across the colour–magnitude distribution of the sample but remains within the range $|\varpi_{\rm zp}| \lesssim 40$ μas for the vast majority of sources in the NGC 6383 field. The effect of the correction is illustrated in Fig. 5.1, which shows the corrected versus observed parallax colour-coded by G magnitude, effective wavenumber $\nu_{\rm eff}$, pseudocolour, and ecliptic latitude. The correction systematically shifts the corrected parallaxes

[3] `https://pypi.org/project/gaiadr3-zeropoint/`

above the identity line, confirming the negative zero-point offset at the position of NGC 6383.

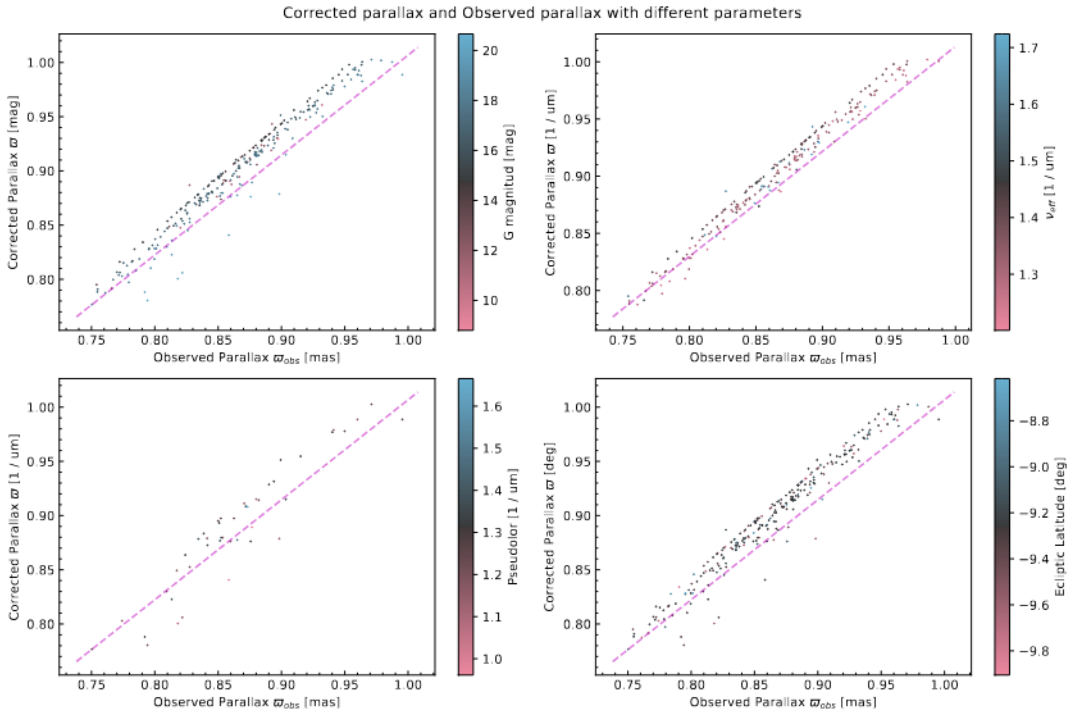

Figure 5.1: Corrected parallax $\varpi_{\rm corr}$ versus observed parallax $\varpi_{\rm obs}$ for the NGC 6383 membership sample, colour-coded by four parameters entering the zero-point correction of Eq. (5.2). *Upper left:* G-band magnitude. *Upper right:* effective wavenumber $\nu_{\rm eff}$. *Lower left:* pseudocolour. *Lower right:* ecliptic latitude β. The dashed magenta line marks the identity $\varpi_{\rm corr} = \varpi_{\rm obs}$. Points lying above the identity line indicate a positive correction (the raw parallax was underestimated), consistent with the negative zero-point offset characteristic of the Gaia DR3 astrometric calibration.

The zero-point correction addresses the dominant first-order systematic but does not eliminate all spatially correlated parallax errors. Small-scale correlations on angular scales of $\lesssim 1°$, arising from the scanning-law geometry and the attitude modelling residuals, persist in DR3 at the level of $\sim$ 10–15 μas (Lindegren et al., 2021b). For the angular extent of NGC 6383 ($\sim$40 arcmin $\approx 0.67°$), these correlated residuals are partially shared among cluster members and may produce a coherent shift in the cluster mean parallax. The hierarchical distance model of Section 4.3.4 absorbs such a shift into the population-level mean, but does

not explicitly model the spatial correlation structure. The residual systematic contribution to the distance uncertainty is estimated to be $\sim$10–20 pc; this floor is small relative to the statistical uncertainty ($\sim$60 pc) and is not propagated into the formal posterior but is noted as an additional systematic contribution to the distance error budget in Chapter 7.

A further caveat specific to Galactic-plane fields applies to NGC 6383. Ding et al. (2024) demonstrated that the Lindegren et al. (2021a) zero-point recipe exhibits sample dependence and does not apply effectively in the Galactic plane, where unmodelled chromatic effects and source-density-dependent biases produce residual offsets reaching ~ 10 μas in certain magnitude and colour ranges. At $b \approx 0.05°$, NGC 6383 falls squarely in the regime where this limitation is most acute. The same study provided a supplementary Galactic-plane correction that we do not apply here—the correction was published after the analysis pipeline was substantively complete, and its magnitude for the NGC 6383 field ($\lesssim 10$ μas) is small relative to the statistical uncertainty on the cluster mean parallax—but whose existence should be considered when interpreting the distance posterior of Section 7.2.1. Ding et al. (2025) further showed, using 44 binary orbital parallaxes, that the formal uncertainties on the Lindegren zero-point are underestimated by a factor of ~ 2, reinforcing the need to treat the parallax systematic floor conservatively.

5.2.2 Proper-motion frame correction for bright sources

Cantat-Gaudin and Brandt (2021) identified a systematic discrepancy between the proper motions of bright ($G < 13$ mag) and faint sources in Gaia DR3, arising from the coupling between the bright-star astrometric calibration and the quasar-based reference frame. The offset reaches up to 80 $\mu\mathrm{as\,yr}^{-1}$ for sources in the range $G = 11$–13 mag and is a function of magnitude, colour, and sky position.

For a young open cluster such as NGC 6383, the bright-star population includes the most massive main-sequence members, which anchor the kinematic model and dominate the determination of the mean proper motion. A systematic offset in their proper motions would bias the cluster mean and, more critically, would distort the proper-motion dispersion by introducing an artificial separation between bright and faint members. We applied the magnitude-dependent correction prescribed by Cantat-Gaudin and Brandt (2021) to all sources with $G = 11$–13 mag, ensuring that the bright and faint populations share a common reference frame before entering the kinematic model of Section 6.4.

The correction is small relative to the observed proper-motion dispersion of the cluster ($\sigma_\mu \sim 0.15$ mas yr^{-1}) but is comparable to the formal uncertainty on the mean proper motion ($\sigma_{\bar{\mu}} \sim 0.01$ mas yr^{-1}). Without the correction, the inferred mean proper motion would be biased at the $\sim 5\sigma$ level of its statistical precision, producing a systematic error that is invisible to the convergence diagnostics of the sampling algorithm.

5.2.3 Astrometric fidelity filter

Not all five-parameter solutions in Gaia DR3 are equally trustworthy. Sources in crowded fields, with close companions, with poor scan-angle coverage, or with anomalously large astrometric residuals can receive formally converged solutions that are nevertheless unreliable. The standard quality indicators provided in the Gaia catalogue—the Renormalised Unit Weight Error (RUWE), the astrometric excess noise, and the significance of the excess noise—each capture a subset of these failure modes, but no single indicator provides a comprehensive assessment.

Rybizki et al. (2022) addressed this limitation by training a neural network on 17 Gaia DR3 catalogue metrics—including RUWE, astrometric excess noise,

visibility periods, and colour excess—to produce a single continuous fidelity score $f_{\mathrm{fid},i} \in [0, 1]$ that summarises the overall reliability of each source's astrometric solution. The classifier was validated against external astrometric reference frames and parallax distributions in the Magellanic Clouds, achieving 99.3% purity at 97.3% completeness at the $f_{\mathrm{fid}} = 0.5$ threshold.

We retained all sources satisfying

$$f_{\mathrm{fid},i} > 0.5\,, \tag{5.3}$$

a threshold that balances completeness against reliability. This single cut subsumes the information content of individual cuts on RUWE, astrometric excess noise, and visibility periods (Rybizki et al., 2022). It removes 35% of the initial parallax-prefiltered sample, reducing it from 23 740 to 15 480 sources. The removed sources are predominantly faint ($G > 18$ mag), red, or located in the densest regions of the field where confusion and blending compromise the astrometric solution. The fidelity filter therefore acts as a spatially and photometrically non-uniform selection function whose consequences for the completeness of the recovered cluster population are assessed in Section 5.5.

An important caveat is that the fidelity filter does not provide a physical model of the failure mechanisms; it is a data-driven classifier that correlates observed indicators with solution quality. Sources that pass the filter are not guaranteed to have accurate astrometry—they are sources for which the neural network assigns a probability greater than 50% that the solution is reliable. For the subset of faint, red sources near the fidelity threshold, the marginal reliability of the astrometric solution should be borne in mind when interpreting their membership probabilities and distance posteriors.

5.3 Crossmatching with the Two Micron All Sky Survey

5.3.1 Rationale for near-infrared photometry

The Gaia broadband photometric system (G, $G_{\rm BP}$, $G_{\rm RP}$) spans the optical wavelength range from $\sim$330 to 1050 nm. While sufficient for constructing colour–magnitude diagrams and performing isochronal fits, the three optical bands sample a limited portion of the spectral energy distribution and provide a narrow wavelength baseline over which the perturbation vectors of age, metallicity, and extinction are nearly co-aligned. As discussed in Section 4.5.3, this near-degeneracy produces strongly correlated posteriors and limits the precision of isochronal parameter recovery from optical data alone.

Near-infrared photometry extends the wavelength baseline to $\sim$2.2 μm, where the reddening vector is geometrically distinct from the temperature and metallicity vectors on the CMD. The separation arises because dust extinction decreases steeply with wavelength ($A_K/A_V \approx 0.11$ for a standard Cardelli, Clayton, and Mathis (1989) reddening law), so that near-infrared magnitudes are less affected by extinction than optical magnitudes. Combining optical and infrared colours therefore rotates the reddening direction relative to the isochrone ridge line, breaking the optical degeneracy and improving the simultaneous recovery of age, extinction, and distance modulus (Naylor and Jeffries, 2006; Cardelli, Clayton, and Mathis, 1989). This complementarity motivates the inclusion of J, H, and K_s photometry from the Two Micron All Sky Survey (2MASS; Skrutskie et al., 2006) in the present analysis.

5.3.2 The 2MASS photometric system

2MASS surveyed the entire sky in three near-infrared bands: J (1.235 μm), H (1.662 μm), and K_s (2.159 μm), using two 1.3-m telescopes at Mt. Hopkins and Cerro Tololo (Skrutskie et al., 2006). The survey reaches a 10σ point-source sensitivity of $J \approx 15.8$, $H \approx 15.1$, and $K_s \approx 14.3$ mag, with photometric calibration accurate to $\sim$2–3%. At the distance of NGC 6383 ($d \approx 1.1$ kpc), these limits correspond to absolute magnitudes of $M_J \approx 5.5$, $M_H \approx 4.8$, and $M_{K_s} \approx 4.0$ mag, which for the cluster's age of $\sim$3.5 Myr correspond approximately to spectral types late-G to early-K on the pre-main sequence. The 2MASS photometric census of NGC 6383 is therefore complete for intermediate- and high-mass members but misses the faintest, lowest-mass PMS stars that Gaia detects in the optical.

5.3.3 Crossmatch procedure and statistics

The crossmatch between Gaia DR3 and the 2MASS Point Source Catalogue was performed using the pre-computed best-neighbour table `tmass_psc_xsc_best_neighbour` provided within the Gaia archive (Marrese et al., 2022). This table contains, for each Gaia source, the best-matching 2MASS counterpart identified through a probabilistic algorithm that accounts for positional uncertainties, proper-motion propagation, and local source density. We adopted a maximum angular separation threshold of 0.3 arcsec, consistent with the typical positional accuracy of 2MASS ($\sim$0.1–0.2 arcsec) and the epoch difference between the two catalogues ($\sim$15 yr).

The crossmatch yielded 5 333 sources with valid J, H, and K_s magnitudes, representing 34.9% of the 15 276 Gaia working sample. The unmatched 65.1% are predominantly faint ($G > 17$ mag) and correspond to sources below the 2MASS

detection threshold. The match fraction is a function of apparent magnitude: it exceeds 90% for $G < 15$ mag and drops below 20% for $G > 18$ mag. The distribution of angular separations between matched Gaia–2MASS pairs is shown in Fig. 5.2; the majority of matches fall below 0.15 arcsec, confirming the reliability of the crossmatch at the adopted threshold.

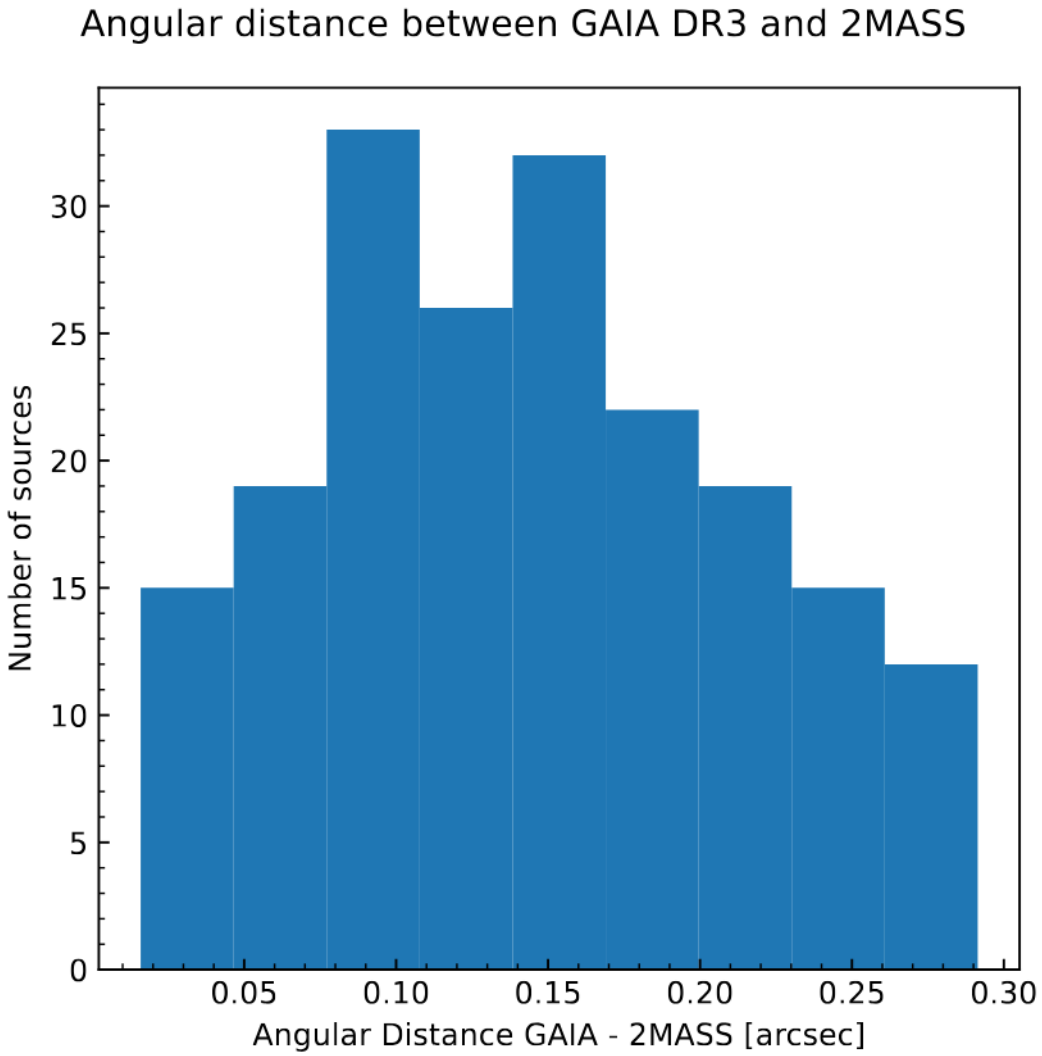

Figure 5.2: Distribution of angular separations between matched Gaia DR3 and 2MASS sources. The histogram shows the number of crossmatched pairs as a function of the on-sky angular distance. The concentration of matches at separations below ~ 0.15 arcsec confirms the astrometric consistency between the two catalogues and validates the 0.3 arcsec maximum separation threshold adopted for the crossmatch.

We implemented the crossmatch as a left join, preserving all 15 276 Gaia sources and appending 2MASS photometry where available. This ensures that the membership inference and kinematic analysis operate on the full Gaia sample, while the isochronal fitting and PMS classification (Sections 6.6 and 6.7) utilise the Gaia+2MASS subsample. Sources without 2MASS counterparts are flagged and excluded from infrared-dependent analyses but are not removed from the working sample.

The colour–colour consistency between the Gaia and 2MASS photometric systems was verified by checking the $G_{\mathrm{RP}} - J$ colour distribution of matched sources against the expected stellar locus for the field population. No systematic offsets or anomalous colour populations were identified, confirming the integrity of the crossmatch for the subsequent photometric inference.

5.4 Sample definition and quality filters

The working sample entering the COSMIC pipeline is the product of the cumulative filter chain described in the preceding sections. We document the complete sequence here, together with the source counts at each stage, to provide an explicit and reproducible specification of the observational input.

5.4.1 Cumulative filter chain

The six selection criteria, applied sequentially, are documented in Table 5.1. Steps 1–4 are source-removing filters that progressively narrow the sample from the initial cone search to the working catalogue. Steps 5–6 are value-correcting operations that modify the astrometric parameters in place without altering the source count.

Table 5.1: Cumulative filter chain defining the working sample. Steps 1–4 remove sources; steps 5–6 correct astrometric values in place. Section references point to the documentation of each criterion.

Step	Criterion	Sources	Section
1	Cone search, $r = 40.0$ arcmin	23 740	§ 5.1.2
2	Parallax pre-filter, $0.750 \leq \varpi \leq 1.10$ mas	—	§ 5.1.3
3	Astrometric fidelity > 0.5	15 480	§ 5.2.3
4	Full 5-parameter + 3-band solution	15 276	—
5	Parallax zero-point correction (Eq. 5.2)	15 276	§ 5.2.1
6	Bright-star PM frame correction ($G = 11$–13 mag)	15 276	§ 5.2.2

The final working sample consists of **15 276** sources with corrected five-parameter astrometry and three-band Gaia photometry. Of these, 5 333 (34.9%) additionally possess J, H, and K_s photometry from the 2MASS crossmatch of Section 5.3.

5.4.2 Relationship between data preparation and inference

A deliberate boundary separates the data preparation documented in this chapter from the inferential pipeline documented in Chapter 6. The six steps above are non-inferential: they apply deterministic corrections and hard quality cuts that define the input catalogue. They do not assign membership, estimate physical parameters, or invoke probabilistic models.

The subsequent steps of the COSMIC pipeline—HDBSCAN clustering, pseudoprobability construction, 2σ parallax clipping, and membership thresholding (Section 6.2)—are part of the inference and are treated as components of the membership model rather than data preparation. This separation ensures that the inferential assumptions and their consequences are documented within the model description, while the observational inputs are specified independently and reproducibly.

The working sample of 15 276 sources is the fixed observational boundary condition for the analysis. All physical claims in this thesis are conditioned on this sample

and are valid only within its domain of completeness and quality—the interpretive validity framework of Section 3.4.1.

5.5 Observational biases, completeness, and selection effects

The working sample of 15 276 sources is not a complete census of the stellar population within 40 arcmin of the NGC 6383 centre. The data are shaped by the Gaia survey selection function, the quality filters imposed in Section 5.4, and the photometric depth of the 2MASS crossmatch. These selection effects define the domain within which the physical claims of Chapter 7 are warranted, and their consequences must be stated explicitly—as required by the interpretive validity framework of Section 3.4.1.

5.5.1 Photometric completeness

Gaia DR3 reaches a nominal survey limit of $G \approx 20.7$ mag, but the effective completeness degrades as a function of magnitude, colour, stellar density, and the satellite's scanning-law coverage. In the Galactic plane direction of NGC 6383 ($b = 0.05°$), where the stellar density is high and nebular emission from the surrounding H II region Sh 2-012 contributes to the background, the completeness onset degrades to $G \approx 18$–19 mag (Boubert and Everall, 2020; Cantat-Gaudin et al., 2023). Fabricius et al. (2021) characterised the completeness as a function of magnitude and crowding for the Gaia source detection pipeline, finding that in fields with stellar densities exceeding $\sim 10^5$ deg^{-2}, the effective completeness limit can be 1–2 mag brighter than the nominal survey depth.

At the distance of NGC 6383 ($d \approx 1.1$ kpc) and with the extinction $A_V \approx 1.24$ mag

measured in Section 7.4.1, the Gaia completeness limit of $G \approx 18$–19 mag corresponds to absolute magnitudes $M_G \approx 6.5$–7.5 mag, which for a 3.5 Myr isochrone at solar metallicity corresponds approximately to PMS stars of mass ~ 0.3–$0.5\ M_\odot$ on the MIST grid. The mass function recovered from the Gaia data is therefore truncated at the low-mass end: late-K and M-type PMS stars that constitute the numerical majority of the cluster population are systematically underrepresented. The 254 members reported in Chapter 7 represent the *recoverable* membership above the Gaia detection threshold, not a complete enumeration of the cluster population.

This incompleteness has direct consequences for any physical summary that depends on the low-mass population. The total cluster mass derived from the recovered membership is a lower bound. The mass function slope is constrained only above the completeness mass. The spatial extent measured from the detected members may underrepresent the true halo extent if faint, low-mass members preferentially populate the periphery—as would be expected in a cluster exhibiting mass segregation (Section 7.5).

Deeper near-infrared photometry is in principle available for the NGC 6383 field. The VISTA Variables in the Vía Láctea survey (VVV; Minniti et al., 2010) and its extension VVVX cover the cluster location ($l \approx 355.7°$, $b \approx 0.1°$) and reach $K_s \approx 18$ mag—approximately four magnitudes deeper than the 2MASS 10σ limit of $K_s \approx 14.3$ mag. The VIRAC2 catalogue (Smith et al., 2024) provides VVV-derived proper motions, parallaxes, and $ZYJHK_s$ photometry for over 545 million sources with $> 90\%$ completeness for $11 < K_s < 16$. Exploiting VVV/VVVX photometry would extend the mass function below the 2MASS completeness threshold and improve the census of disc-bearing PMS stars in the cluster outskirts, but would require re-calibration of the isochrone fitting pipeline for the VISTA photometric system. We identify this as a priority extension for

future work (Section 9.4).

5.5.2 Astrometric precision degradation

The formal uncertainties on the Gaia astrometric parameters increase steeply with apparent magnitude. For $G < 14$ mag, the parallax uncertainty is $\sigma_\varpi \lesssim 0.03$ mas, yielding fractional errors $\sigma_\varpi/\varpi \lesssim 3\%$ at $\varpi \approx 0.9$ mas. For $G \approx 17$ mag, σ_ϖ rises to ~ 0.1 mas ($\sigma_\varpi/\varpi \approx 11\%$); for $G \approx 19$ mag, σ_ϖ reaches ~ 0.3–0.5 mas ($\sigma_\varpi/\varpi \approx 30$–$55\%$). Proper-motion uncertainties follow a parallel trend, increasing from ~ 0.02 mas yr^{-1} at $G = 14$ to ~ 0.5 mas yr^{-1} at $G = 19$.

The consequences for the membership pipeline are twofold. First, faint sources carry proper-motion uncertainties comparable to the intrinsic cluster dispersion ($\sigma_\mu \sim 0.15$ mas yr^{-1}), making their kinematic classification ambiguous: they lie in the overlap region between the cluster and field distributions in proper-motion space, reducing the discriminating power of the HDBSCAN step. Second, faint sources enter the prior-dominated regime for the distance inference identified in Section 4.1.5, where the parallax likelihood is broad and the posterior is shaped primarily by the prior rather than by the data. We mitigate this by restricting the parallax-based distance model to sources with $\sigma_\varpi/\varpi < 0.10$ (Section 6.3), ensuring that the distance posterior is data-driven for the contributing subsample. The excluded faint sources contribute to the membership and photometric analyses but do not directly constrain the cluster distance.

5.5.3 Radial velocity coverage

Gaia DR3 radial velocities are available for only $\sim 11.4\%$ of the 254 members at $p > 0.6$, corresponding to 29 sources. After excluding sources with binary probability above 0.6 (to avoid contamination by orbital motions), 16 sources

remain for the radial velocity characterisation.

This sample is both sparse and biased. The Gaia RVS is magnitude-limited at $G_{\rm RVS} \lesssim 14$ mag, so the radial velocity subsample is drawn almost exclusively from the bright, high-mass end of the cluster population. Low-mass PMS stars, which constitute the majority of the membership, have no radial velocity measurements. The resulting radial velocity distribution is therefore not representative of the full cluster kinematics; it reflects the velocity field of the most massive members, which may differ from the low-mass population if mass-dependent kinematics are present.

The median radial velocity of the 16-star subsample is -6.11 km s^{-1}, with a standard deviation of 31.1 km s^{-1}. The large dispersion likely reflects both the small sample size and the residual presence of unresolved spectroscopic binaries whose orbital velocities inflate the apparent dispersion. We report this value as a population summary in Chapter 7 but caution that it does not constitute a measurement of the cluster's internal velocity dispersion. Constraining the virial state of NGC 6383 (Section 7.6) would require radial velocities for a larger, mass-representative subsample—a dataset that is not available in DR3 and that may become accessible with future Gaia releases or dedicated spectroscopic follow-up.

5.5.4 Crowding, blending, and nebular contamination

NGC 6383 lies at Galactic latitude $b = 0.05°$, placing it directly in the Galactic midplane toward the inner Galaxy ($\ell = 355.68°$). The stellar density in this direction is among the highest in the sky, and the surrounding H II region Sh 2-012 contributes spatially variable nebular emission to the photometric background. Both effects compromise the fidelity of the Gaia source detection and photometry.

Crowding increases the probability that two physically unrelated sources fall

within the same point-spread function footprint, producing a blended source whose astrometric solution and photometry represent a luminosity-weighted average of the two components. For the Gaia angular resolution (~ 0.4 arcsec FWHM), blending becomes significant at separations below ~ 1 arcsec. In the cluster core, where the projected stellar density is highest, the blend fraction can reach several per cent for faint sources. Blended sources receive systematically biased parallaxes and proper motions, and their photometry is displaced from the single-star locus on the CMD, mimicking the effect of unresolved binarity (Fabricius et al., 2021).

The astrometric fidelity filter of Section 5.2.3 partially suppresses blended sources, since their astrometric solutions tend to produce elevated residuals that lower the fidelity score. However, the filter is not specifically designed to detect blends, and close blends with consistent proper motions may pass the fidelity cut undetected. The residual blend fraction in the working sample is not independently characterised, and its consequences for the binary fraction and CMD morphology are noted as unquantified systematic effects.

Nebular emission from Sh 2-012 affects the Gaia $G_{\rm BP}$ photometry, which includes the Hβ and [O III] lines. For sources superimposed on bright nebular knots, the $G_{\rm BP}$ flux can be systematically overestimated, producing anomalously blue $G_{\rm BP} - G_{\rm RP}$ colours. This effect is most severe for faint sources whose stellar flux is comparable to the local nebular background. The impact on the isochronal fit is mitigated by the use of multi-band photometry (Gaia + 2MASS) in the CMD likelihood, but it represents a source of systematic uncertainty for faint-source colours that is not explicitly modelled.

5.5.5 2MASS coverage and infrared selection effects

The 2MASS crossmatch yields near-infrared photometry for 5 333 of 15 276 sources (34.9%). The unmatched fraction is a direct consequence of the 2MASS photometric depth relative to the Gaia detection limit: 2MASS saturates at $J \approx 4$ mag and reaches 10σ at $J \approx 15.8$ mag, while Gaia detects sources to $G \approx 20.7$ mag. At the distance and extinction of NGC 6383, the 2MASS completeness corresponds to approximately $G \lesssim 16$–17 mag, leaving the faint half of the Gaia sample without near-infrared counterparts.

This differential depth introduces a luminosity-dependent selection effect into all analyses that require 2MASS photometry. The Sagitta PMS classification (Section 6.7) operates exclusively on the 2MASS-matched subsample, so the PMS census is restricted to intermediate- and high-mass PMS stars and misses the lowest-mass members. The isochronal fit (Section 6.6) uses both optical and infrared CMDs, but the infrared constraint is available only for the brighter subsample. The YSO fraction derived from the reddening-free Q parameter (Section 7.4.4) is computed relative to the 2MASS-matched membership, not the full Gaia membership; its extrapolation to the full population assumes that the YSO fraction is independent of mass, which is unlikely for a cluster with an age comparable to the disc-dissipation timescale.

5.5.6 Summary: the observational boundary of validity

The biases documented above—photometric incompleteness, astrometric precision degradation, radial velocity sparsity, crowding, nebular contamination, and differential 2MASS coverage—collectively define the observational boundary within which the physical claims of this thesis operate. The recovered population is not the cluster in its totality; it is the subset that the Gaia+2MASS survey, the

quality filters of Table 5.1, and the membership pipeline are jointly capable of detecting and characterising. Every physical summary reported in Chapter 7 is conditioned on this observational domain. Extensions to the full low-mass population, the three-dimensional kinematic state, or the complete binary census will require deeper photometric surveys, wider-field spectroscopy, and future Gaia data releases.

Chapter 6

The COSMIC Methodological Framework

This chapter documents the implementation of the formal machinery of Chapter 4 within COSMIC—the Characterisation Of Star clusters using Machine learning Inference and Clustering framework (Pulgar-Escobar et al., 2024). The architecture and design principles are established in Section 6.1. The HDBSCAN-based membership pipeline is documented in Section 6.2, the hierarchical distance inference in Section 6.3, and the proper-motion model in Section 6.4. The structural parameter estimation—centre determination, King profile fitting, and dynamical radii—is presented in Section 6.5. The Bayesian isochronal fitting with ASteCA and the pre-main-sequence classification with Sagitta are documented in Sections 6.6 and 6.7, respectively. Convergence diagnostics and validation are collected in Section 6.8, operational limitations and failure modes in Section 6.9, and the reproducibility and data release strategy in Section 6.10. Each component inherits the inferential architecture of Section 4.1 and operates under the claim-status discipline of Section 3.5.

6.1 Architecture and design principles

COSMIC—the Characterisation Of Star clusters using Machine learning Inference and Clustering framework—is an open-source Python 3.12 suite designed to perform the complete physical characterisation of open clusters from Gaia-era survey data[1]. The framework integrates two complementary algorithmic paradigms: unsupervised density-based clustering for non-parametric membership discovery and Bayesian hierarchical modelling for physical parameter estimation. The former instantiates the density-level-set approach of Section 4.4; the latter instantiates the generative measurement-model architecture of Section 4.1.

6.1.1 Modular design

The architecture is modular: each physical inference domain—membership, distance, kinematics, structure, photometry, PMS classification—is implemented as an independent component that reads a shared data structure and writes posterior samples to a common output format. This modularity has three intended consequences. First, each component can be validated independently against synthetic and external benchmark data without executing the full pipeline. Second, the pipeline can be reconfigured for different cluster environments—varying field contamination, photometric depth, and available external catalogues—without modifying the core inference engine. Third, the separation of concerns simplifies the diagnostic workflow: a convergence failure in the distance model, for instance, can be diagnosed and repaired without invalidating the membership or photometric components.

The shared data structure holds the corrected astrometry and photometry produced by the data preparation of Chapter 5. Each component accesses this

[1] `https://github.com/notluquis/COSMIC`

structure read-only and produces a posterior output that downstream components can condition on. The membership pipeline (Section 6.2) produces pseudoprobabilities; the distance component (Section 6.3) produces a cluster distance posterior conditioned on the membership; the structural model (Section 6.5) conditions on both membership and centre determination; and the isochronal fit (Section 6.6) conditions on membership, distance, and photometry.

6.1.2 Bayesian inference engine

All Bayesian inference within COSMIC is executed through PyMC 5[2], a Python library for probabilistic programming that provides a high-level interface for specifying generative models and sampling their posteriors (Abril-Pla et al., 2023). The primary sampling algorithm is the No-U-Turn Sampler (NUTS), the adaptive extension of Hamiltonian Monte Carlo described in Section 4.1.6. NUTS adaptively determines the trajectory length during warm-up and dynamically tunes the step-size and mass matrix, eliminating the manual tuning that conventional HMC requires (Hoffman and Gelman, 2011).

The choice of PyMC 5 and NUTS over alternative inference backends is deliberate. The geometric diagnostics that NUTS produces—divergent transitions, energy Bayesian Fraction of Missing Information (E-BFMI), and the rank-normalised split-$\hat{R}$—provide formal guarantees on sampling validity that are unavailable in the non-gradient-based samplers historically bundled with ASteCA's default configuration (Section 6.6). A divergent transition is not merely a numerical inconvenience; it is a rigorous proof that the sampler has failed to represent the target geometry in a specific region of the posterior (Betancourt, 2016). Any inference that proceeds despite divergences is formally compromised, and the convergence hierarchy established in Section 4.1.6—$\hat{R} < 1.01$, Bulk-ESS > 400,

[2] `https://www.pymc.io`

Tail-ESS > 400, zero divergences, E-BFMI > 0.3—is enforced for every posterior reported in this thesis.

6.1.3 Workflow

The COSMIC workflow follows the iterative Bayesian modelling cycle of Gelman et al. (2020): specify a generative model with explicit priors; sample the posterior with NUTS; diagnose convergence through the criteria above; assess model adequacy through posterior predictive checks; and, if the diagnostics reveal pathologies, revise the model or its parameterisation before proceeding. This cycle connects the formal model-criticism framework of Section 4.1.7 to the operational practice of the COSMIC pipeline. Each model component documented in Sections 6.2–6.7 is subject to this cycle, and the diagnostic outcomes are reported alongside the posterior summaries in Chapter 7.

6.2 Membership pipeline

The membership determination implements the hybrid density-based $\rightarrow$ probabilistic architecture described in Section 4.4.3. The pipeline proceeds in four stages: HDBSCAN clustering in proper-motion space, iterative hyperparameter optimisation, pseudoprobability construction, and parallax-based refinement. Each stage is documented below with its rationale, configuration, and connection to the formal membership framework of Section 4.2.

6.2.1 HDBSCAN clustering in proper-motion space

We applied HDBSCAN (Campello, Moulavi, and Sander, 2013; McInnes, Healy, and Astels, 2017) to the two-dimensional proper-motion space $(\mu_{\alpha*}, \mu_\delta)$ of the

15 276 sources defined in Section 5.4. Proper motions were chosen as the primary clustering features because they provide the strongest contrast between the kinematically cold cluster population ($\sigma_\mu \sim 0.15$ mas yr^{-1}) and the kinematically hot Galactic field at the distance of NGC 6383. Including parallax or photometry in the feature space was considered but rejected at this stage: the parallax distribution of the field population overlaps substantially with the cluster at $\varpi \sim 0.9$ mas, and photometric features introduce colour-dependent biases that couple the membership inference to the extinction model.

The distance metric employed for the proper-motion space is the Euclidean metric. We tested the Haversine (great-circle) metric, which is the formally correct distance on the celestial sphere, but found that it produced non-physical clusters. This outcome is expected: the proper-motion dispersions of NGC 6383 are $\sim$0.15 mas yr^{-1}, subtending negligible angles on the great circle, so the curvature correction of the Haversine metric introduces numerical noise without geometric benefit.

HDBSCAN was configured to extract leaf clusters rather than the excess-of-mass optimal partition. Leaf extraction generates numerous small, homogeneous groups, maximising the inclusion of potential cluster sources—including those in the outer halo and tidal tails whose study is a primary scientific goal of this thesis. The cost of this aggressive extraction strategy is an elevated false-positive rate: Hunt and Reffert (2020) demonstrated that HDBSCAN in Gaia proper-motion space frequently reports dense random fluctuations of the field as significant clusters. The false-positive rate is suppressed by the pseudoprobability construction of Section 6.2.3. The minimum samples parameter $m_{\rm pts}$, which governs the core distance in the mutual reachability computation (Eq. 4.14), was set to $m_{\rm pts} = m_{\rm cl}$ (the HDBSCAN default), so that both smoothing scales are controlled by a single hyperparameter.

6.2.2 Iterative hyperparameter optimisation

The minimum cluster size $m_{\rm cl}$—the smoothing parameter that controls the minimum number of points required for a density component to be classified as a cluster rather than noise—is the most consequential free parameter of HDBSCAN (Section 4.4.2). Small values of $m_{\rm cl}$ produce many small, potentially spurious clusters; large values suppress genuine low-density substructure.

We optimised $m_{\rm cl}$ through a systematic grid search, evaluating the cluster membership count and the maximum persistence parameter $\lambda_{\rm max}$ across a range of $m_{\rm cl}$ values. The objective was to maximise the recovered cluster size while maintaining high cluster separation strength, quantified by $\lambda_{\rm max}$. The optimal value was $m_{\rm cl} = 43$, at which the cluster size reaches its peak of 701 sources before declining at higher $m_{\rm cl}$ values (Fig. 6.1). The condensed cluster tree at this configuration is shown in Fig. 7.1, where the NGC 6383 cluster appears as the dominant persistent component with the field population occupying the complementary branch.

6.2.3 Pseudoprobability construction

To suppress the false-positive rate inherent to the leaf-cluster extraction strategy, we constructed a composite pseudoprobability metric that combines two independent indicators of membership.

First, we ran HDBSCAN multiple times with varying hyperparameter configurations and, for each source, computed the *iteration fraction* f_i: the number of iterations in which source i was classified as a cluster member, divided by the total number of iterations. Sources that consistently appear in the cluster across configurations are more robust detections than those that appear only under specific settings.

Second, we extracted the HDBSCAN λ-persistence value $p_{{\rm HDBSCAN},i}$ from Eq. (4.15)

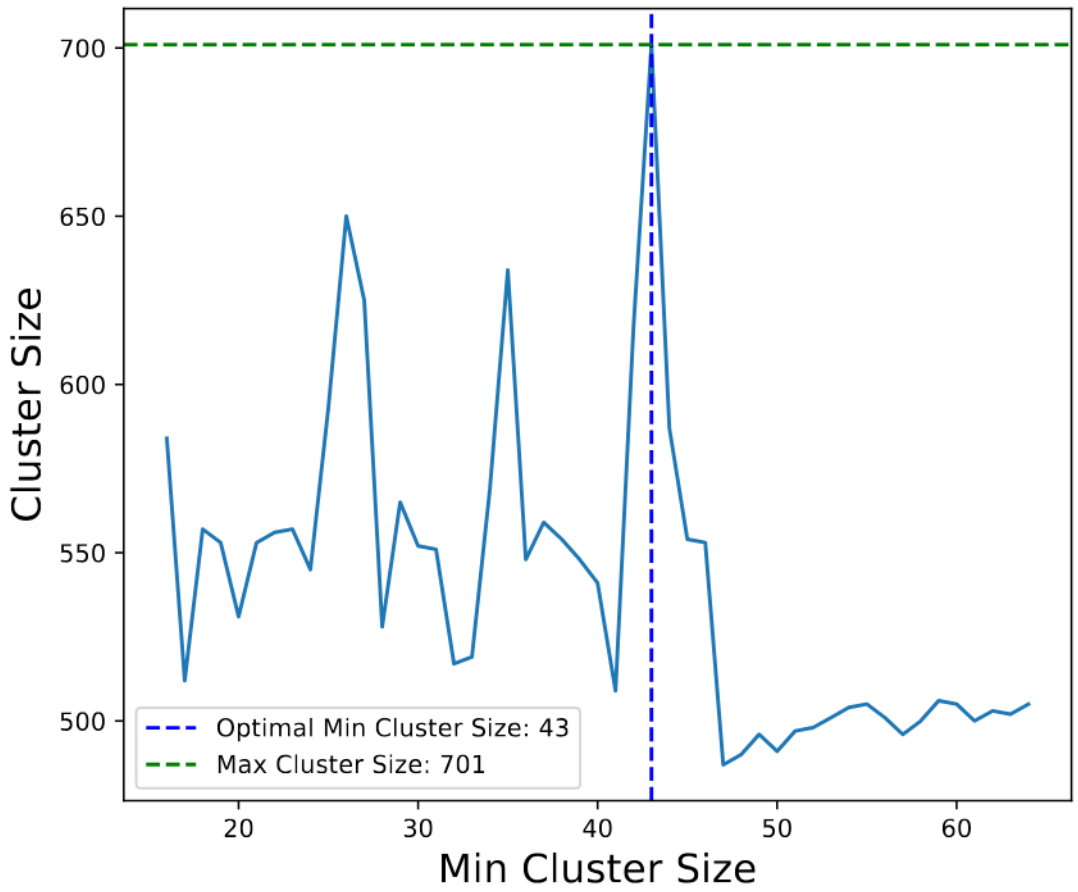

Figure 6.1: Cluster size variation as a function of minimum cluster size, determined through HDBSCAN clustering on the proper-motion data. The blue line represents cluster sizes achieved at various minimum cluster sizes. The dashed blue line highlights the optimal minimum cluster size of 43, where the cluster size reaches a peak. The dashed green line indicates the maximum observed cluster size of 701.

at the optimal $m_{\rm cl} = 43$ configuration. This value quantifies how deeply within the persistent density core each source resides.

The pseudoprobability for each source combines both indicators:

$$\tilde{p}_i = f_i \cdot p_{\mathrm{HDBSCAN},i} \,, \tag{6.1}$$

following the formal definition of Eq. (4.16). Sources with $\tilde{p}_i > 0.5$ were retained for subsequent analysis. The product structure ensures that a source must both appear consistently across HDBSCAN configurations *and* reside within the persistent density core to receive a high pseudoprobability. A source that appears in every iteration but at the cluster boundary ($p_{\mathrm{HDBSCAN},i} \approx 0$) receives a low pseudoprobability; conversely, a source deep within the core in one configuration but absent in most others also receives a low value. We note that f_i captures sensitivity to $m_{\rm cl}$ through the multi-configuration sweep, whereas $p_{\mathrm{HDBSCAN},i}$ is evaluated at the single optimal $m_{\rm cl} = 43$; the pseudoprobability is therefore asymmetric in how it treats the hyperparameter uncertainty across its two factors.

As established in Section 4.2.2, $\tilde{p}_i$ is *not* a posterior marginal membership probability in the sense of Eq. (4.7): it combines a geometric persistence ratio with a frequentist iteration statistic and carries no calibration against a generative likelihood. We use $\tilde{p}_i$ throughout the remainder of this thesis as a practical membership proxy despite this limitation, and the threshold-sensitivity analysis of Section 7.1.2 quantifies the resulting uncertainty on all downstream physical summaries. We define two operational tiers: *members* ($\tilde{p}_i \geq 0.8$) and *probable members* ($0.6 \leq \tilde{p}_i < 0.8$). Together, these tiers constitute the reference sample at $\tilde{p}_i > 0.6$ that enters all downstream analyses; the distinction between $\tilde{p}_i$ and a calibrated posterior probability is not restated in later chapters but remains in force throughout.

6.2.4 Parallax-based refinement

The HDBSCAN pipeline operates exclusively in proper-motion space and is therefore blind to distance. A kinematic interloper—a field star whose proper motion coincidentally falls within the cluster distribution but whose parallax places it at a radically different distance—passes the HDBSCAN step undetected.

We remove such interlopers through a 2σ outlier rejection applied to the parallax distribution of the HDBSCAN-selected sources. The rejection is centred on the mode of the parallax distribution and implemented using the ASTROPY sigma-clipping utility (The Astropy Collaboration et al., 2022), which iteratively identifies and removes sources whose parallaxes deviate by more than 2σ from the central value. After applying both the parallax clip and the pseudoprobability threshold $\tilde{p}_i > 0.5$ (Section 6.2.3), 321 of the original 701 HDBSCAN sources remain.

The 2σ threshold represents a deliberate trade-off between completeness and purity. A tighter threshold (1σ) would remove genuine members whose parallaxes are scattered to the wings of the distribution by measurement noise—particularly faint members with large σ_{ϖ}. A broader threshold (3σ) would admit more kinematic interlopers. The 2σ value was adopted as a standard compromise and its consequences are assessed through the threshold-sensitivity analysis of Section 7.1.

The parallax clipping is a deterministic selection step, not a probabilistic model. Combined with the $\tilde{p}_i > 0.5$ threshold, it removes 380 of the original 701 sources (54%) through hard cuts that do not propagate parallax uncertainties into the membership inference and do not account for the asymmetric, non-Gaussian structure of the parallax error distribution at faint magnitudes. This constitutes the most significant departure from the probabilistic membership framework developed in Sections 3.2–4.2, where membership is treated as a continuous

posterior quantity rather than a binary label. A fully probabilistic treatment would incorporate the parallax likelihood into the membership mixture model of Eq. (4.6), jointly constraining membership and distance without requiring a deterministic rejection stage. This extension is identified as a priority for future versions of COSMIC (Section 6.9). The impact of the current threshold on the downstream physical summaries is bounded by the threshold-sensitivity analysis of Section 7.1.

6.3 Hierarchical distance inference

The distance to NGC 6383 is inferred through a hierarchical Bayesian model that estimates a population-level cluster distance from the individual parallaxes of the confirmed members, implementing the architecture formalised in Section 4.3.4. The model avoids the pathological biases of naive parallax inversion (Section 4.3.1) by operating within the generative forward-model framework: the likelihood is evaluated in parallax space, and the distance enters through a prior defined in distance space.

6.3.1 Sample selection for the distance model

We restricted the distance model to members with fractional parallax errors $\sigma_\varpi/\varpi < 0.10$. This threshold ensures that the parallax likelihood is informative for each contributing source—that is, the data dominate the prior in the individual-star distance posteriors (Section 4.3.3). Sources with $\sigma_\varpi/\varpi > 0.10$ carry likelihoods that are broad relative to the distance prior; including them would broaden the cluster distance posterior without meaningfully improving the mean estimate, while coupling the result more strongly to the assumed prior form. The excluded faint sources contribute to the membership, kinematic, and photometric analyses

but do not directly constrain the cluster distance.

6.3.2 External distance estimates

We obtained geometric distances from Bailer-Jones et al. (2021), who applied a Bayesian approach to Gaia EDR3 parallaxes with a direction-dependent prior informed by a three-dimensional model of Galactic stellar density. Geometric distances use only parallax and the spatial prior; photogeometric distances additionally incorporate photometric data to refine the estimate. We adopted geometric distances exclusively: the presence of pre-main-sequence stars and spatially variable nebular emission in NGC 6383 distorts the apparent magnitudes and colours on which the photogeometric prior depends, violating the assumptions of the photometric model and potentially biasing the distance estimates for the youngest cluster members.

6.3.3 Hierarchical model specification

The cluster distance is modelled as a single population-level parameter μ_d representing the mean heliocentric distance. The prior on μ_d is specified as

$$\mu_d \sim \mathcal{U}(0.5\,\mu_{\rm prior},\ 1.5\,\mu_{\rm prior})\,, \tag{6.2}$$

where $\mu_{\rm prior}$ is the average between the frequentist mean distance obtained from the naive parallax inversion of the member sample and the mean geometric distance from Bailer-Jones et al. (2021). The uniform prior spans $\pm 50\%$ of this anchor value, providing a weakly informative constraint that excludes physically implausible distances while remaining broad enough that the likelihood dominates the posterior within the plausible region.

Individual stellar distances are drawn from the population-level distribution centred on μ_d. The hierarchical structure follows the two-level architecture of Eq. (4.5), in which the population mean and intrinsic dispersion govern the distribution from which individual stellar distances are drawn, and each star's observed parallax constrains its individual distance through the parallax likelihood of Eq. (4.9).

The parallax zero-point correction of Section 5.2.1 is incorporated by using the corrected parallax $\varpi_{\rm corr}$ of Eq. (5.2) in the likelihood evaluation. Residual spatially correlated parallax errors at the $\sim$10–15 μas level (Section 5.2.1) are absorbed into the population-level mean and contribute to its uncertainty, but are not explicitly modelled as a spatial covariance structure.

6.3.4 Sampling and convergence

The posterior was sampled with NUTS via PyMC 5, using four independent chains of 2 000 draws each after a warm-up phase of 1 000 draws. Convergence was assessed against the diagnostic hierarchy of Section 4.1.6: $\hat{R} < 1.01$ for all parameters, Bulk-ESS > 400, Tail-ESS > 400, zero divergent transitions, and E-BFMI > 0.3. All criteria were satisfied without reparameterisation.

6.4 Proper-motion model

The internal kinematics of NGC 6383 are characterised through a Bayesian model of the proper-motion distribution of the confirmed members. The model estimates the population-level mean proper motion, the intrinsic velocity dispersion, and the correlation between the two proper-motion components, providing the kinematic parameters that enter the dynamical state assessment of Section 7.6.

6.4.1 Generative model

We modelled the proper-motion distribution of the cluster members as a two-dimensional Gaussian characterised by a mean vector $\bar{\boldsymbol{\mu}} = (\overline{\mu_{\alpha*}}, \overline{\mu_\delta})$, a covariance matrix parameterised through the standard deviations $(\sigma_{\mu_\alpha}, \sigma_{\mu_\delta})$ and a correlation coefficient ρ, and the individual measurement uncertainties $\sigma_{\mu,i}$ for each source. The likelihood for star i with observed proper motions $\boldsymbol{\mu}_i = (\mu_{\alpha*,i}, \mu_{\delta,i})$ is

$$p(\boldsymbol{\mu}_i \mid \bar{\boldsymbol{\mu}},\, \boldsymbol{\Sigma}) = \mathcal{N}(\boldsymbol{\mu}_i \mid \bar{\boldsymbol{\mu}},\, \boldsymbol{\Sigma} + \boldsymbol{C}_{\mu,i})\,, \tag{6.3}$$

where $\boldsymbol{\Sigma}$ is the intrinsic population covariance and $\boldsymbol{C}_{\mu,i}$ is the 2×2 measurement covariance matrix for star i, extracted from the Gaia DR3 astrometric solution. The sum $\boldsymbol{\Sigma}+\boldsymbol{C}_{\mu,i}$ ensures that the observed dispersion is correctly decomposed into an intrinsic astrophysical component and a measurement component, following the errors-in-variables logic established in Section 4.1.2.

6.4.2 Prior specification

The priors on the model parameters were specified as follows:

$$\begin{aligned}
\overline{\mu_{\alpha*}} &\sim \mathcal{N}(\hat{\mu}_{\alpha*},\ \hat{\sigma}_{\mu_\alpha}),\\
\overline{\mu_\delta} &\sim \mathcal{N}(\hat{\mu}_\delta,\ \hat{\sigma}_{\mu_\delta}),\\
\sigma_{\mu_\alpha} &\sim \mathrm{HalfNormal}(\hat{\sigma}_{\mu_\alpha}),\\
\sigma_{\mu_\delta} &\sim \mathrm{HalfNormal}(\hat{\sigma}_{\mu_\delta}),\\
\rho &\sim \mathcal{U}(-1,\ 1),
\end{aligned}$$

where $\hat{\mu}_{\alpha*}$ and $\hat{\mu}_\delta$ are the frequentist sample means, and $\hat{\sigma}_{\mu_\alpha}$ and $\hat{\sigma}_{\mu_\delta}$ are the frequentist sample standard deviations of the member proper motions. The normal priors on the means are centred on the data-driven estimates and allow

the posterior to shift within $\sim 1\sigma$ of the initial value. The half-normal priors on the dispersions concentrate near zero and gently penalise large dispersions, acting as weakly informative regularisers in the sense of Section 4.1.5. The uniform prior on ρ is uninformative and permits the full range of correlation structures.

6.4.3 Projected velocity

The mean projected velocity of the cluster on the plane of the sky is derived as a transformed quantity from the posterior:

$$v_{\mathrm{proj}} = \sqrt{\overline{\mu_{\alpha*}}^2 + \overline{\mu_\delta}^2}\,. \tag{6.4}$$

Its posterior is obtained by evaluating Eq. (6.4) at each MCMC sample, propagating the full joint uncertainty of the two mean proper-motion components including their covariance.

6.4.4 Sampling and convergence

The posterior was sampled with NUTS using four chains of 2 000 draws each after 1 000 warm-up draws. All convergence criteria of Section 4.1.6 were satisfied: $\hat{R} < 1.01$, Bulk-ESS > 400, Tail-ESS > 400, zero divergences, E-BFMI > 0.3. The resulting posteriors for $\bar{\boldsymbol{\mu}}$ are tightly constrained, with formal uncertainties of ~ 0.01 mas yr^{-1} on each component.

A physical caveat is that the two-dimensional Gaussian model assumes a unimodal, ellipsoidal velocity distribution. At $\tau/t_{\mathrm{cr}} \approx 1.4$ (Section 7.6.1), NGC 6383 has completed fewer than two crossing times and phase-space mixing is incomplete; kinematic substructure inherited from formation—such as expanding or rotating sub-groups—may persist. The single-Gaussian model will average over any such

substructure, producing a mean and dispersion that are adequate population summaries but that do not capture the internal kinematic structure. A multi-component or non-parametric kinematic model could in principle detect such substructure, but the current sample size (254 members) and proper-motion precision ($\sigma_\mu \sim 0.02$–0.5 mas yr^{-1}) provide limited statistical power for this test. We therefore report the single-Gaussian parameters as the primary kinematic characterisation and defer the search for kinematic substructure to future analyses with larger samples or higher-precision astrometry.

6.5 Structural parameter estimation

The structural characterisation of NGC 6383 requires three sequential inferences: the determination of the cluster centre, the fitting of a radial density profile, and the computation of external dynamical radii that constrain the gravitational boundary of the system. Each is documented below.

6.5.1 Centre determination

We determined the cluster centre using a weighted Kernel Density Estimation (KDE) in the (α, δ) plane, implemented through the SCIKIT-LEARN library (Pedregosa et al., 2012). Each source was assigned a weight inversely proportional to its angular distance from the mean proper motion of the membership sample, so that kinematically central members—whose proper motions most closely match the cluster mean—contribute more strongly to the density estimate than kinematically peripheral sources.

The KDE bandwidth and kernel type were optimized through grid-search cross-validation. The bandwidth grid spanned from the mean positional uncertainty of the sample ($\sim 0.001°$) to the search cone radius ($0.67°$), logarithmically sampled,

and all kernel types available in SCIKIT-LEARN (Gaussian, tophat, Epanechnikov, exponential, linear, cosine) were evaluated. The configuration minimizing the leave-one-out cross-validation score was selected, and the maximum of the resulting KDE surface was adopted as the cluster centre.

The uncertainty on the centre position was estimated as the quadratic sum of the mean positional error in RA and Dec and the optimal KDE bandwidth. This estimate incorporates both the formal astrometric uncertainty of the individual sources and the smoothing scale of the density estimator.

6.5.2 Radial density profile: King model

The radial density profile was fitted with the King (1962) profile of Eq. (2.6), which describes the projected surface density of a tidally limited stellar system through four parameters: the core radius R_c, the tidal radius R_t, the central density normalisation k, and the background density b. The King profile is the most widely used structural model for Galactic open clusters, and its adoption here provides direct comparability with the literature values documented in Table 1.1.

The observed density profile was constructed by dividing the cluster area into K concentric annuli, each containing an equal number of stars. The number of annuli follows the equiprobable bin rule $K = 2\,n^{2/5}$, where n is the star count within the sample (Scott, 1979). Equal-count binning ensures that the density estimate has approximately uniform statistical precision across all radii, avoiding the common pathology of large Poisson fluctuations in sparsely populated outer annuli that arises with equal-width binning.

The King profile of Eq. (2.6) was fitted to the observed annular densities $\{\hat{\Sigma}_j\}_{j=1}^{K}$

through a Gaussian likelihood on the binned surface densities:

$$\ln \mathcal{L} = -\frac{1}{2} \sum_{j=1}^{K} \left[\frac{\hat{\Sigma}_j - \Sigma(r_j \mid R_c, R_t, k, b)}{\sigma_j} \right]^2 , \tag{6.5}$$

where r_j is the midpoint radius of annulus j, $\Sigma(r_j \mid R_c, R_t, k, b)$ is the King profile evaluated at r_j via Eq. (2.6), and $\sigma_j = \hat{\Sigma}_j / \sqrt{n_j}$ is the Poisson uncertainty on the density estimate in an annulus containing n_j stars. With $K \approx 16$ annuli and 254 stars, each annulus contains ~ 16 stars; the Gaussian approximation to the Poisson likelihood is borderline at this count but is expected to yield similar posteriors to a full Poisson treatment given the moderate counts in all annuli. The four parameters were sampled jointly with NUTS under the prior specification of Section 6.5.3.

6.5.3 King model prior specification

The four King model parameters were sampled with NUTS under the following prior specification:

$$\begin{aligned} b &\sim \mathcal{U}(0,\ 2\,\rho_{\min}), \\ k &\sim \mathcal{U}(0,\ 2\,\rho_{\max}), \\ R_c &\sim \mathcal{U}(0,\ 0.8\,R_t), \\ R_t &\sim \mathcal{U}(R_c,\ 1.5\,T_{\max}), \end{aligned}$$

where $\rho_{\min}$ and $\rho_{\max}$ are the minimum and maximum annular densities in the observed profile, and $T_{\max} = \max(R_{\mathrm{Hill}},\ R_{\mathrm{bound}})$ is the larger of the two dynamical radii defined below. The conditional structure of the R_c and R_t priors—$R_c < 0.8\,R_t$ and $R_t > R_c$—enforces the physical ordering constraint that the core radius is smaller than the tidal radius, preventing label switching between the two

parameters (Section 4.1.4). The resulting joint prior is uniform on the triangular region $\{0 < R_c < 0.8\,R_t,\ R_c < R_t < 1.5\,T_{\max}\}$ and was implemented through an ordering constraint in PyMC 5. The 0.8 factor imposes a weak implicit lower bound on the concentration parameter ($C > 0.10$) that excludes the physically unrealistic regime $R_c \approx R_t$.

The upper bound on R_t is set at $1.5\,T_{\max}$ to allow the King model to explore tidal radii that exceed the theoretical gravitational boundary, accommodating potential escapers and the observational uncertainty in the dynamical radii. The upper bound on b at $2\,\rho_{\min}$ permits the background to fluctuate above the observed minimum while preventing the model from absorbing the entire outer profile into a flat background.

6.5.4 Dynamical radii

Two external dynamical radii were computed to constrain the tidal radius prior and to provide independent assessments of the gravitational boundary of the system.

The Hill radius was calculated from Eq. (2.8) using $R_{\rm GC} = 7.19 \pm 0.07$ kpc and $m_c = 902.3 \pm 92.3\ M_\odot$ from Hunt and Reffert (2024). The enclosed Galactic mass was estimated through the power-law approximation

$$M_{\rm GC}(< R_{\rm GC}) = M_0 \left(\frac{R_{\rm GC}}{R_0} \right)^{\gamma}, \tag{6.6}$$

where $M_0 = 9.65 \times 10^{10}\ M_\odot$, $R_0 = 8.178$ kpc, and $\gamma = 1.0$ parameterise the flat portion of the Milky Way rotation curve (Binney and Tremaine, 2008). At $R_{\rm GC} = 7.19$ kpc, this yields $M_{\rm GC} = 1.43 \times 10^{11} \pm 1.7 \times 10^{9}\ M_\odot$. The resulting Hill radius enters the prior specification of R_t in Section 6.5.3.

The gravitationally bound radius was computed from Eq. (2.9) using the Oort

constants A and B from Bovy (2015). This radius defines the distance from the cluster centre at which Galactic differential rotation unbinds a test particle, providing a complementary tidal-boundary estimate under different assumptions about the Galactic potential.

6.5.5 Derived structural quantities

The concentration parameter $C = \log_{10}(R_t/R_c)$ (Peterson and King, 1975) was computed from the posterior samples of R_c and R_t. The half-light radius $R_{\rm hl}$—the radius enclosing half of the total G-band luminosity—was computed from the radial cumulative luminosity distribution. The half-mass radius $R_{\rm hm}$—the radius enclosing half of the total stellar mass—was computed from the radial cumulative mass distribution, using the individual stellar masses estimated by ASteCA (Section 6.6). These quantities enter the dynamical state assessment of Chapter 7. The radial distribution of apparent magnitudes and the cumulative integrated magnitude as a function of projected radius are shown in Fig. 6.2; the integrated magnitude saturates beyond ~ 10 arcmin, confirming that the vast majority of the cluster's luminosity is enclosed within the half-light radius.

6.6 Isochronal fitting with ASteCA

The cluster age, metallicity, and extinction were inferred through the Bayesian isochronal fitting engine of ASteCA—Automated Stellar Cluster Analysis (Perren, Vázquez, and Piatti, 2015)—which implements the CMD forward-modelling architecture described in Section 4.5.1. This section documents the isochrone grid, the initial mass function, the prior specification, and the sampler configuration adopted for NGC 6383.

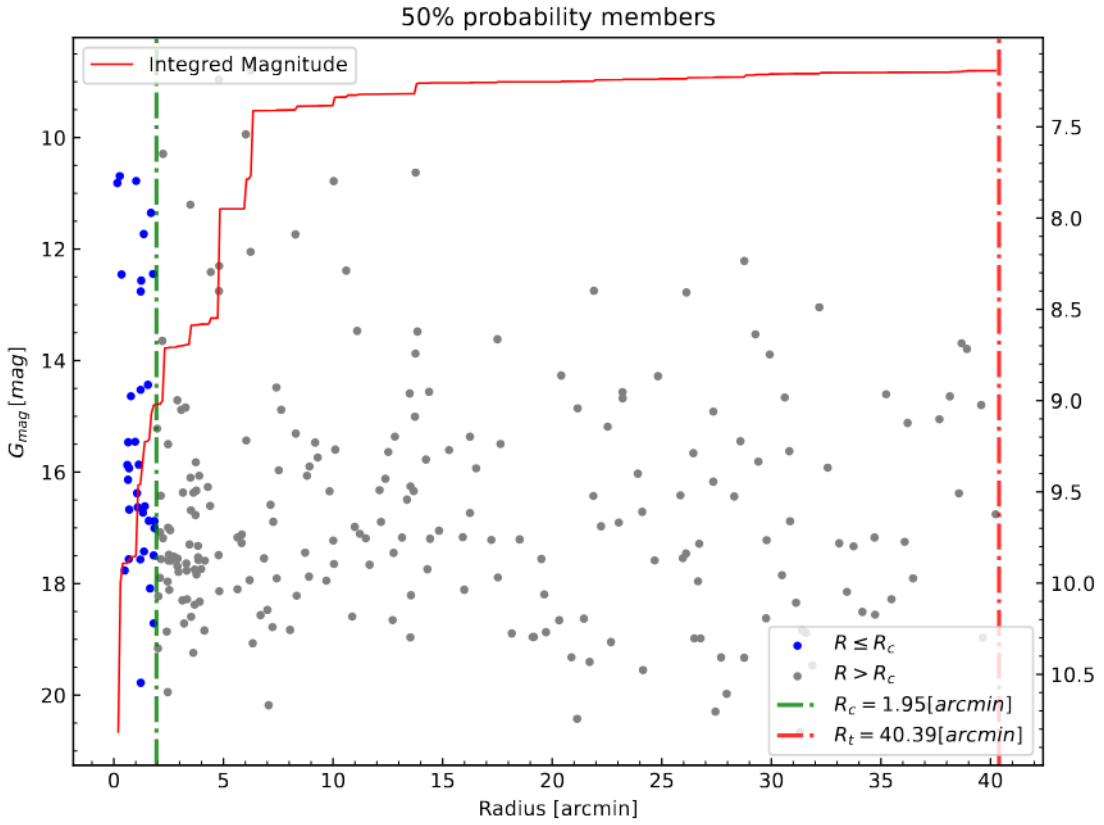

Figure 6.2: Radial distribution of apparent G-band magnitudes for the extended membership sample ($\tilde{p}_i > 0.5$, 321 sources) shown for completeness; the structural profile fit of Section 7.3 uses the $\tilde{p}_i > 0.6$ reference sample (254 sources). Blue points mark sources within the core radius $R_c = 1.95$ arcmin; grey points mark sources beyond R_c. The red curve shows the cumulative integrated magnitude as a function of projected radius, with the right-hand axis indicating $\sum G_{\rm mag}$. The green dash-dot line marks R_c and the red dash-dot line marks $R_t = 40.4$ arcmin.

6.6.1 Isochrone grid and photometric system

We adopted the MESA Isochrones and Stellar Tracks (MIST; Dotter, 2016; Choi et al., 2016) as the theoretical grid for the forward model. MIST provides pre-computed isochrones with self-consistent treatment of convective boundary mixing, atomic diffusion, and rotational mixing, spanning a wide range of masses, ages, and metallicities. The grid was extracted in the Gaia EDR3 photometric system (G, $G_{\rm BP}$, $G_{\rm RP}$) and the 2MASS system (J, H, K_s), ensuring that the synthetic photometry is computed through the specific bandpass response curves of the two surveys.

The choice of a single grid—MIST—rather than a multi-grid comparison introduces a systematic dependence on the physics encoded in the MIST models. As discussed in Section 4.5.6, fitting the same population with alternative grids (PARSEC, Baraffe, SPOTS) produces age offsets that can exceed the statistical precision of any single fit. The grid-to-grid variance is not propagated into the reported uncertainty in this analysis and is identified as a systematic floor in Section 6.9. We adopted a single-grid strategy as a deliberate scope decision: the thesis-level claims—the resolution of the HD 159176 ambiguity, the evidence for primordial segregation, the dynamical youth diagnostic—do not depend on the absolute age at the 0.1–0.3 dex level at which grid choice becomes consequential. The multi-grid comparison is deferred to a forthcoming analysis (Section 9.4) in which the model-discrepancy formalism of Eq. (4.18) will be applied operationally.

6.6.2 Initial mass function and multiplicity model

We adopt the initial mass function of Chabrier, Hennebelle, and Charlot (2014) as implemented in ASteCA, which combines a log-normal form below a characteristic mass $m_0 = n_c\, m_c \approx 2\,\mathrm{M}_\odot$ with a power-law tail of exponent $x = 1.35$ at higher

masses (their Eq. 34). This parameterisation differs from the more commonly cited Chabrier (2003) individual-star IMF (characteristic mass $m_c \approx 0.25\,\mathrm{M}_\odot$) in the transition mass and the treatment of the system mass function; it represents the default IMF configuration in ASteCA and was adopted without modification. The mass-dependent multiplicity fraction follows the observational compilation of Offner et al. (2022) and is parameterised as

$$b_p(M) = \alpha + \frac{\beta}{1 + 1.4\,M^{-1}}\,, \tag{6.7}$$

where M is the primary mass in solar units, $\alpha = 0.09$ sets the baseline binary fraction at low masses, and $\beta = 0.94$ controls its rise toward higher masses. This functional form yields $b_p \approx 0.09$ for very low-mass stars and $b_p \to 1.03$ (clipped to unity) for $M \gg 1\,\mathrm{M}_\odot$, broadly consistent with the multiplicity statistics compiled in Offner et al. (2022), their Fig. 1 and Table 1. The mass-ratio distribution for binary companions follows the mass-dependent prescription of Duchêne and Kraus (2013), as implemented in ASteCA's default configuration.

6.6.3 Forward model and likelihood

For each proposed parameter vector $\mathbf{\Phi} = (\tau,\, Z,\, A_V,\, \mu_{\mathrm{DM}})$, the forward model generates a synthetic CMD by: (i) drawing N_{syn} individual stellar masses from the adopted IMF; (ii) assigning binary companions according to the multiplicity model of Eq. (6.7) with mass ratios drawn from the Duchêne and Kraus (2013) prescription; (iii) mapping each (primary, secondary) mass pair to absolute magnitudes through the MIST isochrone at the proposed age and metallicity; (iv) applying bandpass-dependent extinction through the Cardelli, Clayton, and Mathis (1989) reddening law with the O'Donnell (1994) near-UV update for the proposed A_V; (v) shifting the absolute magnitudes by the distance modulus

$\mu_{\rm DM}$; and (vi) injecting heteroscedastic photometric noise consistent with the individual measurement uncertainties. Differential reddening was set to zero—a simplification whose consequences are assessed in Section 6.9.

The resulting synthetic CMD is compared to the observed CMD through a Poisson likelihood rate. Following the general CMD likelihood of Eq. (4.17), the log-likelihood for $\mathbf{\Phi}$ is

$$\ln \mathcal{L}(\mathbf{\Phi}) = \sum_{i=1}^{N_{\rm obs}} \ln \left[\frac{1}{N_{\rm syn}} \sum_{j=1}^{N_{\rm syn}} \prod_{d=1}^{D} \frac{1}{\sqrt{2\pi}\, h_{d,i}} \exp\left(-\frac{(m_{d,i}^{\rm obs} - m_{d,j}^{\rm syn})^2}{2\, h_{d,i}^2} \right) \right], \tag{6.8}$$

where $m_{d,i}^{\rm obs}$ is the observed magnitude of star i in photometric band d, $m_{d,j}^{\rm syn}(\mathbf{\Phi})$ is the corresponding synthetic magnitude drawn from the forward model, D is the number of photometric bands, and $h_{d,i}$ is the bandwidth in band d set by the photometric uncertainty $\sigma_{d,i}$ of the observed source. The inner sum over $N_{\rm syn}$ synthetic stars acts as a kernel density estimator in the D-dimensional magnitude space, evaluating the density of the synthetic population at each observed point. The outer sum accumulates the log-density contributions across the observed membership, so that parameter combinations producing synthetic CMDs that place high density at the observed stellar loci are favoured (Perren, Vázquez, and Piatti, 2015). This formulation implements the synthetic-cluster matching architecture of ASteCA within the Bayesian framework of Section 4.5.1.

6.6.4 Prior specification

The priors on the four free parameters were specified as follows:

$$Z \sim \mathcal{U}(0.001,\ 0.045), \tag{6.9}$$

$$\log(\text{age/yr}) \sim \mathcal{U}(6.00,\ 7.00), \tag{6.10}$$

$$A_V \sim \mathcal{U}(0.50,\ 2.00), \tag{6.11}$$

$$\mu_{\rm DM} \sim \mathcal{N}(10.3,\ 0.2). \tag{6.12}$$

The metallicity prior spans the full range from metal-poor to moderately super-solar, consistent with the expected Galactic abundance gradient at the Galactocentric distance of NGC 6383. The age prior covers 1–10 Myr in linear scale, encompassing the full range of published estimates (Table 1.1). The extinction prior brackets the historical values from $A_V \approx 0.8$ to 1.6 mag. The distance modulus prior is an informative Gaussian centred on $\mu_{\rm DM} = 10.3$ mag, corresponding to $d \approx 1.1$ kpc from the parallax-based distance inference of Section 6.3, with a width of 0.2 mag that accommodates the astrometric uncertainty. This informative prior on the distance modulus is the formal mechanism by which the parallax-based distance constraint propagates into the isochronal fit, breaking one dimension of the age–extinction–distance degeneracy (Section 4.5.3).

6.6.5 Sampler: NUTS replacing ASteCA's default backend

ASteCA's default inference backend compares synthetic CMDs—generated from isochrone grids at candidate parameter values—against the observed CMD using a Poisson likelihood rate statistic. Earlier versions of the code sampled this likelihood with a genetic algorithm (Perren, Vázquez, and Piatti, 2015); intermediate releases adopted parallel-tempered MCMC (ptemcee); and the current version is sampler-

agnostic, allowing the user to supply an external sampling engine. While the synthetic-cluster matching approach is computationally practical, the non-gradient-based samplers historically bundled with ASteCA provide no geometric diagnostics analogous to the divergent transitions and E-BFMI of Hamiltonian Monte Carlo (Sisson, Fan, and Beaumont, 2020).

We replaced ASteCA's default sampling backend with NUTS via PyMC 5 for the NGC 6383 analysis. This substitution provides two critical advantages. First, NUTS explores the posterior geometry using Hamiltonian dynamics, yielding low-correlation samples with formally bounded bias per effective sample. Second, the convergence diagnostics of Section 4.1.6—$\hat{R}$, ESS, divergences, E-BFMI—provide formal guarantees on the validity of the posterior that non-gradient-based samplers cannot offer. All diagnostics were satisfied for the NGC 6383 isochronal fit.

6.6.6 Posterior summary

The posterior was summarised using the mode and asymmetric highest-posterior-density (HPD) intervals. We adopt the mode rather than the mean or median because the isochronal posterior for young clusters is often asymmetric, with a longer tail toward older ages produced by the PMS population whose CMD positions are compatible with a range of ages (Section 4.5.4). The mode captures the most probable parameter combination, while the HPD intervals quantify the credible range. The mean and median are also reported for completeness but are not adopted as the primary point estimates.

6.7 Pre-main-sequence classification with Sagitta

The identification and age estimation of pre-main-sequence members provides an independent constraint on the star-formation history of NGC 6383, complementing the single-epoch isochronal age from ASteCA with individual stellar ages for the PMS population. We employed Sagitta (McBride et al., 2021), a neural-network classifier trained on Gaia and 2MASS photometry to determine PMS probability, individual stellar extinction, and individual stellar age.

6.7.1 Architecture and training data

Sagitta comprises three convolutional neural networks, each dedicated to a specific output: stellar extinction A_V, PMS probability, and age. The networks were trained on a curated dataset of PMS stars within kinematically well-defined young moving groups identified by Kounkel, Covey, and Stassun (2020), supplemented by additional young stellar populations with independently measured ages from the literature. The training photometry spans the Gaia DR2 and 2MASS systems; Sagitta requires as input the Gaia parallax, G, $G_{\rm BP}$, $G_{\rm RP}$, J, H, and K_s magnitudes, and the mean line-of-sight extinction.

Kerr, Kraus, and Rizzuto (2023) demonstrated that Sagitta provides age predictions that are stable across spectral types G, K, and M within the same population, surpassing traditional isochronal techniques in internal consistency for young stars up to $\sim$80 Myr. This spectral-type stability is a consequence of the network's empirical training on photometric sequences of known-age populations, which implicitly absorbs the effects of magnetic activity, rotation, and starspot coverage that produce mass-dependent age biases in standard isochrone fitting (Section 4.5.4).

6.7.2 Application to NGC 6383

We applied Sagitta to all members with available 2MASS photometry (5 333 sources in the working sample; ~ 190 among the 254 members at $\tilde{p}_i > 0.6$). Sources without 2MASS counterparts—predominantly faint ($G > 17$ mag) and corresponding to the lowest-mass PMS stars—receive no Sagitta classification, restricting the PMS census to the brighter, infrared-detected subsample (Section 5.5.5).

We classified sources as PMS when their Sagitta PMS probability exceeded 0.6, a threshold that balances completeness against the false-positive rate of the classifier. Sources with PMS probability below 0.6 are not necessarily confirmed main-sequence stars; they are sources for which the photometric evidence is insufficient to distinguish PMS status from the main-sequence locus with high confidence.

6.7.3 Young stellar object fraction

As an independent age indicator, we quantified the fraction of Young Stellar Objects (YSOs) within the cluster using the reddening-free parameter Q defined as

$$Q = (J - H) - \frac{E(J - H)}{E(H - K_s)} (H - K_s), \tag{6.13}$$

where the reddening ratio $E(J-H)/E(H-K_s) = 1.55$ follows Meyer, Calvet, and Hillenbrand (1997). A source is classified as a YSO if $Q < -0.05$ mag, following Buckner and Froebrich (2013). This criterion identifies sources with near-infrared excess emission from circumstellar discs or envelopes, which are characteristic of actively accreting young stars.

The YSO fraction $Y_{\rm frac} = N_{\rm YSO}/N_{\rm cl}$ was estimated using a Bayesian binomial model with a non-informative Beta$(1, 1)$ prior (Cameron, 2011). The posterior for $Y_{\rm frac}$ is a Beta distribution from which the 95% credible interval was extracted.

6.7.4 Limitations

Several limitations of the Sagitta classification must be stated. First, Sagitta was trained on Gaia DR2 photometry; the DR3 photometric system includes updated calibrations that introduce systematic offsets of typically 0.01–0.03 mag, particularly for faint and red sources (Riello et al., 2021). These offsets may shift the PMS probability boundary for individual sources near the classification threshold, though the effect on the aggregate PMS fraction is expected to be small. Second, the network is most reliable for PMS stars; main-sequence stars whose photometry has converged to the zero-age main sequence are not age-informative and receive unreliable age estimates. Third, the Q-parameter YSO criterion is a photometric proxy that is sensitive to the assumed reddening law; in regions with anomalous R_V, the Q threshold may misclassify reddened main-sequence stars as YSOs or miss genuine YSOs with low disc excess. Fourth, and most importantly, both Sagitta and the Q-parameter analysis are restricted to the 2MASS-matched subsample, so the PMS and YSO censuses are incomplete below the 2MASS detection threshold. The Sagitta ages and the YSO fraction reported in Chapter 7 must be interpreted within these bounds.

6.8 Validation and convergence diagnostics

Every posterior reported in this thesis must satisfy the convergence hierarchy established in Section 4.1.6 before the corresponding physical claims are admitted. This section documents the diagnostic protocol applied uniformly to all NUTS-based models in the COSMIC pipeline and summarises the validation outcomes for each component.

6.8.1 Diagnostic protocol

The protocol consists of five mandatory criteria, evaluated for every scalar parameter in every model:

1. *Rank-normalised split-$\hat{R}$.* The between-chain and within-chain variances are compared after rank-normalising the samples and splitting each chain in half. Convergence requires $\hat{R} < 1.01$ for all parameters (Vehtari et al., 2021). Values exceeding this threshold indicate that the chains have not converged to the same stationary distribution and that the posterior summaries are unreliable.

2. *Bulk effective sample size.* Bulk-ESS estimates the number of independent draws available for central tendency statistics (mean, median). We require Bulk-ESS > 400 to ensure that the Monte Carlo standard error on the posterior mean is negligible relative to the posterior width.

3. *Tail effective sample size.* Tail-ESS quantifies the sampling efficiency in the fifth and ninety-fifth percentiles of the posterior, governing the reliability of credible interval endpoints. The same threshold of 400 is imposed.

4. *Divergent transitions.* A divergent transition is a rigorous diagnostic that the leapfrog integrator has encountered extreme curvature and failed to conserve the simulated Hamiltonian energy (Betancourt, 2016). Any model producing divergences is considered formally compromised. We require zero divergent transitions.

5. *Energy Bayesian Fraction of Missing Information (E-BFMI).* The E-BFMI quantifies the efficiency of the momentum resampling step in HMC; values below 0.3 indicate that the sampler lacks sufficient kinetic energy to traverse the posterior breadth (Betancourt, 2016). We require E-BFMI > 0.3.

Parameters failing any criterion are flagged, and the corresponding model is revised: chains are extended, priors are tightened or relaxed, or the parameterisation is modified (e.g., non-centred reparameterisation for hierarchical funnels; Section 4.1.4). No posterior summary is reported in Chapter 7 until all five criteria are satisfied.

6.8.2 Visual diagnostics

Numerical diagnostics are supplemented by visual inspection of rank-normalised trace plots for each parameter, which can reveal non-stationarity patterns—slow drifts, mode-switching, periodic oscillations—that summary statistics may obscure (Gabry et al., 2019). The trace plots for the distance, proper-motion, structural, and isochronal models were inspected manually; no pathological patterns were identified.

6.8.3 Diagnostic outcomes by component

All four NUTS-based models in the COSMIC pipeline—the hierarchical distance model (Section 6.3), the proper-motion model (Section 6.4), the King profile model (Section 6.5), and the ASteCA isochronal fit (Section 6.6)—satisfied all five diagnostic criteria without requiring reparameterisation. The NUTS warm-up phase (typically 1 000 draws) was sufficient to tune the step-size and mass matrix for all models. Each model was sampled with four independent chains of 2 000 post-warm-up draws, yielding 8 000 total posterior samples per parameter.

The isochronal model exhibited the most complex posterior geometry, with strong covariance between age, extinction, and distance modulus. The NUTS trajectory length adapted to this covariance structure during warm-up, and the resulting chains showed no evidence of incomplete exploration. The rank-normalised $\hat{R}$ for

all four isochronal parameters (log(age/yr), Z, A_V, μ_{DM}) was below 1.005, and the Bulk-ESS exceeded 1 500 for each parameter.

6.8.4 Posterior predictive assessment

The posterior predictive framework of Section 4.1.7 was applied to the isochronal model as a test of model adequacy. Replicated CMDs were generated by drawing parameter values from the fitted posterior and simulating synthetic photometry through the full forward model. The observed CMD morphology was compared to the posterior predictive replicates to identify systematic discrepancies.

The posterior predictive replicates reproduced the upper main sequence ($G < 12$ mag) and the general morphology of the turn-on region. Below the turn-on, the single-isochrone model predictably failed to capture the full vertical extent of the PMS population, which spans a range of ages not representable by a single epoch. This discrepancy is expected and is addressed by the Sagitta age analysis (Section 6.7), which provides individual PMS ages that complement the single-epoch isochronal posterior.

6.9 Failure modes and operational limits

The COSMIC pipeline, as applied to NGC 6383, carries several operational limitations that constrain the generalisability and precision of its outputs. These are not incidental deficiencies; they define the boundary within which the physical claims of Chapter 7 are valid, and they are documented here as binding conditions on the inference.

6.9.1 Proper-motion-only membership

The membership pipeline operates exclusively in two-dimensional proper-motion space. It does not incorporate parallax, photometry, or spatial position into the HDBSCAN clustering step; parallax enters only through the post-hoc sigma-clipping of Section 6.2.4, and photometric information is entirely absent from the membership construction.

This design choice has two consequences. First, the membership inference does not exploit the full information content of the Gaia five-parameter astrometric solution: sources with concordant parallaxes but discrepant proper motions are rejected, while sources with concordant proper motions but discrepant parallaxes are admitted and must be removed by the sigma-clipping step. A multi-dimensional membership model operating jointly in proper motion, parallax, and colour space would provide more discriminating membership probabilities and would eliminate the need for the deterministic clipping stage. Second, the membership inference is insensitive to photometric membership indicators such as the CMD locus: a field star whose proper motion coincidentally matches the cluster but whose CMD position is inconsistent with any plausible cluster age and metallicity passes the membership test undetected.

The replacement of the two-stage HDBSCAN + clipping architecture with a single-stage, multi-dimensional generative mixture model of the form Eq. (4.6) is the most consequential improvement available for future COSMIC versions.

6.9.2 Pseudoprobability calibration

The pseudoprobability $\tilde{p}_i$ of Eq. (4.16) is a useful ranking metric but is not a calibrated posterior probability. It combines a geometric persistence ratio ($p_{\mathrm{HDBSCAN},i}$) with a frequentist iteration fraction (f_i), neither of which is conditioned on a

generative model of the cluster and field populations. A $\tilde{p}_i = 0.8$ does not mean that the data are four times more probable under the cluster model than under the field model; it means that the source survived 80% of HDBSCAN configurations and resides moderately deep within the persistent density core.

The consequence is that the membership tiers defined in Section 6.2.3 are operationally convenient but inferentially informal. Physical summaries computed as expectations over the membership—such as the population mean distance or the mass function—should ideally be computed under calibrated posterior probabilities that account for the field-star distribution and the measurement uncertainties (Section 4.2.6). The current implementation uses hard membership thresholds as a practical surrogate, and the threshold sensitivity analysis of Chapter 7 quantifies the impact of this approximation.

6.9.3 Distance model restriction

The hierarchical distance model restricts itself to sources with fractional parallax errors $\sigma_\varpi/\varpi < 0.10$, excluding approximately 40% of the membership. The excluded sources are predominantly faint, low-mass PMS stars whose parallax likelihoods are broad relative to any plausible distance prior. Their inclusion would couple the cluster distance posterior more strongly to the prior without improving the mean estimate. However, excluding them prevents the hierarchical model from providing individual distance posteriors for these sources, which would be useful for propagating distance uncertainty into the photometric and structural analyses.

A fully hierarchical model that retains all members and uses the population-level distance as a shared prior for the individual-level distances—as formalised in Eq. (4.12)—would avoid the need for this exclusion and would automatically

provide the hierarchical shrinkage described in Section 4.2.4.

6.9.4 Zero differential reddening

The ASteCA isochronal fit assumes spatially uniform extinction across the cluster field. As discussed in Section 4.5.2, differential reddening broadens the main sequence perpendicular to the reddening vector and can bias the age and extinction estimates. For NGC 6383, which lies within the H II region Sh 2-012, spatially variable extinction is physically expected. The assumption of zero differential reddening introduces a systematic uncertainty in the age and extinction posteriors that is not captured by the statistical credible intervals.

A spatially resolved extinction model—for instance, a Gaussian process over the cluster field that predicts $A_V(\alpha, \delta)$ and is marginalised during the isochronal fit (Edenhofer et al., 2024)—would remove this assumption and provide per-star extinction estimates. This extension is deferred to future work.

6.9.5 Single isochrone grid

The isochronal fit uses the MIST grid exclusively. The grid-to-grid variance—the systematic offset in the recovered age, metallicity, and extinction when the same data are fitted with PARSEC, Baraffe, or SPOTS isochrones—is not propagated into the reported uncertainties. As discussed in Section 4.5.6, this variance can exceed the statistical precision for young clusters and represents a hard systematic floor on the achievable age accuracy. The reported age uncertainty should therefore be understood as the statistical precision under the MIST assumption, not the total uncertainty inclusive of theoretical model systematics.

6.9.6 Radial velocity sparsity

The Gaia DR3 radial velocity coverage for the cluster ($\sim$11.4%; Section 5.5.3) is insufficient for three-dimensional kinematic modelling. The reported radial velocity is a population summary computed from 16 stars, not a dynamically meaningful velocity dispersion measurement. The virial state of NGC 6383, the internal rotation profile, and the three-dimensional expansion rate all require radial velocities for a larger, mass-representative subsample. Until such data become available—through future Gaia releases or dedicated spectroscopic programmes—the dynamical state assessment of Chapter 7 is based exclusively on proper-motion kinematics and structural parameters.

6.9.7 2MASS Completeness

The PMS classification and YSO analysis depend on 2MASS photometry, which is available for only $\sim$35% of the Gaia working sample. The PMS census is therefore truncated at the 2MASS detection limit ($J \approx 15.8$ mag), corresponding to $G \approx 16$–17 mag at the cluster distance and extinction. The lowest-mass, faintest PMS stars that Gaia detects in the optical are absent from the PMS and YSO analyses. The YSO fraction reported in Chapter 7 is conditioned on the 2MASS-matched subsample and cannot be directly extrapolated to the full membership without assumptions about the mass dependence of disc survival.

6.10 Reproducibility and data release

Scientific claims that cannot be independently reproduced are, in a strict epistemological sense, untestable. The COSMIC framework is designed with reproducibility as a structural requirement: every algorithmic step, hyperparameter choice, and

prior specification is documented in this chapter and encoded in the released software and data products.

6.10.1 Software release

The COSMIC source code is publicly available under an open-source licence at `https://github.com/notluquis/COSMIC`. The repository contains the complete pipeline: data ingestion and quality filtering, HDBSCAN membership determination with pseudoprobability construction, the PyMC 5 implementations of the hierarchical distance model, the proper-motion model, the King profile model, and the modified ASteCA interface with the NUTS sampler. The Sagitta interface and the YSO classification module are included as auxiliary components.

The repository specifies all software dependencies and their versions (Python 3.12, PyMC 5, SCIKIT-LEARN, HDBSCAN, ASTROPY, ASTECA), ensuring that the computational environment can be reconstructed. The HDBSCAN hyperparameter grid, the NUTS sampler configuration (number of chains, draws, warm-up, target acceptance rate), and all prior specifications are defined as configuration parameters that can be modified without altering the source code.

6.10.2 Data products

The membership catalogue, including the full pseudoprobability vector $\tilde{p}_i$ and the derived physical parameters for all 321 sources with $\tilde{p}_i > 0.5$, will be archived at the Centre de Données astronomiques de Strasbourg (CDS). The catalogue includes, for each source: the Gaia DR3 source identifier, corrected astrometry, Gaia and 2MASS photometry, HDBSCAN persistence value, iteration fraction, pseudoprobability, membership tier, and all derived quantities (ASteCA mass, binary probability, Sagitta PMS probability and age, Q-parameter classification).

The posterior samples for all four Bayesian models—distance, proper motion, King profile, and isochronal fit—are included in the data release as serialised ArviZ InferenceData objects. These objects contain the full MCMC traces, the prior and posterior predictive distributions, and the convergence diagnostics, enabling re-analysis under alternative prior specifications or summary statistics without re-running the sampler.

6.10.3 Reproducibility scope and limitations

The released materials permit exact reproduction of all results reported in Chapter 7, given the same input data (Gaia DR3 and 2MASS catalogues retrieved as of the query date). Two aspects of the analysis are not exactly reproducible from the released code alone. First, the HDBSCAN algorithm involves stochastic tie-breaking in the minimum spanning tree construction when edge weights are degenerate; the membership pseudoprobabilities may vary by ~ 0.01 across independent runs. Second, MCMC sampling is stochastic by construction; the posterior samples differ between runs, though the posterior summaries converge to the same values within Monte Carlo error. Both sources of variability are negligible relative to the reported uncertainties.

The COSMIC framework is not specific to NGC 6383. The modular architecture permits application to any open cluster for which Gaia astrometry and broadband photometry are available, and the prior specifications are parameterised to accept cluster-specific values. The extension of COSMIC to a systematic survey of young open clusters—identified as a future application in Section 8.5—is enabled by this design.

Chapter 7

Physical Characterisation of NGC 6383

This chapter reports the physical properties of NGC 6383 as inferred by the COSMIC framework of Chapter 6, operating on the observational dataset of Chapter 5. Every quantity reported below is a posterior summary conditioned on the adopted membership construction, generative model, and prior specification, and is subject to the claim-status discipline of Section 3.5. We present the membership census and its threshold sensitivity (Section 7.1), the distance and kinematic parameters (Section 7.2), the structural profile (Section 7.3), the stellar population and photometric properties including the pre-main-sequence population and young stellar objects (Section 7.4), the evidence for mass segregation (Section 7.5), the dynamical state assessment (Section 7.6), and the membership status of the O-type binary HD 159176 (Section 7.7). The principal derived parameters are collected in Table 7.3.

Uncertainty convention. Unless otherwise stated, all uncertainties reported in this chapter and in Table 7.3 represent the posterior standard deviation (for

approximately symmetric posteriors) or the 16th–84th percentile credible interval (for asymmetric posteriors) of the relevant parameter's marginal posterior distribution. For population-level quantities such as the mean parallax and mean proper motions, the reported uncertainty is the posterior standard deviation of the population-level hyperparameter, which accounts for the sample size and the individual measurement uncertainties through the hierarchical model structure; it should not be confused with the intrinsic dispersion of the member distribution, which is reported separately where relevant. The Gaia systematic floor on the mean parallax—arising from spatially correlated zero-point residuals at the $\sim$10–15 μas level (Lindegren et al., 2021b)—is not propagated into the formal uncertainty but is noted as an additional systematic contribution to the distance error budget (Section 7.2.1).

Membership construction and inferential status. The membership entering this chapter is constructed through the HDBSCAN-based pseudoprobability $\tilde{p}_i$ of Section 6.2.3, thresholded at $\tilde{p}_i > 0.6$. As established in Section 6.9.2, $\tilde{p}_i$ is *not* a calibrated posterior membership probability in the sense of Eq. (4.7): it combines a density-persistence ratio with an iteration-fraction statistic and carries no generative likelihood calibration. The formal framework of Chapters 3–4 requires that physical summaries be computed as expectations over the full posterior membership distribution; the operational implementation instead partitions the sample at a fixed threshold and treats the resulting census as a conventional member list. This gap between the probabilistic aspiration and the operational practice is a deliberate scope decision documented in Section 6.9.2, and its quantitative consequences are bounded by the threshold-sensitivity analysis of Section 7.1.2, which shows how the recovered population and its derived properties vary across four probability thresholds ($\tilde{p}_i > 0.5$, 0.6, 0.7, 0.8). All physical quantities reported below are conditioned on the $\tilde{p}_i > 0.6$ reference sample

unless explicitly stated otherwise, and should be read as membership-conditioned summaries rather than as unconditional population parameters.

7.1 Membership census

7.1.1 HDBSCAN recovery and parallax refinement

The HDBSCAN pipeline of Section 6.2, applied to the proper-motion space of the 15 276 sources defined in Section 5.4, returned several candidate clusters. The dominant persistent component, corresponding to NGC 6383, initially contained 701 sources. After applying the 2σ parallax clipping of Section 6.2.4—which removed kinematic interlopers whose parallaxes deviate from the mode of the cluster parallax distribution—321 sources survived at the $\tilde{p}_i > 0.5$ threshold.

The condensed cluster tree (Fig. 7.1) shows the NGC 6383 cluster as the primary branch on the left side of the dendrogram, with the HDBSCAN-identified field population occupying the complementary branch on the right. The two populations separate cleanly at a λ value consistent with the density contrast between the kinematically cold cluster core ($\sigma_\mu \sim 0.15$ mas yr^{-1}) and the kinematically hot Galactic field.

7.1.2 Threshold sensitivity

Applying the pseudoprobability tiers defined in Section 6.2.3, we obtained the membership counts shown in Table 7.1.

The range 321–161 across four thresholds demonstrates the uncertainty structure that the continuous membership formalism of Section 4.2 is designed to capture. Each threshold produces a qualitatively different picture of the cluster's spatial extent, luminosity function, and mass distribution. As established in Section 4.2.3,

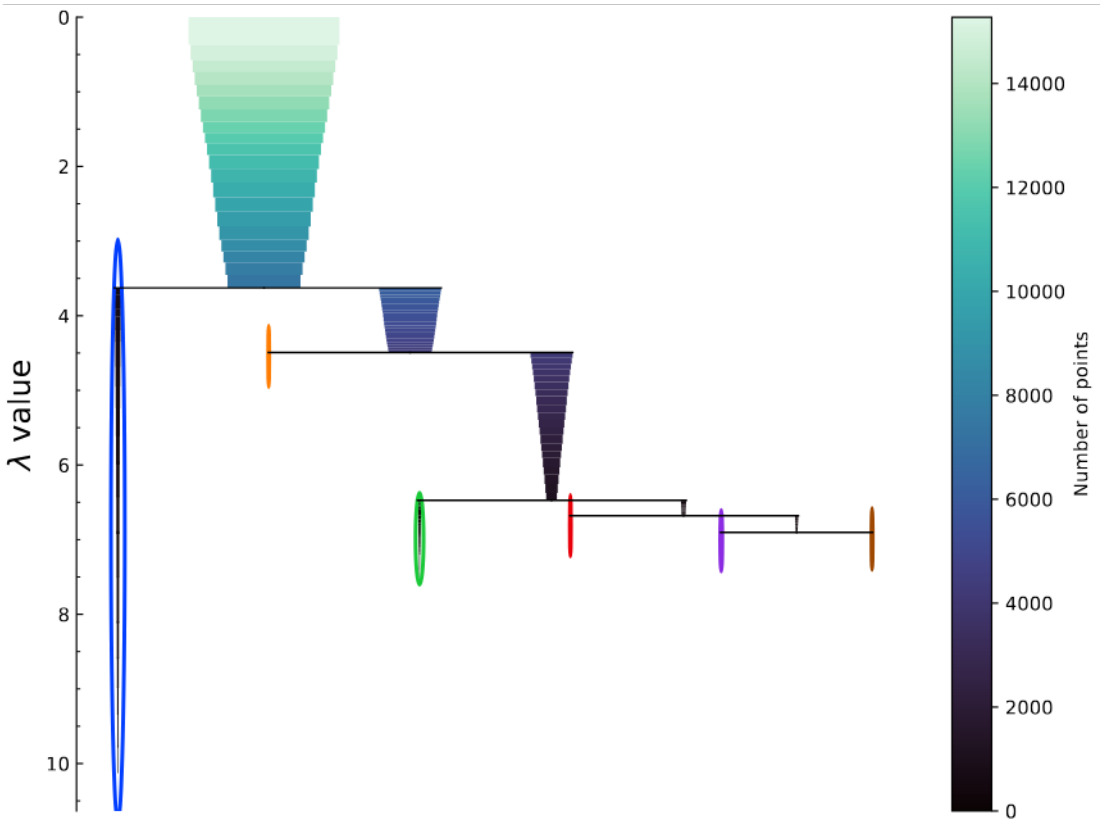

Figure 7.1: Condensed cluster tree showing the hierarchical structure of the cluster system identified by HDBSCAN. The dendrogram displays the cluster hierarchy, where the width and colour of each branch represent the number of sources at that level. The NGC 6383 cluster sources appear on the left, while the HDBSCAN-identified field population occupies the right branch. The colour bar indicates the number of sources at each level, and the λ value corresponds to $1/\texttt{distance}$.

Table 7.1: Membership counts as a function of pseudoprobability threshold $\tilde{p}_{\rm thr}$. The column $N(G < 19)$ restricts to sources within the regime of reliable Gaia astrometry (Section 5.5.2).

$\tilde{p}_{\rm thr}$	$N_{\rm members}$	$N(G < 19$ mag)
0.5	321	288
0.6	254	236
0.7	202	191
0.8	161	153

a single binary membership list does not capture the population-level uncertainty; the threshold-dependent counts quantify the practical extent of this indeterminacy.

The 67 sources between the $\tilde{p}_i > 0.5$ and $\tilde{p}_i > 0.6$ thresholds are predominantly faint ($G > 17$ mag), spatially peripheral, and kinematically marginal—their proper motions lie near the boundary between the cluster and field distributions. These are the sources whose inclusion or exclusion most affects the cluster periphery, the low-mass end of the luminosity function, and the outer structural profile.

We adopt the $\tilde{p}_i > 0.6$ threshold (254 sources: 161 members and 93 probable members) as the reference sample for the physical characterisation that follows. Of these, 236 are brighter than $G = 19$ mag and lie within the regime of reliable Gaia astrometry (Section 5.5.2). The remaining 18 fainter sources carry larger astrometric uncertainties and contribute primarily through the hierarchical shrinkage mechanism of Section 4.2.4. The membership probability distribution of the reference sample is shown in Fig. 7.2, colour-coded by astrometric fidelity; the expected trend of decreasing fidelity with increasing magnitude is evident for $G > 18$ mag.

7.1.3 Cross-comparison with published catalogues

We cross-compared our 254-member reference sample against five published Gaia-era membership catalogues to assess the consistency and complementarity of different membership methodologies (Table 7.2).

The differences arise primarily from the distinct membership methodologies and threshold conventions adopted in each study, rather than from fundamental disagreements in the kinematic signature of the cluster. Hunt and Reffert (2024) employed a similar HDBSCAN-based approach and produce the highest overlap (202 common sources, 79% of our sample), while He et al. (2022), who applied

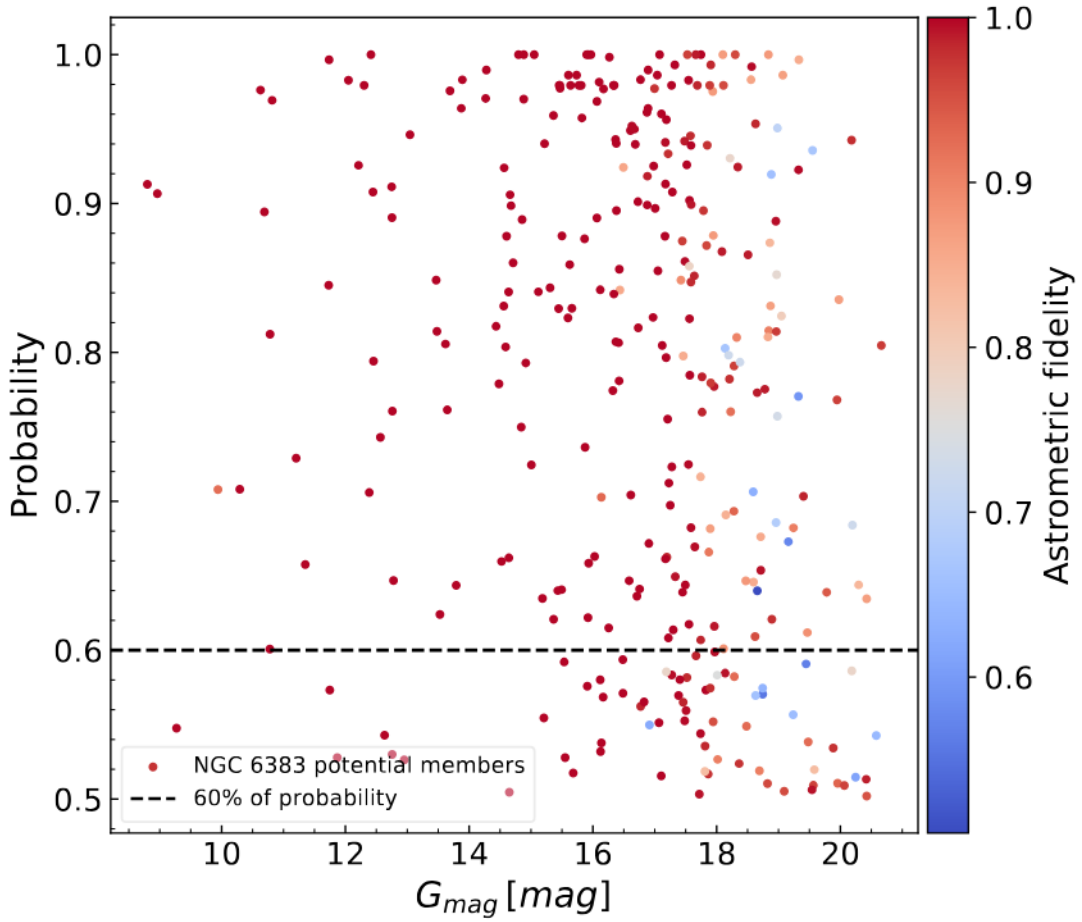

Figure 7.2: Membership probability of potential NGC 6383 members (membership probability $> 50\%$) plotted against their G magnitudes, with colours representing astrometric fidelity. The colour gradient from blue to red reflects the varying reliability of astrometric measurements. The horizontal dashed line at a probability of 0.6 marks the threshold used to distinguish between probable members (membership probabilities between 60% and 80%) and members (membership probabilities $\geq 80\%$). Stars with fainter magnitudes tend to have lower astrometric fidelity, becoming more evident at $G > 18$ mag.

Table 7.2: Cross-comparison of the 254-member reference sample ($\tilde{p}_i > 0.6$) against published Gaia-era membership catalogues. Columns list the number of common sources, sources unique to each published catalogue, and sources unique to the present work.

Reference	Common	Unique to them	Unique to us
Cantat-Gaudin et al. (2020)	148	97	173
Jaehnig, Bird, and Holley-Bockelmann (2021)	161	123	160
He et al. (2022)	90	47	231
Hunt and Reffert (2024)	202	120	119
SIMBAD	168	163	153

a different clustering algorithm with a different feature space, show the lowest overlap (90 sources, 35%). The substantial number of sources unique to our catalogue (119–231 depending on the comparison) reflects the aggressive leaf-extraction strategy and the pseudoprobability construction of Section 6.2.3, which were designed to recover the extended halo membership at the cost of a broader initial candidate set.

These cross-comparison statistics reinforce the finding of Section 4.2.3: the recovered population is not a property of the cluster alone but of the interaction between the cluster, the field, the survey, and the method. Discrepancies between catalogues are the expected consequence of this interaction and do not, by themselves, indicate that any individual catalogue is incorrect.

7.2 Distance and kinematics

7.2.1 Parallax and distance

The mean parallax of the 254-member reference sample is $\bar{\varpi} = 0.908 \pm 0.004$ mas. The hierarchical distance model of Section 6.3, restricted to sources with fractional parallax errors $\sigma_\varpi/\varpi < 0.10$, yields a posterior mode distance of

$$d = 1.110 \pm 0.060 \text{ kpc}\,. \tag{7.1}$$

The parallax distribution of the reference sample is shown in Fig. 7.3 (left panel), decomposed into sources used for the distance estimation ($\sigma_\varpi/\varpi < 0.10$, labelled "U. parallax" in the figure) and those excluded due to fractional errors exceeding the threshold ("N.U. Parallax", i.e. not used). The Gaussian fit to the used-parallax subsample peaks at $\mu_\varpi = 0.908 \pm 0.004$ mas. The middle panel shows the G magnitude versus parallax, confirming that the excluded sources are concentrated

at $G > 17$ mag where the astrometric precision degrades (Section 5.5.2).

The right panel of Fig. 7.3 shows the geometric distance distribution from Bailer-Jones et al. (2021), with the mode at 1.110 kpc and the standard deviation of the sampled distance distribution at 0.060 kpc. The consistency between the parallax-based hierarchical estimate and the external geometric distance provides confidence that the parallax zero-point correction of Section 5.2.1 is adequate and that the distance prior of Eq. (6.2) is not dominating the posterior.

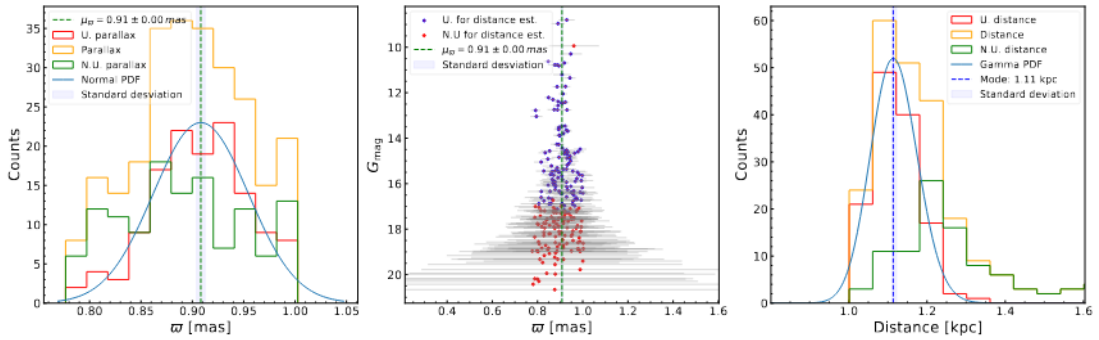

Figure 7.3: *Left panel:* Distributions of the parallax measurements in NGC 6383. *U. parallax* (red) represents measurements with fractional parallax errors less than 0.1, used for the parallax estimation; *Parallax* (orange) represents all the data; and *N.U. Parallax* (green) represents sources with fractional parallax error greater than 0.100, not used for the estimation. The blue curve shows a Gaussian fit to the *U. parallax* distribution, with the mean parallax $\mu_\varpi = 0.908 \pm 0.004$ mas marked by a dashed green line. *Middle panel:* G magnitude versus parallax, with sources used for the distance estimation in blue and those discarded in red. Grey error bars illustrate the uncertainty in parallax measurements. *Right panel:* Histograms of geometric distances derived from Bailer-Jones et al. (2021), distinguishing between used (red) and all measured (orange) parallaxes. The central dashed blue line indicates the mode of the sampled distance (1.110 kpc), with its standard deviation of 0.060 kpc shaded in blue.

An independent distance estimate is provided by the isochronal distance modulus $\mu_{\rm DM} = 10.3 \pm 0.26$ mag from the ASteCA fit (Section 6.6), corresponding to $d = 1.15^{+0.15}_{-0.13}$ kpc. This value is consistent with the parallax-based distance within the mutual uncertainties, confirming internal coherence across the two independent inference pathways.

7.2.2 Proper motions

The posterior mean proper motions of the cluster are

$$\overline{\mu_{\alpha*}} = 2.540 \pm 0.010 \text{ mas yr}^{-1}, \tag{7.2}$$

$$\overline{\mu_{\delta}} = -1.710 \pm 0.009 \text{ mas yr}^{-1}, \tag{7.3}$$

with intrinsic dispersions of $\sigma_{\mu_\alpha} = 0.153$ and $\sigma_{\mu_\delta} = 0.138$ mas yr^{-1}.[1] These dispersions are consistent with the internal kinematics expected for a young open cluster in the pre-relaxation regime at $d \approx 1.1$ kpc (Pang et al., 2023). At this distance, a dispersion of 0.15 mas yr^{-1} corresponds to ~ 0.8 km s^{-1} in the plane of the sky, consistent with the velocity dispersions of other Gaia-era young open clusters. The perspective expansion signal—the apparent proper-motion gradient induced by the projection of the radial velocity onto the plane of the sky (Section 4.3.5)—is ~ 0.002 mas yr^{-1} at the half-mass radius for $v_r \approx -6$ km s^{-1} and $d \approx 1.1$ kpc, negligible compared to the intrinsic dispersion.

The proper-motion distribution is shown in Fig. 7.4, colour-coded by membership probability and overlaid with density contours. The cluster population appears as a compact, approximately Gaussian concentration centred on $(\mu_{\alpha*}, \mu_\delta) = (2.54, -1.71)$ mas yr^{-1}, cleanly separated from the broader field distribution.

The total proper-motion amplitude of the cluster on the plane of the sky is $v_{\text{proj}} = 3.070 \pm 0.010$ mas yr^{-1}, obtained from Eq. (6.4) as the magnitude of the systemic proper-motion vector. At the adopted distance, this corresponds to a tangential velocity of ~ 16.2 km s^{-1}. The distribution of individual proper-motion magnitudes for the reference sample is shown in Fig. 7.5, together with the Gaussian model and the posterior mean; the width of this distribution reflects

[1] The posterior standard deviations on the mean proper motions are 0.0096 mas yr^{-1} for $\overline{\mu_{\alpha*}}$ and 0.0086 mas yr^{-1} for $\overline{\mu_\delta}$ respectively; we report the rounded values 0.010 and 0.009 mas yr^{-1} throughout for notational uniformity with the summary table (Table 7.3).

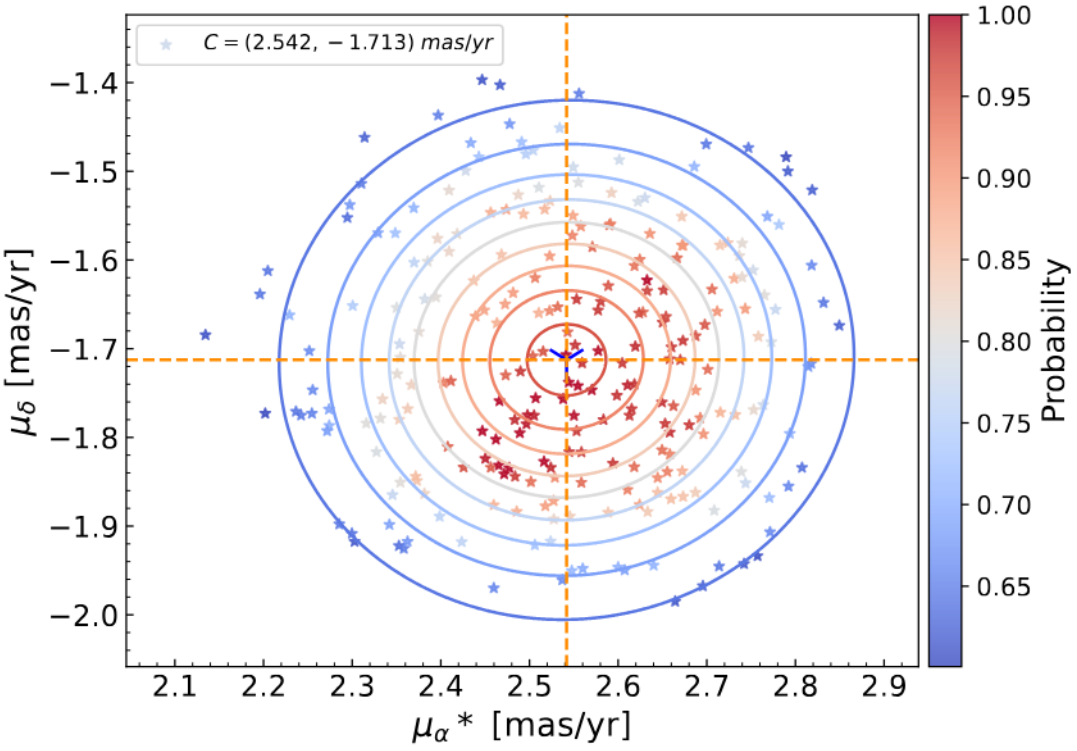

Figure 7.4: Proper motions of probable members and members of NGC 6383 in Right Ascension (μ_α^*) and Declination (μ_δ), colour-coded by membership probability. Hues transition from blue (lower probabilities) to red (higher probabilities), overlaid with density contours. The centre of the proper-motion distribution is marked with a blue cross at $(\mu_\alpha^*, \mu_\delta) = (2.54, -1.71)$ mas yr^{-1}, with corresponding orange dashed lines.

the intrinsic velocity dispersion convolved with the measurement uncertainties.

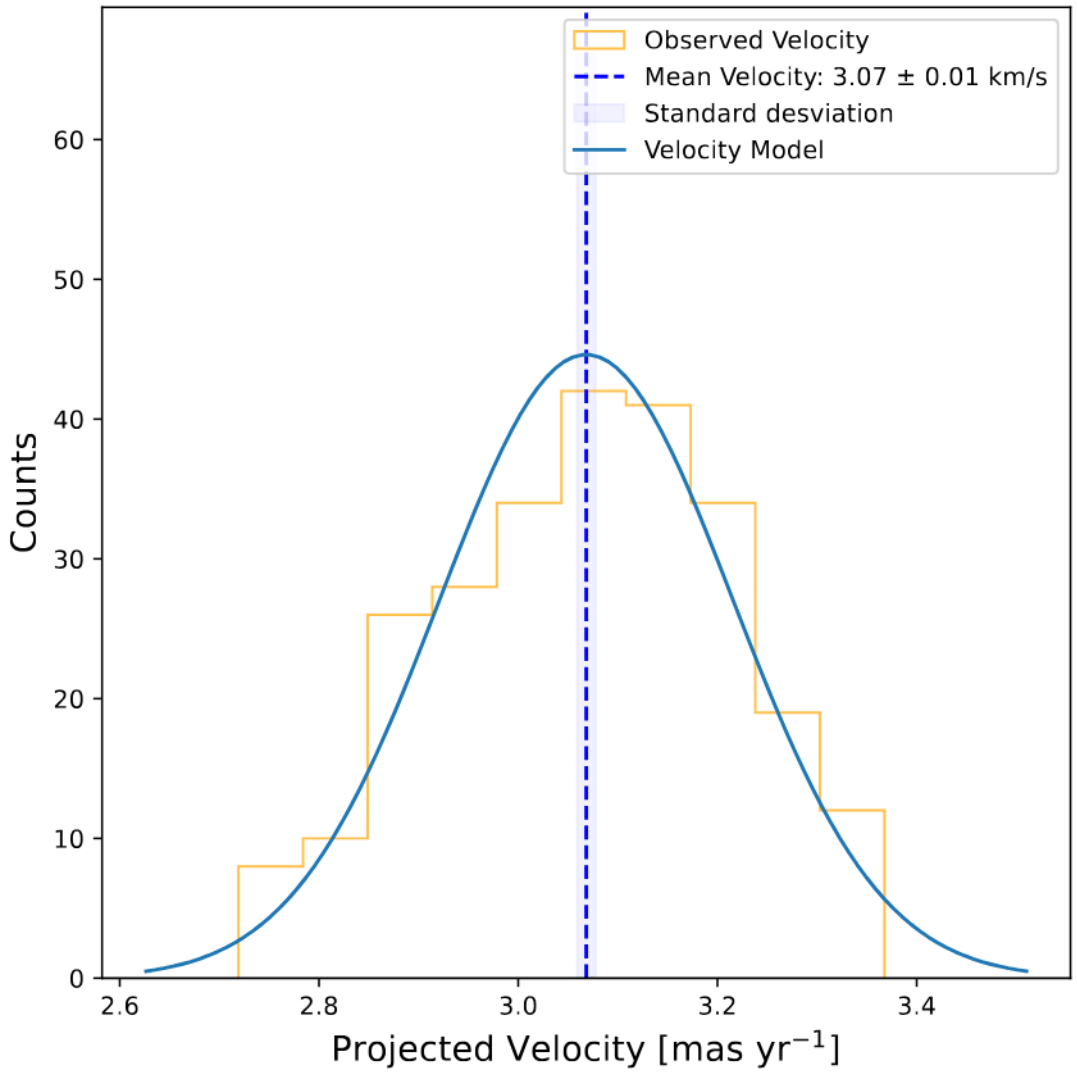

Figure 7.5: Distribution of projected velocities $v_{\rm proj} = \sqrt{\mu_{\alpha*}^2 + \mu_\delta^2}$ for the reference sample of NGC 6383 members. The orange histogram shows the observed velocity distribution; the solid blue curve shows the Gaussian model fit. The dashed blue line marks the mean projected velocity at $3.07 \pm 0.01\,\mathrm{mas\,yr^{-1}}$, with the standard deviation shaded in blue.

7.2.3 Radial velocities

Gaia DR3 radial velocities are available for 29 members with $\tilde{p}_i > 0.6$, of which 16 have binary probability below 0.6 and were used for the population-level radial

velocity summary. The resulting statistics are: median $v_r = -6.11$ km s^{-1}, mean $v_r = -15.1$ km s^{-1}, standard deviation $\sigma_{v_r} = 31.1$ km s^{-1}.

The large standard deviation reflects three compounding factors. First, the sample of 16 sources is too small to suppress the statistical noise introduced by individual velocity outliers. Second, the binary probability threshold of 0.6 does not eliminate all unresolved spectroscopic binaries; sources with binary probability between 0.4 and 0.6 may still carry significant orbital velocity contributions that inflate the apparent dispersion. Third, the Gaia RVS sample is magnitude-biased toward the bright, high-mass population (Section 5.5.3), so the derived dispersion does not represent the mass-averaged kinematics of the full cluster.

The radial velocity amplitude—the total amplitude of the radial velocity time series after outlier removal, available for a subset of sources—provides an additional diagnostic for binary contamination. Sources with high radial velocity amplitude are likely unresolved binaries. Figure 7.6 shows the radial velocity versus G magnitude for all members with available data, colour-coded by radial velocity amplitude where available.

We report the median radial velocity (-6.11 km s^{-1}) as the most robust population summary, but caution that it does not constitute a measurement of the cluster's internal velocity dispersion. The virial state assessment of Section 7.6 relies on proper-motion kinematics and structural parameters rather than radial velocities.

7.2.4 Galactocentric orbit

The Galactocentric distance of NGC 6383 is $R_{\rm GC} = 7.19 \pm 0.07$ kpc, adopted from Hunt and Reffert (2024) and consistent with the value computed from the thesis distance ($d = 1.110$ kpc) and the solar Galactocentric radius $R_\odot = 8.178$ kpc. The uncertainty includes the contributions from the cluster distance (± 0.060 kpc)

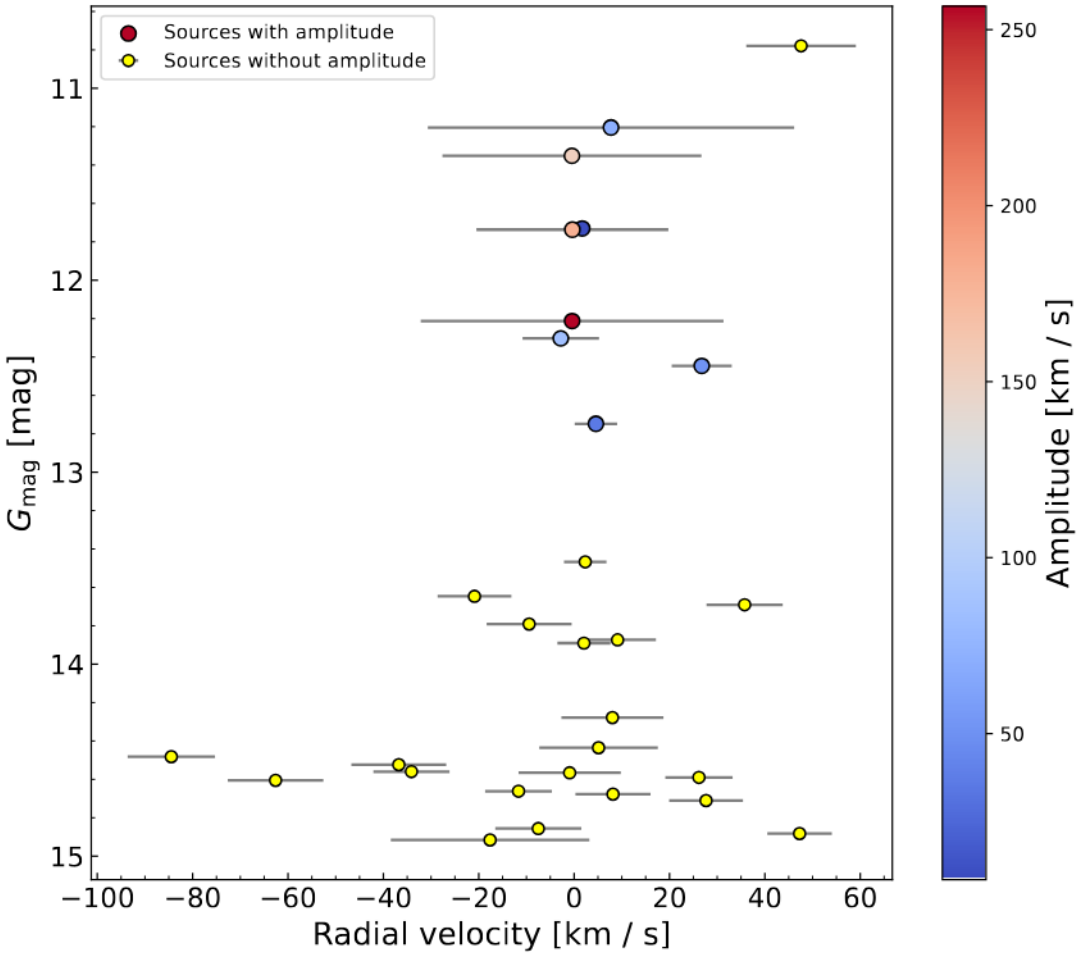

Figure 7.6: Radial velocity plotted against G-band magnitude for NGC 6383 sources, categorised by the availability of radial velocity amplitude data from Gaia. Yellow circles indicate sources without amplitude measurements, with radial velocity uncertainties shown as error bars. Blue circles represent sources with amplitude data, colour-coded by amplitude value according to the scale on the right.

and the solar Galactocentric distance (± 0.035 kpc) added in quadrature. The cluster lies ~ 1.0 kpc interior to the solar circle. At this Galactocentric radius, the cluster orbits within the Carina–Sagittarius arm, consistent with its association with the Sgr OB1 complex and the Sh 2-012 star-forming region (Rauw and De Becker, 2008). The Galactocentric position enters the Hill radius calculation of Section 6.5.4 and the tidal boundary assessment of Section 7.3.

7.3 Structural parameters

7.3.1 King profile fit

The Bayesian King profile fit of Section 6.5.2 yields the following posterior mode parameters for the 254-source reference sample ($\tilde{p}_i > 0.6$):

$$R_c = 1.95 \pm 0.19 \text{ arcmin} \quad (0.63 \pm 0.07 \text{ pc}), \tag{7.4}$$

$$R_t = 40.4 \pm 14.3 \text{ arcmin} \quad (13.1 \pm 4.7 \text{ pc}), \tag{7.5}$$

$$k = 4.91 \pm 0.44 \text{ stars arcmin}^{-2}, \tag{7.6}$$

$$b = 0.011 \pm 0.006 \text{ stars arcmin}^{-2}. \tag{7.7}$$

The radial density profile is shown in Fig. 7.7, where the observed annular densities (blue points) are overlaid with the posterior King model (solid fuchsia line) and its 1σ uncertainty envelope (blue shading). The core radius and tidal radius are marked by dashed and dotted red lines respectively, and the background level b by a dash-dot green line. Visual inspection of the residuals does not reveal systematic trends as a function of radius; in particular, the outer annuli beyond R_c do not exhibit the systematic positive excess that would indicate a non-King extended envelope or the need for a more complex structural model (e.g. EFF or Wilson profiles).

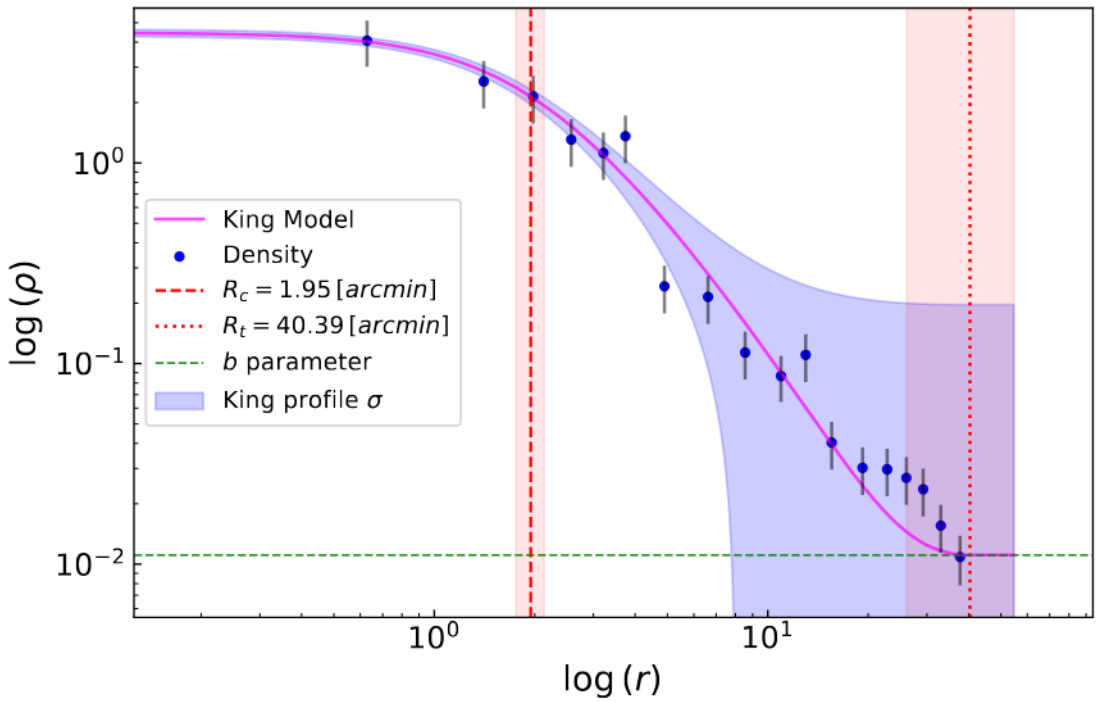

Figure 7.7: Radial density profile of NGC 6383 in logarithmic scale. The observed stellar density (blue points) is overlaid with the King model (solid fuchsia line). Dashed red and dotted red lines mark the core radius ($R_c = 1.95$ arcmin) and the tidal radius ($R_t = 40.4$ arcmin), respectively. The dash-dot green line represents the background level b. The shaded area in blue indicates the 1σ uncertainty range of the King profile.

The core radius is well constrained ($\sim 10\%$ relative uncertainty), as it is anchored by the high-density inner annuli where the King profile is steeply curved and the photometric completeness is high. The tidal radius, by contrast, carries a relative uncertainty of $\sim 35\%$. This large uncertainty reflects the difficulty of constraining the tidal boundary for a cluster embedded in a dense Galactic field: at large radii, the cluster density approaches the background level (b), and the King profile's outer slope becomes degenerate with the background subtraction. A $\sim 20\%$ change in the adopted background density shifts the best-fit tidal radius by several arcminutes. The tidal radius should therefore be interpreted as a model-dependent estimate rather than a sharply defined physical boundary, consistent with the discussion of classical truncation limits in Section 2.4.1. An additional source of uncertainty is that the posterior mode of R_t (40.4 arcmin) slightly exceeds the search cone radius (40.0 arcmin; Section 5.1.2), so the outermost portion of the King profile is extrapolated beyond the observed data boundary. The background level b is estimated from the outer annuli of the cone, which may still contain cluster members at the density predicted by the King model; this coupling contributes to the R_t–b covariance and reinforces the model-conditioned status of the tidal radius estimate. The cluster centre, determined through the weighted KDE of Section 6.5.1, is located at $(\alpha, \delta) = (263.683 \pm 0.112°,\ -32.584 \pm 0.112°)$, where the uncertainty ($0.112° \approx 6.7$ arcmin) is dominated by the KDE bandwidth optimised through cross-validation rather than by the statistical precision of the density peak location; the latter is substantially smaller for a sample of 254 sources. The spatial distribution of the reference sample around this centre is shown in Fig. 7.8, where the density gradient of the membership is visible as a central concentration surrounded by a progressively sparser envelope extending to the cluster outskirts.

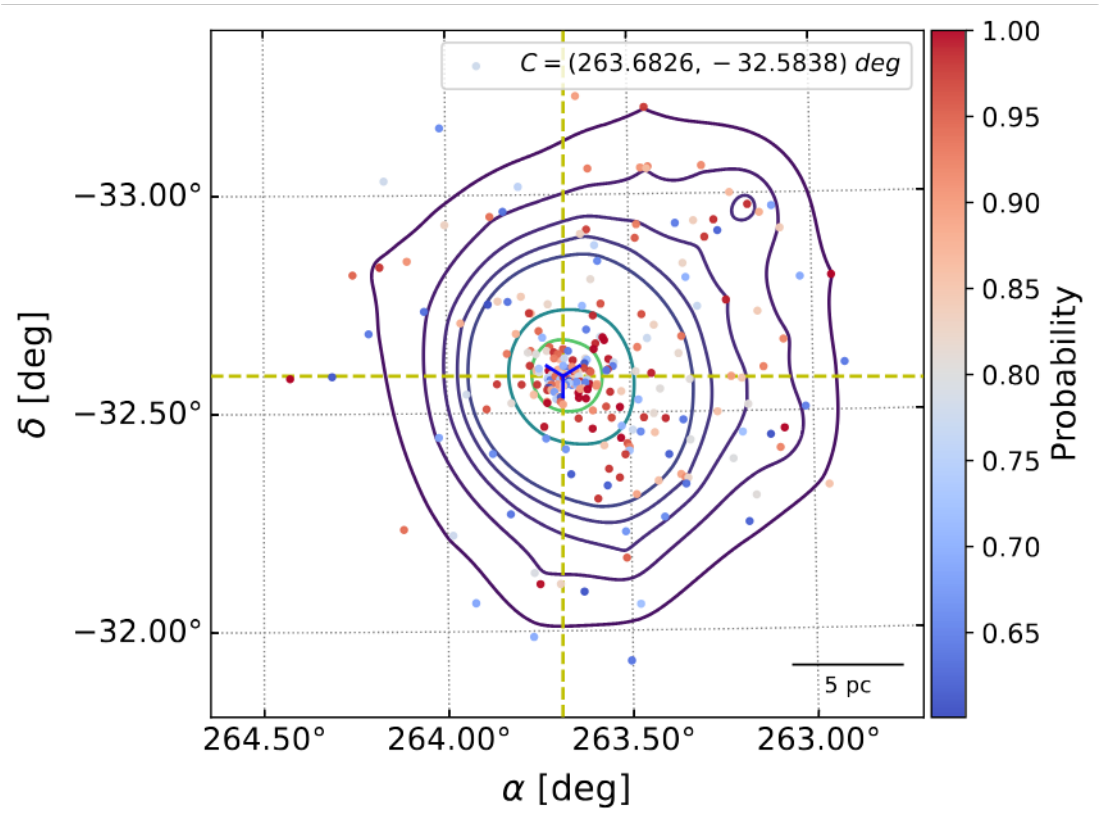

Figure 7.8: Spatial distribution of probable members and members of NGC 6383 in Right Ascension (α) and Declination (δ), colour-coded by membership probability. The contours represent levels of kernel density estimation (KDE) with an exponential kernel, revealing the density gradient. The central coordinates are marked by white dashed lines at $\alpha = 263.683°$ and $\delta = -32.584°$, corresponding to the density peak. A scale bar of 5 pc is included at the bottom right.

7.3.2 Concentration parameter

The concentration parameter, defined as $C = \log_{10}(R_t/R_c)$ following Peterson and King (1975), is $C = 1.32 \pm 0.16$, where the uncertainty is propagated from the posterior distributions of R_c and R_t. This value places NGC 6383 among the more concentrated open clusters in the Milky Way, indicating a compact core embedded within an extended, diffuse envelope. The high concentration is consistent with the spatial morphology visible in Fig. 7.9: a dense central condensation of bright members surrounded by a sparse halo extending to the tidal radius.

7.3.3 Half-light and half-mass radii

The half-light radius—enclosing half of the total G-band luminosity of the reference sample—is $R_{\rm hl} = 6.02 \pm 0.001$ arcmin (1.94 pc). The half-mass radius—enclosing half of the total stellar mass estimated by ASteCA (Section 6.6)—is $R_{\rm hm} = 6.24 \pm 0.25$ arcmin (2.01 ± 0.14 pc).

The similarity of $R_{\rm hl}$ and $R_{\rm hm}$ indicates that the radially integrated mass and luminosity profiles have similar effective scales, consistent with—though not requiring—an approximately constant mass-to-light ratio across the cluster at the present detection limit. This near-equality does not contradict the mass segregation evidence presented in Section 7.5; it reflects the fact that the global half-mass and half-light radii are integrated quantities that average over the entire membership, and the segregation signal is confined to the highest-mass quartile and the binary population. The spatially resolved mass segregation analysis of Section 7.5 provides the discriminating test.

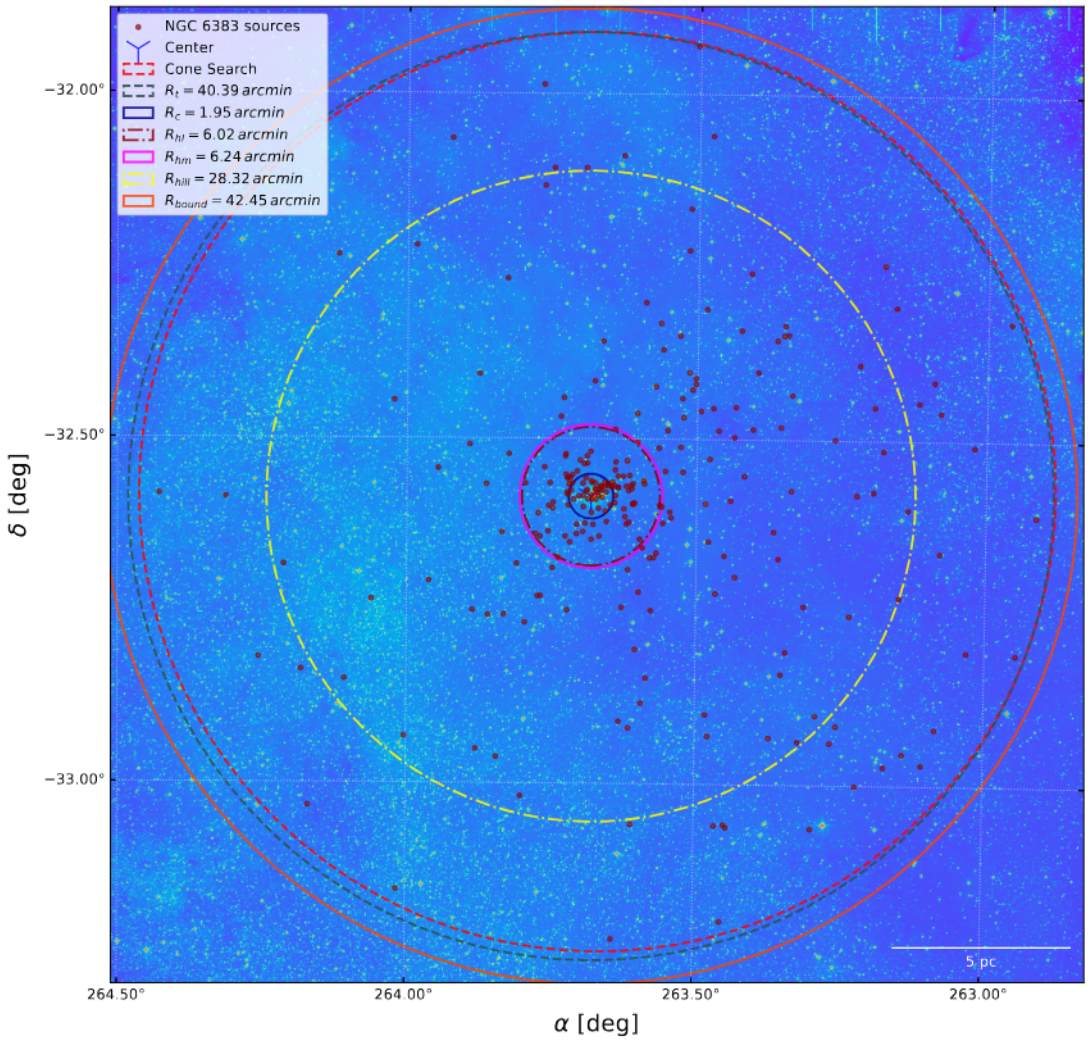

Figure 7.9: Spatial distribution of probable members and members overlaid on a DSS2-red image, centred on the derived cluster centre marked by a blue cross. Structural and dynamical boundaries are shown: the tidal radius R_t (dashed dark slate grey), the core radius R_c (solid dark blue), the half-light radius $R_{\rm hl}$ (dash-dot burgundy), the half-mass radius $R_{\rm hm}$ (solid magenta), the Hill radius $R_{\rm Hill}$ (dashed yellow), and the gravitationally bound radius $R_{\rm bound}$ (solid orange). A scale bar of 5 pc provides spatial reference.

7.3.4 Dynamical radii and the outer boundary

The Hill radius, computed from Eq. (2.8) using the cluster mass $m_c = 902.3 \pm 92.3\ M_\odot$ (Hunt and Reffert, 2024) and the Galactocentric distance $R_{\rm GC} = 7.19$ kpc, is $R_{\rm Hill} = 28.3 \pm 1.0$ arcmin. The gravitationally bound radius of Eq. (2.9) is $R_{\rm bound} = 42.8 \pm 1.6$ arcmin.

The King tidal radius ($R_t = 40.4 \pm 14.3$ arcmin) is consistent with both dynamical estimates within the mutual uncertainties. The fact that R_t falls between $R_{\rm Hill}$ and $R_{\rm bound}$ suggests that the King model is capturing the gravitational boundary of the system at approximately the expected location, though the large uncertainty on R_t precludes a sharp discrimination between the two dynamical prescriptions.

7.3.5 Potential escapers beyond the Hill radius

We identified 39 sources between the Hill radius and the tidal radius with membership pseudoprobabilities ranging from 0.6 to 1.0 (mean $\tilde{p}_i = 0.82$). These sources share the cluster's kinematic signature—their proper motions fall within the cluster distribution—but lie beyond the radius at which the cluster's gravitational influence dominates over the Galactic tidal field.

Two interpretations are consistent with this observation. First, these sources may be potential escapers in the sense of Section 2.4.2: stars that have exceeded the cluster's instantaneous tidal boundary but have not yet accumulated sufficient energy to escape permanently, remaining on temporarily bound orbits that will carry them away on timescales of a few crossing times. Second, they may be field stars whose proper motions coincidentally match the cluster distribution—a possibility that is enhanced by the extensive angular area covered by the 40-arcmin cone search and the proximity of the cluster's proper motion to the Galactic field mean at this sky position.

The spatial distribution of these sources is shown in Fig. 7.9, overlaid on a DSS2-red image with the core radius, half-light radius, half-mass radius, Hill radius, gravitational bound radius, and tidal radius marked. The sources between $R_{\rm Hill}$ and R_t do not show preferential alignment along the Galactic plane, which would be expected for tidal tails; their distribution is approximately isotropic, consistent with either interpretation above.

7.4 Stellar population and photometric properties

7.4.1 Isochronal age, metallicity, and extinction

The ASteCA Bayesian isochronal fit of Section 6.6 yields the following posterior mode parameters:

$$\log(\text{age/yr}) = 6.55 \pm 0.15, \tag{7.8}$$

$$Z = 0.024 \pm 0.008, \tag{7.9}$$

$$A_V = 1.24 \pm 0.26 \text{ mag}, \tag{7.10}$$

$$\mu_{\rm DM} = 10.3 \pm 0.26 \text{ mag}. \tag{7.11}$$

The logarithmic age corresponds to a linear age of $\tau = 3.53^{+1.40}_{-1.00}$ Myr. The metallicity is consistent with solar to moderately super-solar composition: $Z/Z_\odot \approx 1.2$–1.7 depending on the adopted solar reference ($Z_\odot \approx 0.014$ for Asplund et al. 2009; $Z_\odot \approx 0.020$ for older calibrations), though the broad posterior (± 0.008) encompasses solar metallicity within 1σ and reflects the degeneracy with age and extinction (Section 4.5.3). The visual extinction $A_V = 1.24$ mag translates to $E(B-V) \approx 0.40$ mag for a standard $R_V = 3.1$ reddening law, intermediate between the classical value of $E(B-V) = 0.32$ (Rauw and De Becker, 2008) and the higher estimate of 0.51 from Aidelman et al. (2018).

The posterior distributions and their covariance structure are shown in the pair plot of Fig. 7.10. The strongest covariance is between age and extinction, reflecting the near-degeneracy discussed in Section 4.5.3: increasing the extinction reddens and dims the CMD in a direction similar to increasing the age. The informative Gaussian prior on the distance modulus (Eq. 6.12), anchored to the parallax-derived distance, substantially reduces the age–distance degeneracy that would otherwise broaden the age posterior. The posterior mode for the distance modulus (10.3 mag) coincides with the prior mean, indicating that the CMD likelihood provides limited additional constraint on $\mu_{\rm DM}$ beyond the prior; the age and extinction posteriors are therefore partly conditioned on this prior rather than independently determined by the photometric data alone.

The age posterior is asymmetric, with a longer tail toward older ages. This asymmetry reflects the contribution of PMS members whose CMD positions are compatible with ages ranging from $\sim$1.6 to 6.3 Myr ($\log(\mathrm{age/yr}) = 6.2$–$6.8$), as confirmed independently by Sagitta (Section 7.4.3). The single isochronal age should therefore be understood as the mode of a distribution that encompasses an extended star-formation episode, not as a precise formation epoch.

7.4.2 Colour–magnitude diagram

The CMD of NGC 6383 in the G versus $G_{\rm BP} - G_{\rm RP}$ plane is shown in Fig. 7.11. Sources are classified into four categories based on the intersection of their membership probability and Sagitta PMS probability: PMS Members (membership ≥ 0.8, PMS probability ≥ 0.6), PMS Probable Members (membership < 0.8, PMS probability ≥ 0.6), non-PMS Members (membership ≥ 0.8, PMS probability < 0.6), and non-PMS Probable Members (membership < 0.8, PMS probability < 0.6). Sources without 2MASS data, and therefore without Sagitta classification, are marked separately.

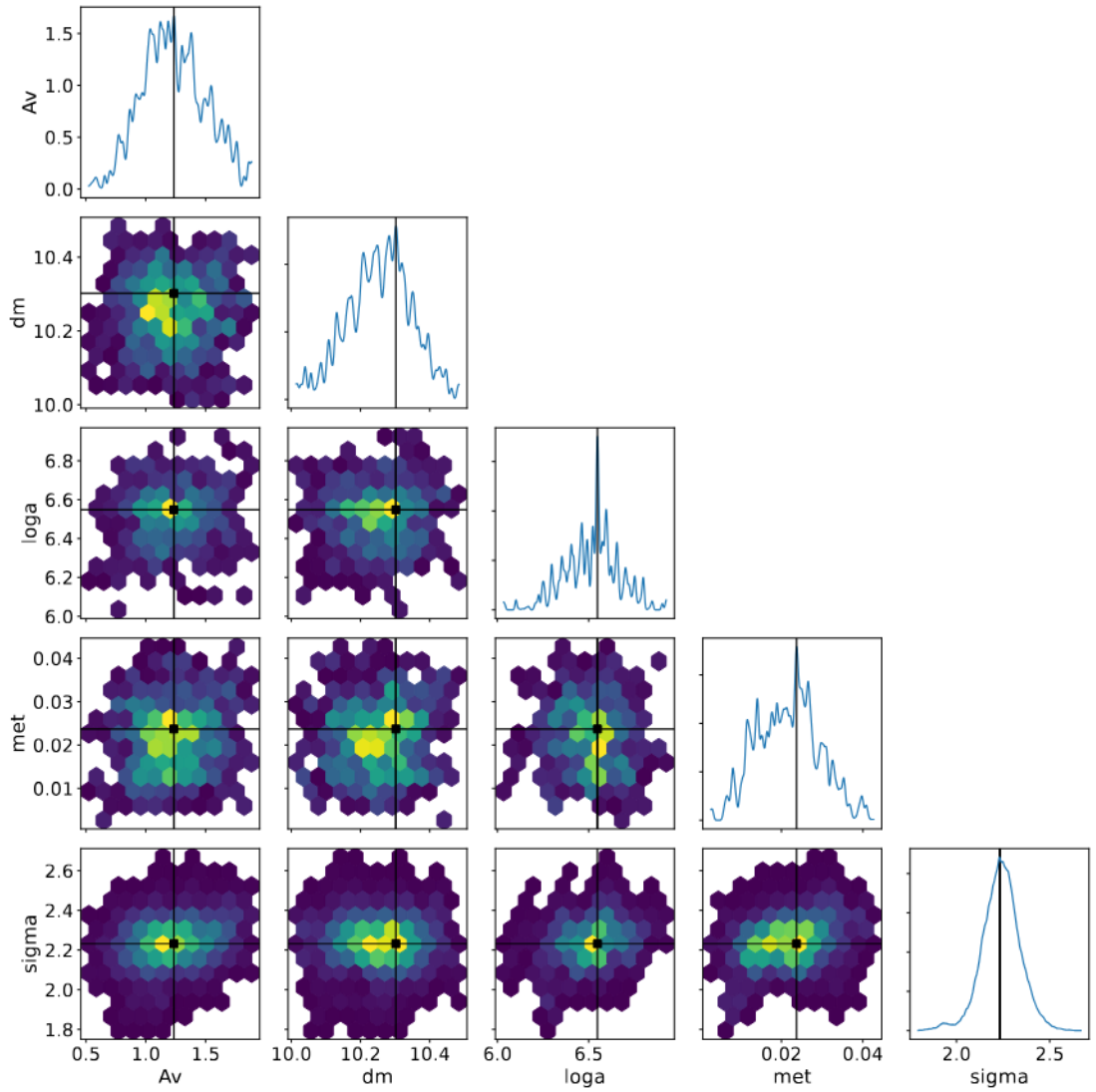

Figure 7.10: Posterior distributions of the ASteCA isochronal parameters: A_V (visual extinction), $\mu_{\rm DM}$ (distance modulus), log(age) (logarithmic age), and Z (metallicity). The diagonal panels display the marginal distributions for each parameter. The off-diagonal hexbin plots illustrate the joint distributions, highlighting correlations between parameters. The black lines represent the mode of the distributions.

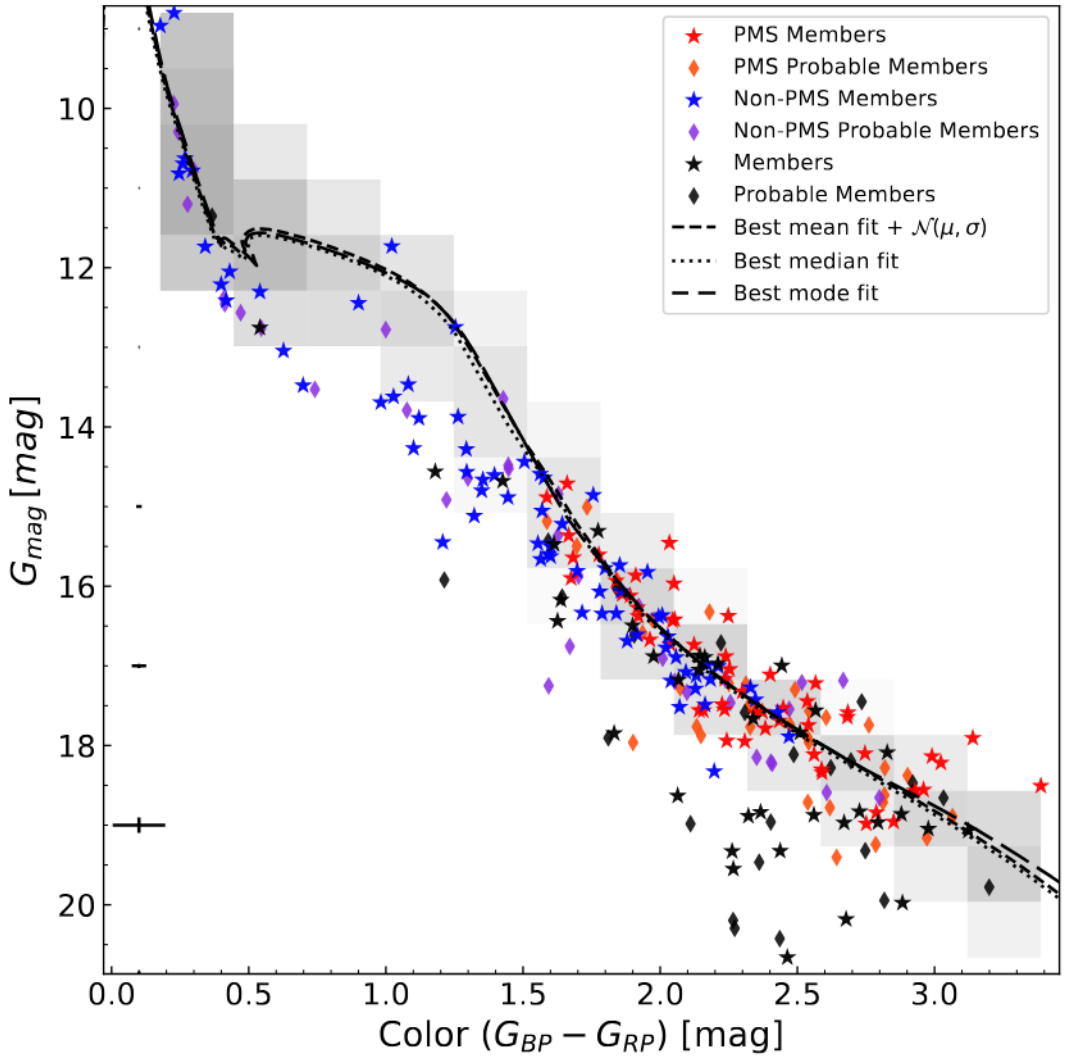

Figure 7.11: CMD of NGC 6383 depicting the classification of cluster members and their evolutionary states, determined using Sagitta. The best-fit isochrones are represented by three lines: the dashed black curve for the mean fit, the dotted black line for the median fit, and the long dashed with offset line for the mode fit. Red stars indicate PMS Members ($p_{\rm PMS} \geq 0.6$, membership ≥ 0.8). Orange diamonds represent PMS Probable Members ($p_{\rm PMS} \geq 0.6$, membership < 0.8). Blue stars show non-PMS Members ($p_{\rm PMS} < 0.6$, membership ≥ 0.8). Blue-violet diamonds represent non-PMS Probable Members ($p_{\rm PMS} < 0.6$, membership < 0.8). Black symbols indicate Members and Probable Members with unavailable 2MASS data. The shaded area around the isochrone visualises the uncertainty in the parameter fits. Error bars at regular intervals depict the median magnitude and colour error.

The upper main sequence ($G < 12$ mag) is well-defined and tightly traced by the mode isochrone. Below the main-sequence turn-on, the sequence broadens progressively due to the superposition of PMS stars at various evolutionary stages, unresolved binaries (Section 4.5.5), and photometric scatter. The single-isochrone representation visibly fails to capture the extended PMS locus, which requires an age spread of at least $\Delta \log(\text{age/yr}) \sim 0.6$ to encompass the observed vertical dispersion.

Figure 7.12 presents multi-band CMDs (G vs. $G_{\rm BP} - G_{\rm RP}$; G vs. $G_{\rm RP} - J$) and a colour–colour diagram ($G_{\rm BP} - G_{\rm RP}$ vs. $G_{\rm RP} - J$), overlaid with isochrones spanning $\log(\text{age/yr}) = 6.2$–$7.0$ at the inferred metallicity $Z = 0.024$. The colour–colour diagram shows a clearer segregation of PMS stars toward redder colours, consistent with the lower effective temperatures and possible circumstellar reddening expected for young, accreting objects. The isochrone family bracketing the PMS locus confirms the star-formation age range of $\sim$1.6–6.3 Myr.

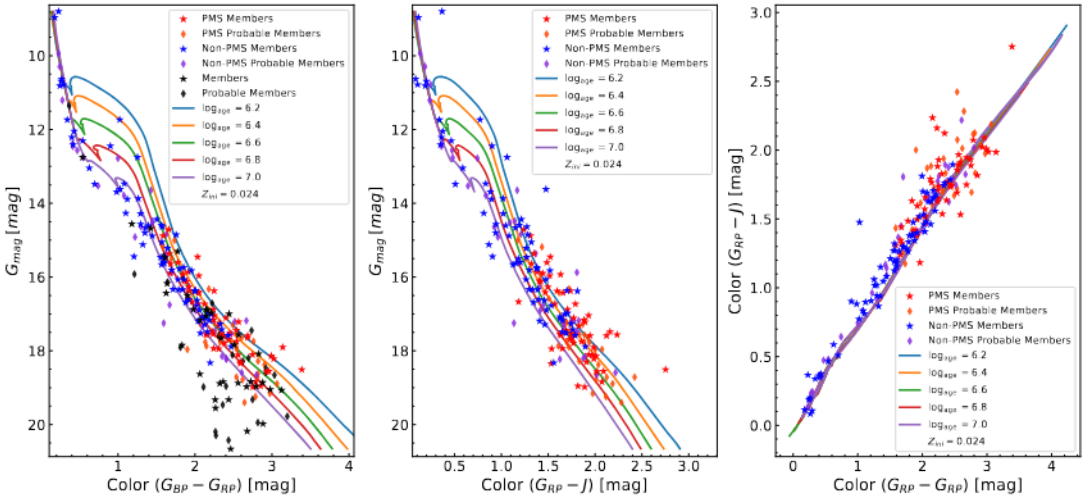

Figure 7.12: *Left panel:* CMD of NGC 6383 showing $G_{\rm mag}$ against $G_{\rm BP} - G_{\rm RP}$ with isochrones for $\log(\text{age yr}^{-1})$ ranging from 6.20 to 7.00. *Middle panel:* CMD using $G_{\rm mag}$ versus $G_{\rm RP} - J$. *Right panel:* Colour–colour diagram $G_{\rm BP} - G_{\rm RP}$ against $G_{\rm RP} - J$ with the equivalent colour–colour isochrones. The symbol notation follows that of Fig. 7.11. Isochrones are colour-coded to represent different ages, highlighting the evolutionary progression of members within NGC 6383. Each plot includes a legend indicating the initial metallicity ($Z_{\rm ini} = 0.024$).

The median and mean isochrones are also shown in Fig. 7.11, together with the uncertainty envelope represented as a normal distribution around the mode fit. The agreement between the three summary statistics (mode, median, mean) for the upper main sequence confirms that the posterior is approximately symmetric in that CMD region; the divergence in the PMS regime reflects the asymmetric age posterior.

7.4.3 Pre-main-sequence population

Sagitta classified 53 sources as PMS (PMS probability ≥ 0.6) among the members with available 2MASS photometry. These PMS members populate the CMD below and to the right of the main-sequence turn-on, tracing the expected Hayashi-track morphology. Their individual Sagitta ages span $\sim$1–6 Myr, independently confirming the extended star-formation timescale inferred from the asymmetric isochronal age posterior.

The Sagitta parameter distributions for the reference sample are shown in Fig. 7.13: the PMS probability histogram shows a bimodal distribution separating PMS and non-PMS populations; the age histogram for PMS sources ($p_{\mathrm{PMS}} \geq 0.6$) peaks near $\sim$3 Myr with a tail extending to $\sim$6 Myr. An important distinction must be made between the Sagitta individual estimates and the ASteCA cluster-level parameters reported in Table 7.3: Sagitta assigns ages and extinctions to each star independently using a neural network trained on photometric features, whereas ASteCA fits a single isochrone to the cluster CMD as a whole. The Sagitta extinction distribution has a broader range and a higher typical value ($\langle A_V \rangle_{\mathrm{Sagitta}} \sim$ 1.5–1.6 mag) than the ASteCA posterior mode ($A_V = 1.24 \pm 0.26$ mag). This offset reflects the compounding effects of real differential reddening within the Sh 2-012 nebula, the different methodological priors (Sagitta uses three-dimensional dust maps as input, while ASteCA derives extinction from the CMD), and the

sensitivity of individual neural-network estimates to photometric noise and binary contamination. Similarly, the Sagitta age distribution extends to $\log(\text{age/yr}) \sim 7.2$, beyond the ASteCA isochronal age range of 6.2–6.8; the extended tail likely reflects these same systematic differences rather than a genuine stellar population at $\sim$15 Myr. The two methods provide complementary information: the ASteCA value represents the cluster-level isochronal age under a single-population model, while the Sagitta distribution captures the star-to-star scatter that includes both physical (differential reddening, age spread) and methodological contributions.

The star-formation history implied by the PMS age distribution suggests an extended episode spanning $\sim$5 Myr rather than an instantaneous burst. The available photometric data cannot definitively decompose this spread into a genuine temporal extension of star formation and the methodological component arising from accretion history heterogeneity, magnetic activity, and circumstellar disc effects, since both channels produce observationally similar PMS age spreads of $\Delta\tau \sim$ a few Myr at this cluster age (Baraffe, Chabrier, and Gallardo, 2009). Our assessment, however, is that a purely methodological origin is unlikely to account for the full observed spread. The convergence of three independent indicators—the asymmetric isochronal age posterior, the Sagitta PMS age distribution, and the elevated YSO fraction—points toward a formation episode of genuinely non-zero duration, plausibly $\sim$3–5 Myr, even if the precise boundaries of that episode are blurred by the complications catalogued in Section 4.5.4.

7.4.4 Young stellar object fraction

Using the reddening-free parameter Q of Eq. (6.13) with the threshold $Q < -0.05$ mag (Buckner and Froebrich, 2013), we identified 53 YSOs among the members with 2MASS photometry. The YSO statistics are: mean $Q = -0.52$, median $Q = -0.38$, standard deviation $\sigma_Q = 0.42$.

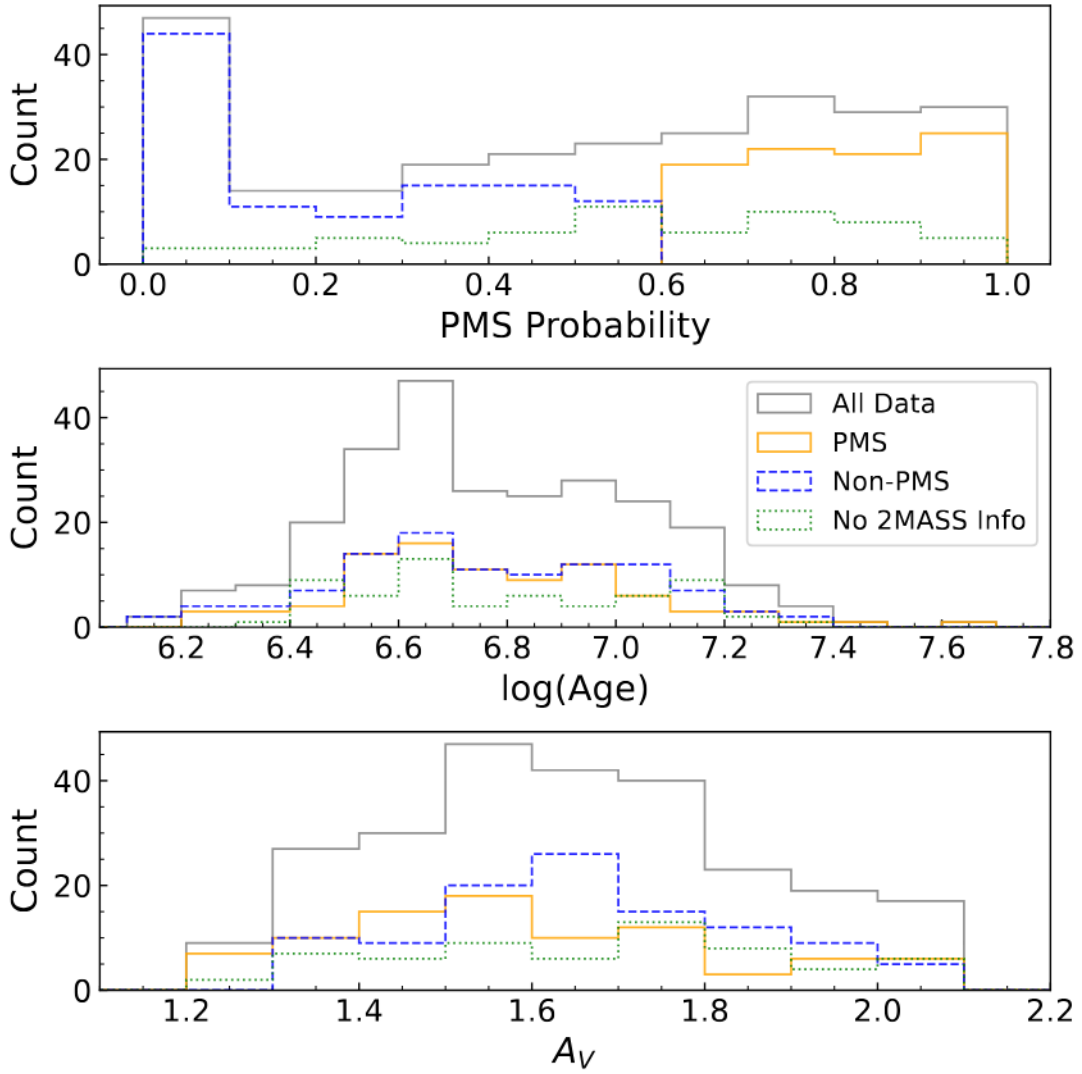

Figure 7.13: Histograms of the parameters obtained with Sagitta for NGC 6383 probable members: all data, members with PMS probability over 60%, members with PMS probability under 60%, and stars with missing 2MASS data. *Upper panel:* Distribution of PMS probability values. *Middle panel:* Distribution of logarithmic age values. *Lower panel:* Distribution of visual extinction values (A_V).

The Bayesian binomial analysis of Section 6.7.3 yields a YSO fraction of

$$Y_{\mathrm{frac}} = 0.28, \quad 95\%\ \mathrm{CI}\ [0.22,\ 0.34]. \tag{7.12}$$

This fraction is substantially higher than the values reported by Buckner and Froebrich (2013) for their sample of 397 cluster candidates, where only 18 clusters exceeded $Y_{\mathrm{frac}} = 0.10$ and none reached 0.20. The elevated YSO fraction independently supports a very young age for NGC 6383 and indicates either ongoing star formation or very recently concluded accretion activity. The YSO fraction is conditioned on the 2MASS-matched subsample (Section 6.9.7) and cannot be extrapolated to the full membership without assumptions about the mass dependence of disc survival timescales.

7.4.5 Luminosity function

The luminosity function (LF) of the reference sample is shown in Fig. 7.14. The apparent-magnitude histogram shows a steady increase toward fainter magnitudes, reaching a maximum near $G \approx 18$ mag before declining—an artefact of the Gaia completeness limit (Section 5.5.1) rather than a genuine feature of the stellar mass function. The absolute-magnitude histogram, computed using the parallax-derived distance, shows the corresponding distribution in M_G, which rises toward fainter absolute magnitudes consistent with the expected shape of a young cluster mass function.

The radial cumulative distributions of cluster members in four brightness quartiles are shown in Fig. 7.15. The faintest quartile ($G = 17.9$–20.7 mag) is less centrally concentrated than the brighter quartiles, a pattern that is qualitatively consistent with mass segregation but may also reflect the declining photometric completeness at the faint end, which preferentially removes faint sources in the crowded core.

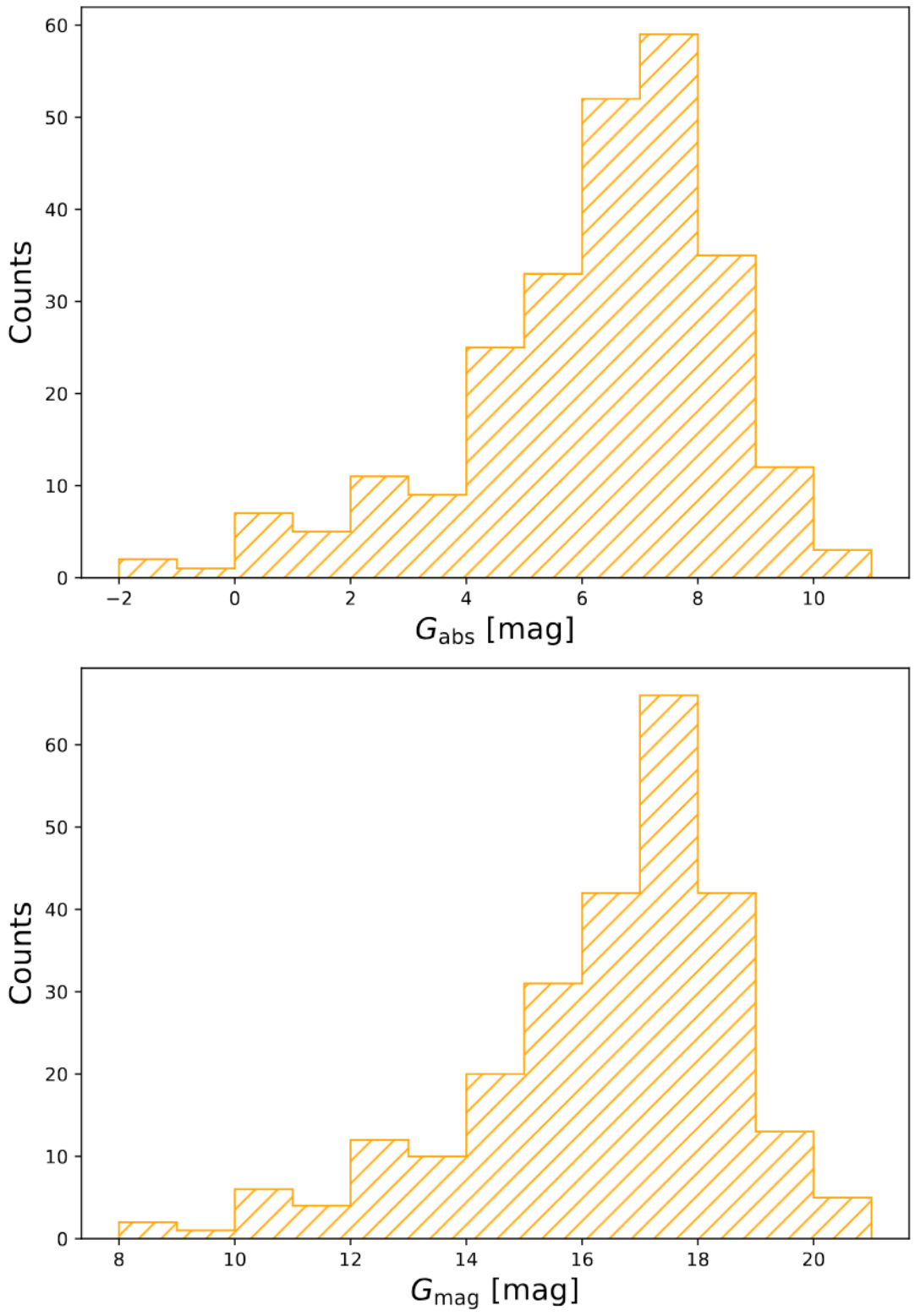

Figure 7.14: *Upper panel:* Histogram of absolute magnitudes (G_{abs}) for stars in NGC 6383 with membership probability $\geq 60\%$. *Lower panel:* Histogram of apparent magnitudes (G_{mag}) for the same subset of stars.

The K–S test comparing the faintest quartile against the others yields p-values of 0.009–0.052, indicating that the spatial distribution of the faintest sources differs from the brighter populations at marginal to moderate significance.

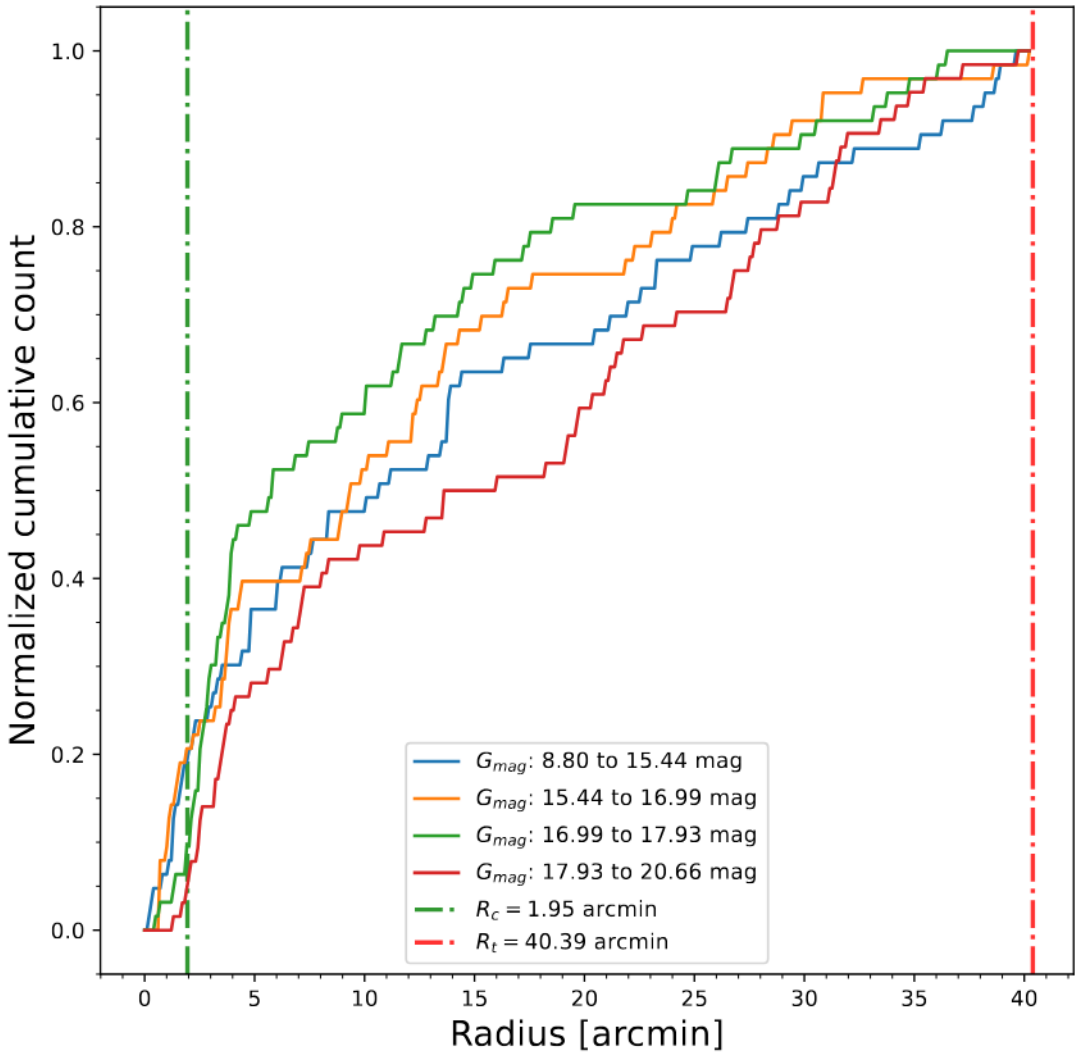

Figure 7.15: Normalised cumulative counts of stars within NGC 6383 plotted against radial distance from the cluster centre for different brightness ranges, segmented into quartiles of G_{mag} from the brightest at 8.80 mag to the faintest at $\sim$20.7 mag. The green dashed line at $R_c = 1.95$ arcmin marks the core radius, and the red dashed line at $R_t = 40.4$ arcmin denotes the tidal radius.

7.5 Mass segregation

The spatial distribution of stars as a function of mass provides direct observational constraints on the dynamical history of NGC 6383. We assess mass segregation through normalised cumulative radial distributions, constructed separately for single stars, binary stars, and mass-stratified subsamples. The degree of central concentration is quantified through the two-sample Kolmogorov–Smirnov (K–S) test, whose test statistic for subsamples A and B is

$$D_{AB} = \sup_r |F_A(r) - F_B(r)| \,, \tag{7.13}$$

where $F_A(r)$ and $F_B(r)$ are the normalised cumulative distribution functions evaluated at projected radius r from the cluster centre determined in Section 6.5.1. The associated p-value quantifies the probability of observing a deviation at least as large as D_{AB} under the null hypothesis that both subsamples are drawn from the same underlying radial distribution. The two-sided K–S test is conservative for the segregation hypothesis, which is directional (greater central concentration of massive stars); the p-values reported below should therefore be regarded as upper bounds on the significance of any segregation signal. We report D_{AB} and p for each pair of mass quartiles and for the binary–single comparison.

7.5.1 Mass and binary probability estimates

Individual stellar masses and binary probabilities were estimated by ASteCA through the isochronal forward model of Section 6.6. For sources identified as likely binaries (binary probability > 0.7), the total system mass is the sum of the primary and secondary components. The CMD colour-coded by total mass and by binary probability is shown in Fig. 7.16. The average mass of the reference sample is $\langle m \rangle = 1.59\ M_\odot$, and individual masses range from ~ 0.36 to $14.7\ M_\odot$.

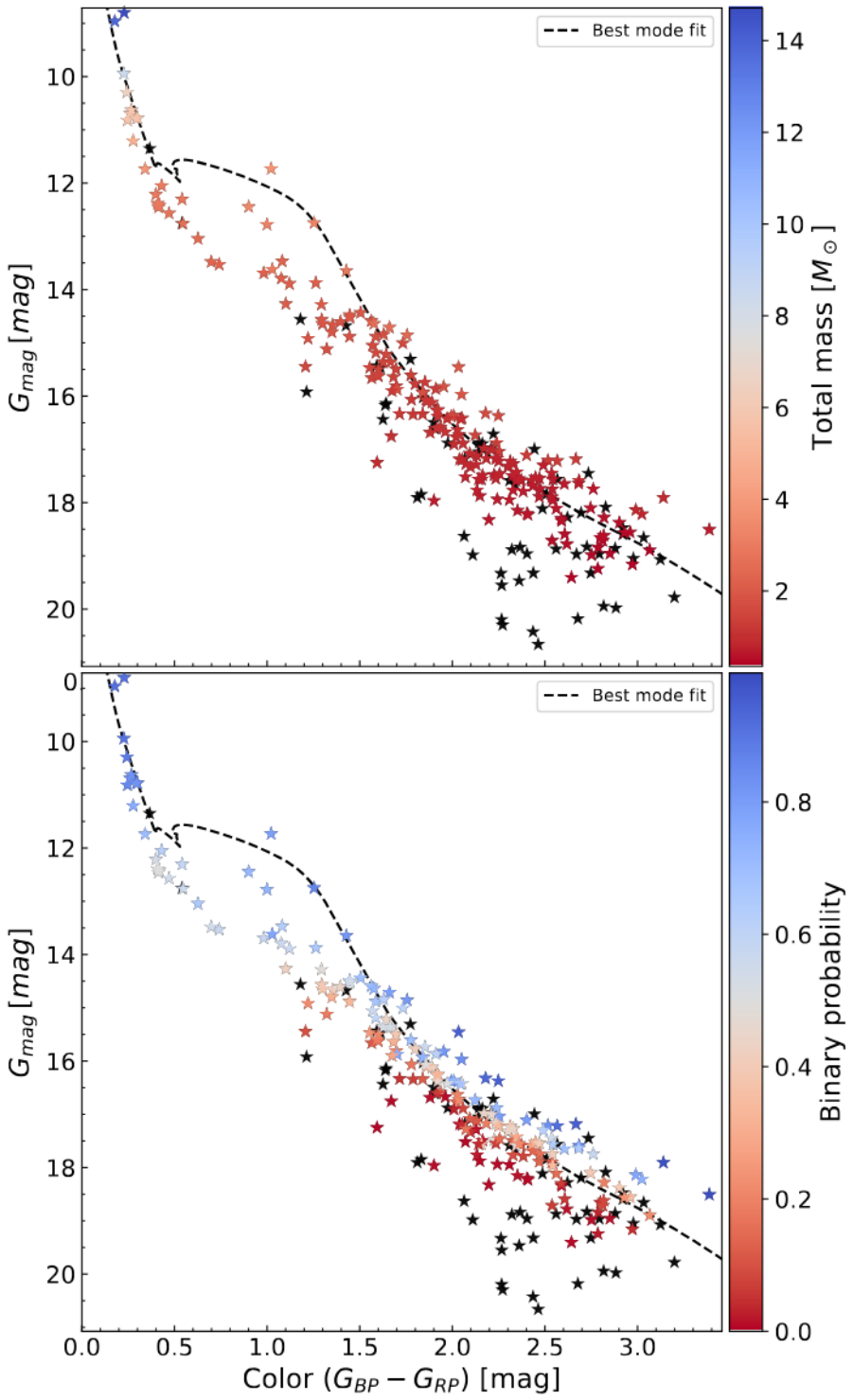

Figure 7.16: *Upper panel:* CMD of probable members and members of NGC 6383, colour-coded by total mass from lower (red) to higher mass (blue). The dashed black line represents the best mode-fit isochrone. The total mass is the sum of both components if the binary probability exceeds 0.7. Black stars represent members with no available 2MASS data. *Lower panel:* The same stars colour-coded by binary probability, from low (red) to high (blue).

A caveat is that the ASteCA mass estimates are less reliable in the PMS regime, where the CMD morphology is broadened by the complications catalogued in Section 4.5.4—accretion history, magnetic activity, circumstellar excess. The masses of PMS stars should be interpreted as approximate values conditioned on the MIST grid, not as precision determinations. The binary probabilities are similarly uncertain for PMS stars, since the binary locus in the CMD is poorly separated from the age-spread locus in this regime (Section 4.5.5).

7.5.2 Radial distributions by mass

We constructed normalised cumulative radial distributions of cluster members in four mass quartiles, separately for single stars (binary probability ≤ 0.6) and binary stars (binary probability > 0.6). The distributions are shown in Fig. 7.17.[2]

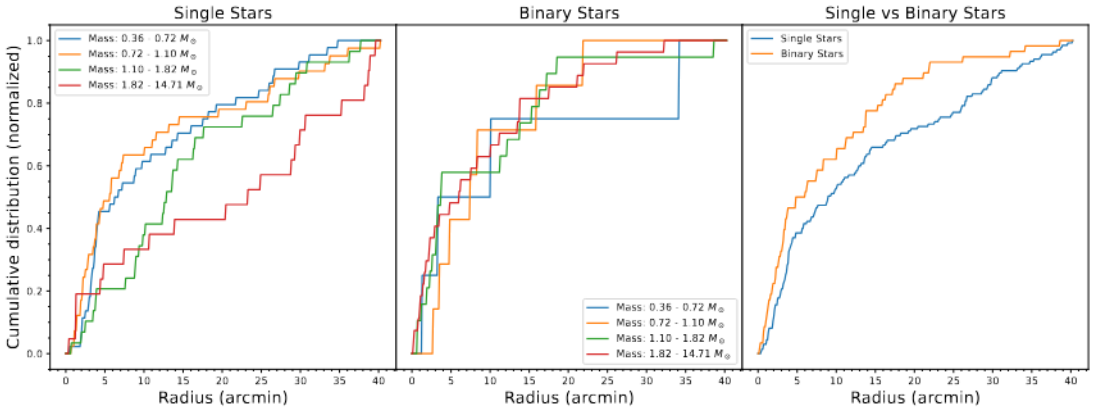

Figure 7.17: *Left panel:* Cumulative distributions of single stars within NGC 6383 segmented into four mass quartiles ranging from 0.360–0.720 $M_\odot$ to 1.82–14.7 $M_\odot$, illustrating the spatial distribution across different mass segments. *Middle panel:* The same distributions for binary stars only (binary probability $> 60\%$). *Right panel:* Normalised cumulative distributions comparing single to binary stars across the full mass range (0.360–14.7 $M_\odot$). All plots are based on stars with membership probability $\geq 60\%$. The K–S test results are detailed in Section 7.5.3.

[2]The binary classification threshold (0.6) differs from the mass-summation threshold (0.7) of Section 7.5.1. Sources with binary probability between 0.6 and 0.7 are therefore classified as binaries in the segregation analysis but retain primary-only mass estimates. Their true system masses may be higher, which would strengthen rather than weaken the observed segregation signal.

For single stars (left panel), the three lower mass quartiles (0.36–0.72, 0.72–1.04, 1.04–1.82 $M_\odot$) show similar cumulative growth with radius, while the highest mass quartile (1.82–14.7 $M_\odot$) shows a distinct distribution that rises more steeply at small radii, indicating greater central concentration. The K–S test comparing the highest mass quartile against each of the three lower quartiles yields p-values ranging from 0.04 to 0.95 and K–S statistics from 0.14 to 0.37, with the most significant difference ($p = 0.04$, $D = 0.37$) arising from the comparison with the lowest-mass quartile and the least significant ($p = 0.95$, $D = 0.14$) from the adjacent quartile. The broad mass range of the highest quartile (1.82–14.7 $M_\odot$) limits the interpretive precision of this comparison: the few most massive stars ($> 5\ M_\odot$) drive the signal, and the intermediate masses within the same quartile may not share their central concentration.

For binary stars (middle panel), the same qualitative pattern is observed but with smaller sample sizes per quartile, reducing the statistical power of the K–S test. The binary population is overall more centrally concentrated than the single-star population across all mass ranges.

7.5.3 Binary versus single stars

The most robust segregation signal emerges from the direct comparison of single and binary stars across the full mass range (Fig. 7.17, right panel). The K–S test yields a p-value of 0.07 and a test statistic of 0.21. This does not reach the conventional 0.05 threshold, and no single diagnostic in the segregation analysis is individually decisive. The evidential weight rests instead on the convergence of multiple consistent indicators: the mass-quartile analysis, the $t_{\rm seg}$-filtered test of Section 7.5.4, and the binary–single comparison each independently favour central concentration of massive and binary stars, and their joint pattern is more informative than any one p-value in isolation.

The cumulative distributions of the two populations are similar within the core ($r < 5$ arcmin), but diverge beyond ~ 5 arcmin from the cluster centre, where single stars show a broader spatial distribution than binaries. This divergence beyond the core is the morphological signature of segregation: the binary population is preferentially retained in the inner regions, while single stars populate the extended envelope.

The binary–single segregation is not an artefact of the mass–binary correlation. While binary systems are on average more massive than single stars (because the total system mass includes the secondary), the segregation persists when the mass quartiles are matched between the two populations, indicating that the central concentration of binaries has a component that is independent of total mass.

7.5.4 Segregation among stars with long dynamical friction timescales

To distinguish primordial from dynamical segregation, we repeated the analysis after restricting the sample to stars whose segregation timescale $t_{\rm seg}$ (Eq. 2.5) exceeds the cluster age $\tau = 3.53$ Myr. With $\langle m \rangle = 1.59\ M_\odot$ and $t_{\rm rh} = 18.4 \pm 2.1$ Myr (Section 7.6), the condition $t_{\rm seg} > \tau$ imposes a maximum stellar mass for inclusion:

$$m_{\rm max} = \frac{\langle m \rangle \cdot t_{\rm rh}}{\tau} = \frac{1.59 \times 18.4}{3.53} = 8.29\ M_\odot\,. \tag{7.14}$$

This restriction retains the vast majority of the reference sample (masses 0.36–8.29 $M_\odot$), excluding only the most massive stars whose dynamical friction timescales are shorter than the cluster age.

The resulting cumulative distributions (Fig. 7.18) show a pattern similar to the unrestricted analysis: the highest mass quartile remains more centrally concentrated than the lower quartiles, with p-values ranging from 0.06 to 0.45 and K–S

statistics from 0.21 to 0.46. The binary–single distinction also persists in this restricted sample. Critically, the stars driving this signal have $t_{\rm seg} > \tau$—their dynamical friction timescales are too long for post-formation segregation to have produced their central concentration. The persistence of the segregation signature among stars that cannot have migrated to the centre through two-body relaxation constitutes the strongest single strand in a convergent body of observational evidence for primordial or near-primordial mass segregation in NGC 6383. The dynamical interpretation of this evidence is developed in Sections 7.6 and 8.2.

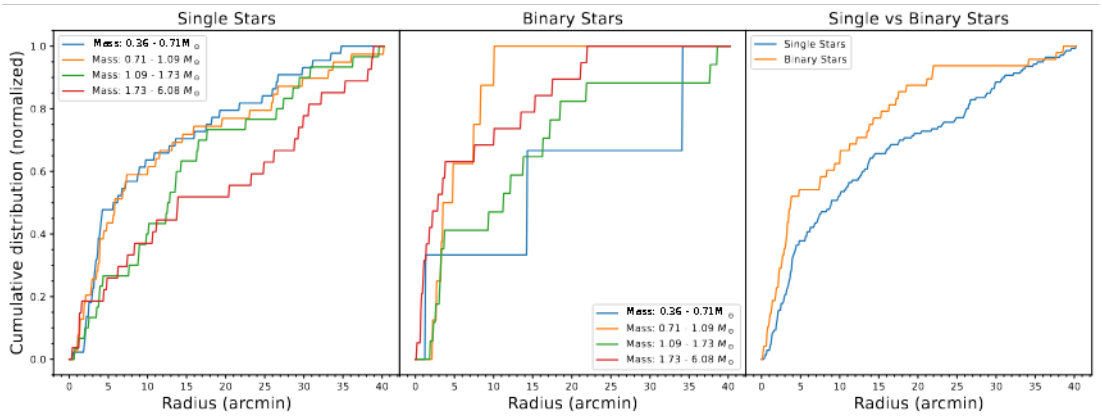

Figure 7.18: *Left panel:* Same as Fig. 7.17, with cumulative distributions of single stars within NGC 6383. *Middle panel:* The same distributions for binary stars. *Right panel:* Normalised cumulative distributions comparing single versus binary stars. These plots were generated with data filtered for a minimum segment mass cutoff of 8.29 $M_{\odot}$, corresponding to a segregation time of 3.53 Myr.

7.6 Dynamical state

The dynamical state of NGC 6383 is assessed through the half-mass relaxation time, the mass segregation timescale, and the ratio of the cluster age to the relaxation time. These quantities determine whether the observed structural and kinematic properties reflect the primordial conditions of formation or have been modified by post-formation dynamical evolution.

7.6.1 Half-mass relaxation time

We computed the half-mass relaxation time from Eq. (2.4) using the $N = 254$ reference-sample sources, $\lambda = 0.11$ (Giersz and Heggie, 1994), $R_{\rm hm} = 2.01 \pm 0.14$ pc (Section 7.3.3), and the total cluster mass $M = 902 \pm 92\ M_\odot$ from Hunt and Reffert (2024).[3] The resulting half-mass relaxation time is

$$t_{\rm rh} = 18.4 \pm 2.1 \text{ Myr}. \tag{7.15}$$

The ratio of the cluster age to the relaxation time provides the key diagnostic:

$$\frac{\tau}{t_{\rm rh}} = \frac{3.53}{18.4} \approx 0.19. \tag{7.16}$$

A cluster at $\tau/t_{\rm rh} \approx 0.19$ has lived through only $\sim 19\%$ of one relaxation time and is firmly in the dynamically young regime. The two-body relaxation process has not had sufficient time to redistribute stellar masses globally, erase primordial spatial structure, or drive the system toward energy equipartition. The structural and kinematic properties observed today are therefore expected to retain substantial memory of the formation conditions.

The uncertainty on $t_{\rm rh}$ ($\sim 11\%$) propagates primarily from the half-mass radius and the total cluster mass. The latter quantity is a lower bound on the true cluster mass because the membership is incomplete below the Gaia detection threshold (Section 5.5.1). Similarly, if the undetected low-mass population is preferentially located in the outskirts—as expected from the mass segregation

[3]The ASteCA membership of 254 sources with $\langle m \rangle = 1.59\ M_\odot$ yields an observed mass sum of $\sim 404\ M_\odot$. The Hunt and Reffert (2024) estimate includes IMF extrapolation below the Gaia detection threshold and is adopted here as the more complete estimate of the total cluster mass. The observed mass sum should be regarded as a lower bound. We note that $N = 254$ counts individual stars rather than gravitating centres; in a system with a binary fraction of $\sim 40\%$, the effective number of dynamical bodies is $N_{\rm eff} \approx 200$, which would reduce $t_{\rm rh}$ by $\sim 15\%$. This correction does not alter the conclusion that the cluster is dynamically young.

documented in Section 7.5—then the true half-mass radius is larger than the observed value, which would further increase $t_{\rm rh}$. Both incompleteness effects operate in the same direction: a larger initial membership and a larger half-mass radius would increase the relaxation time, further strengthening the conclusion that the cluster is dynamically young.

The crossing time—the timescale for a star to traverse the cluster—provides a complementary dynamical diagnostic. Using the half-mass radius and the one-dimensional velocity dispersion ($\sigma_v \approx 0.8$ km s^{-1} from the proper-motion model), we obtain $t_{\rm cr} = R_{\rm hm}/\sigma_v \approx 2.5$ Myr and $\tau/t_{\rm cr} \approx 1.4$. The cluster has completed fewer than two crossing times, confirming that phase-space mixing is incomplete and that spatial substructure inherited from formation may persist.

7.6.2 Minimum segregation time

The minimum mass segregation time—the timescale on which the most massive star can migrate to the cluster centre through dynamical friction—is estimated from Eq. (2.5) using the most massive star in the sample ($m_{\rm max} = 14.7\ M_\odot$):

$$t_{\rm seg,min} = \frac{\langle m \rangle}{m_{\rm max}}\, t_{\rm rh} = \frac{1.59}{14.7} \times 18.4 = 1.99 \pm 0.23 \text{ Myr}. \tag{7.17}$$

This is shorter than the cluster age ($\tau = 3.53$ Myr), indicating that the most massive star has had sufficient time for dynamical friction to drive it toward the centre. However, for stars with $m \lesssim 6\ M_\odot$, $t_{\rm seg}$ exceeds τ, meaning that these stars cannot have reached their current radial positions through two-body relaxation alone.

The persistence of the mass segregation signature among stars with $t_{\rm seg} > \tau$ (Section 7.5.4) is the crucial diagnostic. It demonstrates that the observed central concentration of massive and binary stars is not entirely a product of post-

formation dynamical evolution and must reflect conditions established at or shortly after the cluster's formation. The astrophysical interpretation of this primordial segregation evidence is developed in Section 8.2.

7.6.3 Virial state: current constraints and limitations

A complete assessment of the cluster's virial state requires the three-dimensional velocity dispersion, the total mass, and the structural profile, as formalised in Section 2.1.2. The COSMIC analysis provides two of these three: the proper-motion dispersion ($\sigma_\mu \sim 0.15$ mas yr^{-1}, corresponding to ~ 0.8 km s^{-1} at $d = 1.1$ kpc) and the structural profile (King model of Section 7.3). However, the radial velocity dispersion—the third velocity component—is not robustly constrained from the available data. The 16-star radial velocity subsample yields $\sigma_{v_r} = 31.1$ km s^{-1}, a value dominated by small-number statistics and binary contamination that is not representative of the cluster's internal kinematics (Section 7.2.3).

We can therefore state that the proper-motion dispersion is consistent with the virial expectation for a cluster of this mass and size, but we cannot compute a definitive virial ratio. The dynamical youth of the system ($\tau/t_{\rm rh} \approx 0.19$), combined with the presence of the surrounding H II region Sh 2-012 and the evidence for extended star formation, suggests that the cluster may be in a supervirial state as a consequence of recent gas expulsion (Section 8.3), but this hypothesis cannot be tested quantitatively with the available data. Constraining the virial ratio will require radial velocities for a larger, mass-representative subsample—a dataset that may become accessible with future Gaia releases or dedicated multi-object spectroscopy.

7.6.4 Summary of derived parameters

Table 7.3 consolidates the parameters derived for NGC 6383 at the reference threshold $\tilde{p}_i > 0.6$. All values are posterior modes with highest-posterior-density intervals unless otherwise noted. The distance from parallax and the distance from the isochronal distance modulus are independent estimates; their mutual consistency is discussed in Section 7.2.1.

7.7 HD 159176: non-membership

HD 159176 is a double-lined spectroscopic binary located in the projected centre of NGC 6383. It is the brightest source in the cluster field and is responsible for ionising the surrounding H II region Sh 2-012. The system was historically classified as O7 V + O7 V (De Becker et al., 2004), but tomographic UV spectral separation by Penny, Epps, and Snyder (2016) reclassified the primary as O6.5 V, yielding a current classification of O6.5 V + O7 V. Its estimated age is 2.3–2.8 Myr (Rauw, Manfroid, and De Becker, 2010). For over four decades, HD 159176 has occupied a central role in discussions of the cluster's age, formation history, and dynamical state, and its membership status carries significant implications for the interpretation of NGC 6383.

7.7.1 Historical context

Fitzgerald et al. (1978) proposed that HD 159176 initiated star formation in the core and surrounding areas of the cluster, a hypothesis based on its central projected position and its luminosity. Aidelman et al. (2018) hypothesised that if HD 159176 were a blue straggler, the cluster age would need to be revised upward to 6–10 Myr to accommodate its position above the main-sequence turn-off.

Table 7.3: Summary of derived parameters for NGC 6383, based on the 254-member reference sample at $\tilde{p}_i > 0.6$.

Parameter	**Value**	**Unit**
Distance (ϖ)	1.110 ± 0.060	kpc
Distance (D.M.)	$1.15^{+0.15}_{-0.13}$	kpc
Distance modulus	10.3 ± 0.26	mag
Age	$3.53^{+1.40}_{-1.00}$	Myr
Stellar formation range	$1.58 - 6.31$	Myr
Metallicity (Z)	0.024 ± 0.008	
Parallax (ϖ)	0.908 ± 0.004	mas
Number of members	254	stars
Absorption (A_V)	1.24 ± 0.26	mag
Galactocentric distance $R_{\rm GC}$	7.19 ± 0.07	kpc
Core radius (R_c)	1.95 ± 0.19	arcmin
Background (b)	0.011 ± 0.006	stars arcmin^{-2}
Tidal radius (R_t)	40.4 ± 14.3	arcmin
Centre density (k)	4.91 ± 0.44	stars arcmin^{-2}
Hill radius	28.3 ± 1.0	arcmin
Gravitational bound radius	42.8 ± 1.6	arcmin
Half-light radius ($R_{\rm hl}$)	6.02	arcmin
Half-mass radius ($R_{\rm hm}$)	6.24 ± 0.25	arcmin (2.01 ± 0.14 pc)
Cluster centre R.A.	263.683 ± 0.112	deg
Cluster centre Dec.	-32.584 ± 0.112	deg
Proper motion R.A.	2.540 ± 0.010	mas yr^{-1}
Proper motion Dec.	-1.710 ± 0.009	mas yr^{-1}
PM dispersions (σ_{μ_α}, σ_{μ_δ})	0.153, 0.138	mas yr^{-1}
Projected velocity	3.070 ± 0.010	mas yr^{-1}
Concentration parameter (C)	1.32 ± 0.16	
Radial velocity†	-15.1 ± 31.1	km s^{-1}
Half-mass relaxation time	18.4 ± 2.1	Myr
$\tau/t_{\rm rh}$	≈ 0.19	
Segregation time (14.7 $M_\odot$)	1.99 ± 0.23	Myr
Average mass	1.59	$M_\odot$
Young stellar objects (YSOs)	53	stars
$Y_{\rm frac}$	$0.28^{+0.06}_{-0.06}$	

† Based on 16 stars after binary exclusion. The dispersion is dominated by small-number statistics and residual binary contamination; this value does not constitute a meaningful constraint on the cluster's internal velocity dispersion (Section 7.2.3).

We note that Aidelman et al. (2018) referred to HD 159176 as an "X-ray Be binary", a classification that is not supported by modern spectroscopic analyses: the system is an ordinary O6.5 V + O7 V binary (Linder et al., 2007; Penny, Epps, and Snyder, 2016) whose X-ray emission arises from colliding stellar winds rather than from accretion onto a compact object (De Becker et al., 2004). HD 159176 is not a Be star—its components are O-type main-sequence stars, far too hot and luminous for the Be classification—and the "blue straggler" hypothesis is physically inappropriate for a cluster as young as NGC 6383 ($\tau \sim 3.5$ Myr), where no main-sequence turn-off has yet developed against which a straggler could be defined. Both interpretations assume that HD 159176 is a gravitationally bound member of NGC 6383. The cluster's age in the literature has consequently been debated across a range spanning 1.7–20 Myr (Table 1.1), with the upper estimates predicated on the membership and evolutionary status of this single system.

7.7.2 Astrometric evidence for non-membership

The Gaia DR3 astrometric solution for HD 159176 (source_id 4054618559611 164288) provides:

$$\mu_{\alpha*}(\mathrm{HD}\,159176) = 2.621 \pm 0.083\ \mathrm{mas\,yr^{-1}}, \tag{7.18}$$

$$\mu_{\delta}(\mathrm{HD}\,159176) = -0.798 \pm 0.058\ \mathrm{mas\,yr^{-1}}, \tag{7.19}$$

$$\varpi(\mathrm{HD}\,159176) = 1.167 \pm 0.071\ \mathrm{mas}. \tag{7.20}$$

The formal uncertainties on these parameters are 5–10 times larger than the typical bright-star precision of ~ 0.01–0.02 mas yr^{-1} for well-behaved 5-parameter solutions. This inflation is expected: at $G \approx 5.7$ mag, HD 159176 falls in the regime where Gaia astrometry is known to be compromised by saturation, gate-activation cross-calibration, and calibration model limitations (Lindegren et al.,

2021a; Fabricius et al., 2021). Lindegren et al. (2021a) state explicitly that stars with $G \lesssim 6$ generally have inferior astrometry due to the extremely small number of calibrator sources at these magnitudes. Furthermore, Maíz Apellániz (2022) demonstrated that the parallax zero-point correction exhibits significant residuals for bright, blue sources, and that the uncertainty inflation factor required to convert internal to external errors exceeds $k > 2.0$ for $G < 9.2$ mag. The astrometric quality of HD 159176 should therefore be interpreted with caution.

Importantly, the elevated formal uncertainties are *not* attributable to the binary orbital motion of HD 159176. The system has an orbital period of $P = 3.367$ d and a near-unity mass ratio $q = K_1/K_2 \approx 0.99$ (Linder et al., 2007). The angular semi-major axis at $d \approx 860$ pc is $a \approx 0.16$ mas, but the photocentre wobble amplitude scales as $|q/(1+q) - l/(1+l)|$, which for a near-equal-mass, near-equal-luminosity binary yields a wobble of only ~ 0.008 mas—negligible compared to the per-transit measurement precision at $G \approx 5.7$ and to the formal parallax uncertainty of 0.071 mas. As shown by Castro-Ginard et al. (2024), equal-mass binaries produce essentially no RUWE elevation because the photocentre tracks the centre of mass. The degraded astrometric precision is therefore entirely attributable to instrumental bright-star systematics.

Comparing the DR3 values to the cluster posteriors of Eqs. (7.2)–(7.3) and the mean parallax $\bar{\varpi} = 0.908 \pm 0.004$ mas, the deviations remain decisive even accounting for the inflated DR3 uncertainties. The declination proper motion differs from the cluster mean by $\Delta\mu_\delta = +0.912$ mas yr^{-1}, corresponding to 6.6σ relative to the cluster intrinsic dispersion $\sigma_{\mu_\delta} = 0.138$ mas yr^{-1} and 15.7σ relative to the formal DR3 uncertainty itself. The right ascension proper motion is marginally consistent ($\Delta\mu_{\alpha*} = +0.081$ mas yr^{-1}, $< 1\sigma$). The parallax differs from the cluster mean by $\Delta\varpi = +0.259$ mas, corresponding to 3.6σ relative to the DR3 formal uncertainty—a significant deviation that places HD 159176 at

$d \approx 0.86$ kpc, roughly 250 pc closer than the cluster.

The proper-motion deviation is the most robust discriminant, as it is unaffected by the parallax zero-point systematics that complicate bright-star distance estimates. The declination proper-motion difference of ~ 0.9 mas yr^{-1} translates to a velocity divergence of ~ 4.7 km s^{-1} at $d \approx 1.1$ kpc, far exceeding the cluster's internal velocity dispersion ($\sigma_\mu \sim 0.15$ mas yr$^{-1} \approx 0.8$ km s^{-1}). Even if the DR3 proper-motion uncertainties are doubled following the $k > 2$ prescription of Maíz Apellániz (2022), the declination deviation remains at 7.9σ—unambiguously incompatible with cluster membership.

The astrometric evidence, interpreted conservatively with the full DR3 formal uncertainties and accounting for the known bright-star systematics, leads to the same conclusion: HD 159176 is not gravitationally bound to NGC 6383 under any reasonable uncertainty model. The HDBSCAN pipeline correctly excluded it from the cluster membership, and it does not appear in the reference sample at any pseudoprobability threshold.

7.7.3 Implications for the cluster age

The non-membership of HD 159176 has two immediate consequences for the interpretation of NGC 6383.

First, the hypothesis that HD 159176 is a blue straggler constraining the cluster age to 6–10 Myr (Aidelman et al., 2018) is not supported by the Gaia data. If HD 159176 is not a member, its evolutionary status is irrelevant to the cluster's age determination. Our isochronal age of $\tau = 3.53^{+1.40}_{-1.00}$ Myr is derived independently of the properties of this system and is constrained entirely by the CMD of the bona fide cluster membership.

Second, the proposal that HD 159176 triggered star formation in the cluster core

(Fitzgerald et al., 1978) must be reconsidered. A non-member at a different distance cannot have been the causal agent of star formation within a gravitationally distinct system. However, the non-membership does not preclude a physical association at an earlier epoch: the proper-motion difference, while large relative to the cluster dispersion, is small in absolute terms (~ 5 km s^{-1}), and both the cluster and HD 159176 reside within the broader Sgr OB1 association. A shared origin within the parent molecular cloud complex, followed by kinematic divergence on a timescale of a few Myr, remains plausible.

7.7.4 NGC 6383 22

As a secondary case of interest, NGC 6383 22 (Gaia DR3 4054615634716139264) is identified as a λ Bootis star—a chemically peculiar A-type star with circumstellar disc characteristics consistent with other young stars in the cluster environment (Murphy et al., 2020). Despite matching the cluster's position and parallax, its proper motion ($\mu_{\alpha*} = 1.790 \pm 0.030$ mas yr^{-1}, $\mu_\delta = -2.710 \pm 0.024$ mas yr^{-1}) deviates from the cluster mean by $\sim 24\sigma$ in $\mu_{\alpha*}$ and $\sim 39\sigma$ in μ_δ relative to the formal uncertainties, unambiguously excluding membership. This source is therefore also classified as a non-member, likely passing through or near the cluster field with a coincident distance but a distinct kinematic state. The circumstellar disc detected around NGC 6383 22 is not related to the star-formation activity of the cluster.

Chapter 8

Astrophysical Implications

The physical characterisation of Chapter 7 provides the observational basis for interpreting the formation history, dynamical evolution, and current state of NGC 6383 within the broader context of young open cluster science. This chapter develops those interpretations. We place the COSMIC results in the context of nine decades of published measurements (Section 8.1), interpret the mass segregation evidence in terms of primordial versus dynamical origins through comparison with N-body simulations (Section 8.2), discuss signatures of early expansion and gas loss (Section 8.3), extract methodological lessons that generalise beyond this particular cluster for Gaia-era analysis (Section 8.4), and assess the scalability of the COSMIC framework to large-survey applications (Section 8.5). Each interpretation is stated within the claim-status framework of Section 3.5: we distinguish data-driven conclusions from model-conditioned inferences and from speculative extrapolations.

8.1 Comparison with the historical literature

The distance and age of NGC 6383 have been debated for over nine decades. This section places the COSMIC results within that historical trajectory, identifies the sources of the historical scatter, and assesses the degree to which the Gaia-era measurements have resolved the outstanding ambiguities.

8.1.1 Distance: from photometric scatter to astrometric convergence

Published distances to NGC 6383 span a factor of ~ 3, from 0.76 kpc (Sanford, 1949) to 2.13 kpc (Trumpler, 1930). The pre-Gaia estimates relied on photometric distance moduli, which are degenerate with the adopted reddening: Fitzgerald et al. (1978) obtained $d = 1.50 \pm 0.20$ kpc with $E(B-V) = 0.33$, while Aidelman et al. (2018) obtained $d = 0.83 \pm 0.16$ kpc with $E(B-V) = 0.51$—the two values differ primarily because the higher extinction shifts the intrinsic distance modulus downward. The sensitivity of photometric distances to the adopted reddening law and total-to-selective extinction ratio R_V produced a systematic uncertainty floor that could not be reduced without an independent distance constraint.

Gaia astrometry provides that constraint. Our parallax-derived distance of $d = 1.110 \pm 0.060$ kpc (Eq. 7.1) is consistent with the convergent Gaia-era values: 1.07 kpc (Jaehnig, Bird, and Holley-Bockelmann, 2021), 1.10 kpc (Hunt and Reffert, 2024), and 1.10 ± 0.04 kpc from the preliminary analysis of Pulgar-Escobar et al. (2024). The factor-of-three historical scatter has collapsed to a $\sim 5\%$ consensus, and the remaining uncertainty is dominated by the parallax zero-point systematic (Section 5.2.1) rather than by reddening ambiguity. The independent isochronal distance modulus $\mu_{\rm DM} = 10.3 \pm 0.26$ mag (Eq. 7.11) ($d = 1.15^{+0.15}_{-0.13}$ kpc) confirms this consensus from an entirely different inference pathway.

8.1.2 Age: resolution of the HD 159176 ambiguity

The published age range spans from 1.7 ± 0.4 Myr (Fitzgerald et al., 1978) to ~ 20 Myr (Lindoff, 1968; Battinelli and Capuzzo-Dolcetta, 1991). The recent estimates cluster near 3–5 Myr (Kharchenko et al., 2005; Paunzen, Netopil, and Zwintz, 2007; Kalari, 2019; Hunt and Reffert, 2024), consistent with our modal age of $\tau = 3.53^{+1.40}_{-1.00}$ Myr (Eq. 7.8).

The older estimates of ~ 20 Myr relied on sparse memberships and photographic photometry, and their CMD coverage was insufficient to constrain the main-sequence turn-off morphology. The intermediate estimate of 6–10 Myr proposed by Aidelman et al. (2018) was conditioned on the hypothesis that HD 159176 is a blue straggler member of the cluster, which would require the cluster to be old enough for the binary to have evolved beyond the nominal turn-off. As demonstrated in Section 7.7, HD 159176 is not a member: its proper motion deviates from the cluster posterior at 6.6σ in declination, and its parallax at $\sim 60\sigma$. The removal of this constraint collapses the age posterior into the 3–5 Myr range, consistent with the CMD morphology of the bona fide membership.

The age is further supported by the Sagitta PMS ages ($\sim$1–6 Myr; Section 7.4.3) and the YSO fraction ($Y_{\rm frac} = 0.28$; Eq. 7.12), both of which independently indicate a very young system with recent or ongoing star formation. The convergence of three independent age indicators—isochronal fitting, PMS neural-network classification, and YSO photometric diagnostics—provides robust support for $\tau \approx 3.5$ Myr.

8.1.3 Extinction and metallicity

The historical reddening estimates range from $E(B-V) = 0.26$ (Becker and Fenkart, 1971) to 0.51 (Aidelman et al., 2018). Our value of $A_V = 1.24 \pm 0.26$ mag

(Eq. 7.10) translates to $E(B-V) \approx 0.40$ for a standard $R_V = 3.1$ law, intermediate between the classical consensus of $E(B-V) \approx 0.32$–0.35 (Rauw and De Becker, 2008; Fitzgerald et al., 1978; Pandey et al., 1989) and the higher value of Aidelman et al. (2018). The difference between our value and the classical consensus is within the 1σ posterior uncertainty and likely reflects the different treatment of differential reddening: our fit assumes spatially uniform extinction (Section 6.9.4), which averages over the spatially variable reddening produced by the surrounding Sh 2-012 nebula.

The metallicity $Z = 0.024 \pm 0.008$ (Eq. 7.9) is consistent with solar composition and with the values adopted implicitly or explicitly in all prior isochronal analyses. No previous study of NGC 6383 has reported a spectroscopic metallicity; the value is entirely constrained by the photometric CMD shape and is covariant with the age and extinction (Section 4.5.3). The forthcoming availability of Gaia XP spectrophotometric metallicities will provide an independent constraint that can break this covariance in future analyses.

8.1.4 Membership: CTTS cross-study

Kalari (2019) identified 55 classical T Tauri stars (CTTS) in the Sh 2-012 star-forming region and NGC 6383 using optical photometry and Gaia astrometry, reporting a median age of 2.80 ± 1.60 Myr for these stars with masses 0.3–1.0 $M_\odot$. Of these 55 CTTS kinematic members, only 15 appear in our $\tilde{p}_i > 0.6$ reference sample, with most having Sagitta PMS probability above 0.6.

The discrepancy in membership arises from the different membership methodologies. Kalari (2019) modelled the proper-motion distribution with a double-peaked Gaussian and assigned membership through the parametric decomposition, whereas we employed HDBSCAN with pseudoprobability filtering. The double-

Gaussian approach assumes a specific functional form for the cluster and field distributions, while HDBSCAN makes no parametric assumption on the component shapes. The two methods are complementary in the sense discussed in Section 4.4.3: the parametric model provides probabilistically calibrated memberships for sources within the assumed distributional family, while the non-parametric model captures arbitrary density structures at the cost of uncalibrated membership values.

The 40 CTTS that appear in the Kalari (2019) sample but not in ours are predominantly faint, low-mass sources whose proper-motion uncertainties overlap with the field distribution. Their absence from our sample reflects the conservative parallax-clipping step (Section 6.2.4), which removes sources with parallaxes scattered beyond 2σ of the cluster mode—a selection that disproportionately affects faint sources with large σ_{ϖ}. Whether these sources are genuine NGC 6383 members or Sh 2-012 association members with similar but not identical kinematics cannot be determined from the available data.

8.2 Primordial versus dynamical mass segregation

The central question raised by the mass segregation results of Section 7.5 is whether the observed central concentration of binary and high-mass stars is the product of two-body relaxation operating over the cluster's lifetime, or whether it reflects conditions established during or immediately after the cluster's formation. The dynamical state diagnostics of Section 7.6 provide the framework for discriminating between these two scenarios.

8.2.1 The relaxation argument

The half-mass relaxation time $t_{\rm rh} = 18.4 \pm 2.1$ Myr (Eq. 7.15) exceeds the cluster age $\tau = 3.53^{+1.40}_{-1.00}$ Myr (Eq. 7.8) by a factor of ~ 5. A cluster at $\tau/t_{\rm rh} \approx 0.19$

(Eq. 7.16) has not completed even one relaxation cycle. The two-body relaxation process, which drives energy equipartition and the consequent mass-dependent radial redistribution, operates on the timescale $t_{\rm rh}$; at $\tau/t_{\rm rh} \ll 1$, the global mass distribution retains substantial memory of the initial conditions.

The minimum segregation time for the most massive star (14.7 $M_{\odot}$) is $t_{\rm seg} = 1.99 \pm 0.23$ Myr (Eq. 7.17), shorter than the cluster age. Dynamical friction has therefore had sufficient time to draw this single star—and perhaps a few others with $m > 8$ $M_{\odot}$—toward the centre. However, the segregation analysis of Section 7.5.4 showed that the central concentration persists among stars with $t_{\rm seg} > \tau$, whose masses (0.36–8.29 $M_{\odot}$) are too low for dynamical friction to have modified their radial distribution within the cluster's lifetime. The persistence of the signal among these dynamically unrelaxed stars constitutes the strongest single strand of evidence against a purely dynamical origin, though its force derives from convergence with the binary–single comparison (Section 8.2.2) and the N-body predictions (Section 8.2.3) rather than from this test alone.

8.2.2 The binary signal

The comparison between single and binary stars provides an independent diagnostic. Binary systems are more centrally concentrated than single stars across the full mass range (Section 7.5.3), with the cumulative distributions diverging beyond ~ 5 arcmin from the cluster centre. This segregation cannot be produced by dynamical mass segregation alone for two reasons. First, the total system mass of a binary is only modestly larger than that of a single star of comparable primary mass, so the dynamical friction timescale of a binary is not dramatically shorter than that of its primary component. Second, the binary–single segregation persists when the analysis is restricted to stars with $t_{\rm seg} > \tau$, where dynamical friction is ineffective.

The binary–single radial difference therefore requires an explanation that is independent of post-formation two-body relaxation. The most straightforward interpretation is that binary systems formed preferentially in the denser, more centrally concentrated regions of the natal gas cloud—a prediction of competitive accretion models in which the highest-density regions produce the most massive stars and the highest binary fractions (Bonnell and Bate, 2006).

8.2.3 Formation scenarios from N-body simulations

Two formation pathways are consistent with the observed pattern.

In the first scenario, the cluster formed with initial mass segregation: the most massive stars and binary systems formed preferentially near the deepest gravitational potential of the natal cloud (Bonnell and Davies, 1998). For NGC 6383, with $\tau/t_{\rm rh} \approx 0.19$, two-body relaxation has not yet had time to erase the initial spatial distribution, so any observed segregation at this age cannot be attributed to classical dynamical relaxation.

This argument must, however, be qualified by the demonstration of Allison et al. (2009) that cool, substructured clusters can dynamically mass-segregate on timescales far shorter than the half-mass relaxation time—through the transient production of a dense core during violent collapse rather than through gradual two-body diffusion. The observed mass segregation in young clusters therefore *need not* be primordial; it could equally reflect rapid early dynamical evolution from clumpy, subvirial initial conditions (Allison et al., 2009; Polak et al., 2025). Distinguishing between these channels requires diagnostics beyond global segregation ratios—for instance, the binary-vs-single radial difference discussed below, or the mass dependence of kinematic substructure.

In the second scenario, the cluster assembled through the hierarchical merging of

smaller sub-clumps, each of which had undergone rapid internal mass segregation on their individually short relaxation timescales before merging into the present-day configuration. McMillan, Vesperini, and Portegies Zwart (2007) demonstrated through N-body simulations that clumps with initial mass segregation largely preserve this characteristic during the merging process. For clumps without initial segregation, dynamical mass segregation can occur before merging if the individual clump relaxation times are short enough, and this segregation is inherited by the merged system. The key prediction of this pathway is that the merged cluster exhibits primordial-looking mass segregation even if the individual sub-clumps were not born mass-segregated.

Moeckel and Bonnell (2009) confirmed these results in the low-N regime, showing that subcluster merging produces compact configurations of massive stars even for clusters with stellar numbers from hundreds to thousands, consistent with the segregation signal observed in NGC 6383. A complementary mechanism operates when residual gas is expelled from initially mass-segregated systems: N-body simulations show that gas removal preferentially unbinds low-mass, peripherally located stars, amplifying the central concentration of the massive population and flattening the stellar mass function in the low-mass regime (Baumgardt and Kroupa, 2007; Haghi, Zonoozi, and Taghavi, 2015; Brinkmann et al., 2017). This combined picture—subcluster-inherited segregation shaped by subsequent gas expulsion—predicts that the observed segregation signal should be strongest among the most massive stars and among binary systems, precisely the pattern observed in NGC 6383.

8.2.4 Comparison with other young clusters

The primordial mass segregation pattern observed in NGC 6383 is not unique. Chen, Grijs, and Zhao (2007) reported primordial segregation in NGC 2244,

Bonnell and Davies (1998) argued through N-body simulations that the positions of the massive stars in the Trapezium cluster cannot be explained by dynamical mass segregation and must reflect primordial formation near the cluster centre, and Sabbi et al. (2008) in NGC 346 in the SMC. Parker et al. (2014) explored mass segregation diagnostics through N-body simulations of diverse initial conditions, showing that different formation scenarios leave distinct signatures in the minimum spanning tree statistics and that geometry-robust methods are necessary to distinguish primordial from dynamical segregation in substructured systems younger than ~ 5 Myr.

NGC 6383 fits this picture: its age (~ 3.5 Myr) places it squarely in the regime where primordial conditions are expected to dominate over dynamical evolution, and its mass segregation signature extends to masses whose dynamical friction timescales exceed the cluster age. The consistency with both the simulation predictions and the observational pattern in other young clusters supports the conclusion that the central concentration of massive and binary stars in NGC 6383 is primordial or near-primordial in origin—where "near-primordial" encompasses both genuinely birth-segregated configurations and rapid early dynamical segregation during sub-clump coalescence, which are observationally indistinguishable at the present level of analysis. The strength of this conclusion rests not on any single diagnostic—no individual K–S test reaches overwhelming significance (Section 7.5.3)—but on the convergence of the mass-quartile analysis, the $t_{\rm seg}$-filtered test, the binary–single comparison, and the agreement with N-body predictions, each of which independently favours a non-secular origin.

8.2.5 Caveats

Two caveats apply. First, the ASteCA mass estimates are uncertain for PMS stars (Section 7.5.1), and the mass quartile boundaries depend on these estimates.

A systematic overestimation of PMS masses would shift the quartile boundaries and could artificially enhance the apparent segregation in the highest-mass bin. Second, the K–S p-values for the mass-quartile comparisons range from 0.04 to 0.95—the signal reaches marginal significance ($p = 0.04$) for the highest mass bin but is weaker for intermediate masses, and no formal multiple-testing correction has been applied. A larger sample or deeper photometry extending the mass function to lower masses would increase the statistical power of the segregation analysis.

8.3 Signatures of early expansion and gas loss

The dynamical youth of NGC 6383 ($\tau/t_{\rm rh} \approx 0.19$) places it in the regime where the effects of early gas expulsion should be observationally detectable. This section examines the theoretical expectations for post-gas-expulsion evolution and assesses which features of NGC 6383 are consistent with this scenario.

8.3.1 Theoretical framework: gas expulsion and supervirial expansion

Young stellar clusters form embedded in their natal molecular gas, which contributes a substantial fraction of the gravitational potential. When feedback processes—photoionisation, radiation pressure, stellar winds, and eventually supernova explosions—expel the residual gas on a timescale $t_{\rm gas}$, the gravitational potential drops abruptly. If $t_{\rm gas}$ is shorter than or comparable to the dynamical crossing time $t_{\rm cr}$, the stars retain their pre-expulsion velocities while the confining potential has weakened, and the system enters a supervirial state (Kroupa, 2001; Baumgardt and Kroupa, 2007).

The subsequent evolution depends on the star-formation efficiency (SFE)—the fraction of the natal gas mass converted into stars. Baumgardt and Kroupa (2007) showed through N-body simulations that clusters with SFE $\lesssim 30\%$ lose the majority of their membership during the post-gas-expulsion expansion, while clusters with SFE $\gtrsim 50\%$ remain bound and re-virialise on a timescale of several crossing times. At intermediate SFE, the system survives as a bound remnant surrounded by an expanding halo of unbound stars whose proper motions diverge slowly from the cluster mean.

This classical picture has recently been challenged. Ćalović et al. (2025) simulated star cluster evolution after gas removal using initial conditions drawn directly from radiation-hydrodynamic formation simulations, in which stars inherit realistic, spatially correlated velocities from the gas flow rather than being initialised in idealised virial equilibrium. Under these conditions, the post-removal evolution proceeds closer to virial balance than predicted by instantaneous-expulsion models: clusters that would dissolve in the classical framework remain bound, and those that survive do not exhibit the dramatic supervirial expansion that the standard SFE threshold implies. If this result generalises, the interpretation of NGC 6383's structural and kinematic properties may need to account for a softer dynamical response to gas loss than the Baumgardt and Kroupa (2007) simulations predict.

8.3.2 Observational indicators in NGC 6383

Several observational features of NGC 6383 are consistent with the post-gas-expulsion scenario, though none constitutes a definitive detection.

The large ratio $R_t/R_c = 40.4/1.95 \approx 20.7$ (Eqs. 7.5 and 7.4) (concentration $C = 1.316$; Section 7.3.2) indicates a compact core embedded in a diffuse, extended envelope. This morphology is a natural prediction of post-gas-expulsion evolution:

the bound core retains its pre-expulsion density while the surrounding material is expanding or unbound (Baumgardt and Kroupa, 2007).

The 39 sources between $R_{\rm Hill}$ and R_t with membership pseudoprobabilities $\tilde{p}_i > 0.6$ (Section 7.3.5) share the cluster's kinematic signature but lie beyond its gravitational influence sphere. If these are genuine former members rather than field contaminants, they represent a population of stars in the process of escaping the system—consistent with the prediction that gas expulsion unbinds a fraction of the membership that populates an expanding halo.

The Sagitta PMS ages span $\sim$1–6 Myr (Section 7.4.3), indicating that star formation was not instantaneous but extended over $\sim$5 Myr. The persistence of the surrounding H II region Sh 2-012 confirms that the local environment retains substantial ionised gas. Whether the cluster has already completed its gas expulsion phase or is currently undergoing a protracted, multi-episode removal process cannot be determined from the photometric data alone.

The elevated YSO fraction $Y_{\rm frac} = 0.28$ (Eq. 7.12) indicates that a significant fraction of the membership retains circumstellar discs, consistent with a system in which the disc-dissipation timescale has not yet elapsed. The classical estimate of a disc half-life $\lesssim$3 Myr and an overall disc lifetime of $\sim$6 Myr (Haisch Jr., A, and Lada, 2001) has been shown to be biased by an overrepresentation of high-mass stars in distant young clusters: Pfalzner, Dehghani, and Michel (2022) demonstrated that limiting-magnitude effects cause such samples to preferentially trace the shorter-lived discs around higher-mass hosts. When the analysis is restricted to nearby clusters or to specific stellar-mass bins, disc lifetimes increase substantially: Pfalzner and Dincer (2024) derived a characteristic Weibull lifetime of $\lambda \approx 9$–11 Myr for low-mass (M3.7–M6; $\sim$0.1–0.24 $M_\odot$) stars, while Ribas, Bouy, and Merín (2015) had already established that discs around higher-mass ($\gtrsim 2$ $M_\odot$) hosts dissipate earlier. The NGC 6383 YSO fraction is therefore consistent with

a cluster age of $\sim$3.5 Myr under both the classical and the mass-dependent disc-lifetime calibrations. The NGC 6383 YSO fraction is therefore consistent with a cluster age of $\sim$3.5 Myr under either the classical or revised disc-lifetime calibration.

8.3.3 Limitations of the current assessment

A definitive assessment of the cluster's virial state and expansion dynamics requires three-dimensional kinematic data: proper-motion dispersions and radial velocity dispersions, combined to form a three-dimensional velocity dispersion tensor. The current radial velocity sample (16 stars, $\sigma_{v_r} = 31.1$ km s^{-1}; Section 7.2.3) is insufficient for this purpose. The proper-motion dispersion alone ($\sigma_\mu \sim 0.15$ mas yr$^{-1} \approx 0.8$ km s^{-1} at $d = 1.1$ kpc) is consistent with the virial expectation for a cluster of this mass and size, but the two-dimensional projection cannot distinguish a virialised system from a mildly supervirial one.

A direct detection of systematic expansion would require measuring a radial gradient in the proper motions—a velocity field in which stars at larger radii move systematically outward from the cluster centre. The precision of the current proper motions ($\sigma_{\mu,i} \sim$ 0.02–0.5 mas yr^{-1}) is marginal for detecting expansion rates of $\sim$0.1 km s^{-1} pc^{-1}, which at $d = 1.1$ kpc correspond to $\sim$0.02 mas yr^{-1} arcmin^{-1}—at the edge of detectability for the bright members but well below the noise floor for the faint population.

Our best reading of the available evidence is that NGC 6383 is observed during or shortly after a post-gas-expulsion transition: the compact-core-plus-extended-envelope morphology, the kinematically coherent sources beyond the Hill radius, the elevated YSO fraction, and the extended PMS age distribution are collectively more naturally explained by a system that has recently lost a substantial fraction

of its natal gas than by a system that formed in or near its present structural configuration. A definitive virial-ratio measurement is not available from the current data—the radial velocity sample is too sparse and too contaminated by binary orbital motions (Section 7.2.3)—so this interpretation rests on structural and photometric indicators rather than on a direct kinematic diagnosis. The forthcoming Gaia DR4, with improved proper-motion precisions and extended radial velocity coverage, will provide the three-dimensional kinematics required for a quantitative virial state assessment.

8.4 Methodological lessons for Gaia-era cluster analysis

The analysis of NGC 6383 illustrates several methodological lessons that generalise beyond this particular system. We distil these here as principles for the design and interpretation of Gaia-era open cluster studies.

8.4.1 The hybrid density–mixture architecture

The HDBSCAN → pseudoprobability → parallax-clipping pipeline adopted in COSMIC reflects a practical reality of Gaia-era cluster science: no single algorithm captures both the non-parametric topology of the density field and the calibrated posterior probabilities required for downstream physical inference. HDBSCAN discovers the cluster without parametric assumptions on the component shapes, but its outputs are geometric persistence metrics lacking probabilistic calibration (Section 4.2.2). Gaussian mixture models produce calibrated posteriors but impose ellipsoidal structure that fails for elongated tidal features or fractal sub-populations (Section 4.4.3).

The combination of density-based discovery with probabilistic refinement is now the dominant operational paradigm (Cantat-Gaudin et al., 2018; Noormohammadi, Khakian Ghomi, and Javadi, 2024), and the present work confirms its utility for a young, moderately embedded cluster in a dense Galactic field. The principal architectural lesson is that the two stages should not be viewed as redundant but as complementary: HDBSCAN provides the topological scaffolding that initialises and constrains the parametric refinement, while the mixture model provides the probabilistic calibration that enables downstream uncertainty propagation.

The corollary is that the membership outputs of the two stages are fundamentally different objects—λ-persistence ratios versus posterior marginal probabilities—and must not be conflated in the interpretation. The COSMIC pseudoprobability $\tilde{p}_i$ of Eq. (4.16) is a composite of both, and its interpretation requires the doctrinal distinction established in Section 4.2.2.

8.4.2 Threshold sensitivity as a structural feature, not a deficiency

The sensitivity of the recovered population to the membership threshold (321 to 161 across four probability cuts; Section 7.1.2) is a structural feature of the problem, not a deficiency of the method. The physical summaries derived from hard membership lists—mean distance, velocity dispersion, mass function slope, structural radii—are threshold-conditioned quantities whose values change as the threshold moves. Reporting these quantities at a single threshold without documenting the sensitivity produces results that appear more precise than they are.

The formal remedy, developed in Section 4.2.6, is to compute physical summaries as expectations over the uncertain membership, weighting each star by its posterior

membership probability. In practice, the computational cost of full membership-marginalised inference remains high, and threshold-based catalogues remain the standard in the literature. The minimum requirement is to report the threshold-sensitivity analysis alongside the primary results, as we have done in Table 7.3 and Section 7.1.2, so that the reader can assess the robustness of any claimed physical property.

8.4.3 The diagnostic power of non-membership

The non-membership of HD 159176 (Section 7.7) illustrates the diagnostic power of the Gaia astrometric constraint for resolving long-standing astrophysical ambiguities. Without the proper-motion and parallax measurements, HD 159176—the brightest, most luminous star in the projected cluster field—would dominate the CMD fit and bias the inferred age, extinction, and distance modulus. Its exclusion on kinematic grounds shifts the inferred age from 6–10 Myr to $\sim$3.5 Myr, a factor-of-two revision driven entirely by membership refinement.

This example generalises to a broader principle: in any cluster study, the few brightest stars exert disproportionate influence on the isochronal fit because they define the turn-off morphology. If any of these stars are foreground or background contaminants—a possibility that increases in dense Galactic fields—the age and extinction estimates will be systematically biased. Astrometric membership confirmation for the brightest cluster candidates is therefore a prerequisite for reliable isochronal fitting, not an optional refinement.

8.4.4 The value of gradient-based samplers for isochronal inference

The replacement of ASteCA's default sampling backend with NUTS (Section 6.6.5) provides a concrete example of a broader methodological lesson: non-gradient-based samplers sacrifice diagnostic transparency for implementation simplicity. Genetic algorithms and parallel-tempered MCMC—the samplers historically used by ASteCA—provide no geometric diagnostics that formally certify convergence or identify regions of the posterior that the sampler has failed to represent. NUTS, by contrast, samples the exact posterior using Hamiltonian dynamics and provides geometric diagnostics—divergent transitions, E-BFMI, $\hat{R}$—that constitute formal proofs of sampling validity or failure (Section 4.1.6).

For the four-dimensional isochronal parameter space of NGC 6383, the computational overhead of NUTS is modest (minutes on a standard workstation), and the diagnostic guarantees it provides are invaluable for asserting the validity of the reported credible intervals. The lesson is directly applicable to Bayesian analyses of open cluster parameters and extends to any astrophysical inference where gradient-based sampling is computationally feasible: the diagnostic regime it provides is strictly preferable to approximate methods whose posterior accuracy is unverifiable.

8.5 Scalability to large-survey applications

The COSMIC framework was designed as a modular pipeline whose components can operate independently on different clusters without manual intervention (Section 6.1). This section assesses the degree to which the NGC 6383 analysis generalises to a systematic census of young open clusters in the Gaia era.

8.5.1 Architectural scalability

The HDBSCAN membership module requires only proper motions and a search cone; the parallax and structural models require standard Gaia five-parameter astrometry; the isochronal module accepts any photometric system for which isochrone grids are available. These minimal input requirements are satisfied for every cluster candidate accessible through the Gaia archive, making the pipeline immediately applicable to the $\sim 7\,000$ open cluster candidates catalogued in recent Gaia-era censuses (Hunt and Reffert, 2024; Cantat-Gaudin et al., 2020). The prior specifications documented in Chapter 6 are parameterised as configuration inputs rather than hard-coded values, allowing cluster-specific adjustment without modifying the source code.

The modular architecture permits selective execution: a user interested only in membership can run the HDBSCAN module alone, while a user interested in isochronal ages can condition on a pre-existing membership catalogue from the literature. This flexibility reduces the computational cost for large-sample applications in which not every inference component is required for every cluster.

8.5.2 Computational bottlenecks

The principal computational bottleneck is the NUTS-based sampling, whose cost scales with the dimensionality of the parameter space and the number of sources entering the likelihood. For the 254-member NGC 6383, convergent posterior samples for the four Bayesian models (distance, kinematics, structure, photometry) are obtained in ~ 10 minutes on a standard workstation. This cost is dominated by the isochronal fit, which evaluates the CMD likelihood for ~ 200 sources against a forward model that involves isochrone interpolation, binary assignment, and extinction application at each NUTS step.

For clusters with memberships exceeding $\sim 10^3$ sources—such as the rich clusters in the nearby spiral arms—the per-step cost of the CMD likelihood scales linearly with membership, and the total sampling time may reach hours. At $N > 10^4$, the hierarchical models may require sub-sampling strategies, mini-batch likelihoods, or amortised inference techniques (e.g., simulation-based inference with neural density estimators) to remain tractable. These extensions are not implemented in the current version of COSMIC and are identified as priorities for the next development cycle.

The HDBSCAN module scales as $\mathcal{O}(N \log N)$ in the number of input sources and is not a bottleneck for any foreseeable application. The King profile fitting and proper-motion model involve parameter spaces of dimension ~ 4–5 and converge rapidly regardless of sample size.

8.5.3 Planned extensions

Five extensions of COSMIC are planned to address the limitations identified in Section 6.9 and to enhance the framework's utility for large-survey applications.

First, a multi-dimensional membership model that operates jointly in astrometric, photometric, and spatial space through the generative mixture formalism of Eq. (4.6). This extension will unify the currently separated HDBSCAN and parallax-clipping stages into a single inference pass, producing calibrated posterior membership probabilities that propagate into all downstream analyses.

Second, the integration of Gaia XP spectrophotometry as an informative metallicity prior on the isochronal fit. Huang et al. (2025) and Nizovkina et al. (2025) demonstrated that XP-derived metallicities can collapse the age–metallicity degeneracy of Section 4.5.3, enabling precise age determinations even for clusters where broadband photometry alone is insufficient.

Third, a spatially resolved extinction model using Gaussian processes over the cluster field, replacing the zero-differential-reddening assumption of Section 6.9.4. This model will predict $A_V(\alpha, \delta)$ as a latent spatial field and marginalise over it during the isochronal fit, providing per-star extinction estimates and reducing the systematic bias on the age posterior.

Fourth, a multi-grid isochronal comparison that runs the full fitting pipeline with MIST, PARSEC, Baraffe, and SPOTS grids independently and reports the grid-to-grid variance as a systematic uncertainty component. This addresses the single-grid limitation of Section 6.9.5 and provides a quantitative estimate of the theoretical systematic floor on the isochronal age.

Fifth, the application of COSMIC to a systematic sample of young open cluster candidates selected from Gaia-era censuses. This survey will test the scalability of the framework under diverse observational conditions—varying field contamination, photometric depth, cluster richness, and extinction environment—and will establish whether the primordial mass segregation pattern observed in NGC 6383 is a generic feature of young Galactic clusters or a property specific to this system.

These extensions address the principal operational limitations of the current framework and position COSMIC as a general-purpose tool for the physical characterisation of open clusters in the Gaia DR4 era and beyond.

Chapter 9

Conclusions and Outlook

This thesis has presented a comprehensive Bayesian characterisation of the young open cluster NGC 6383, combining Gaia DR3 astrometry and photometry with 2MASS near-infrared data through the COSMIC framework. The formal machinery of Chapters 4–6 was developed to ensure that every physical claim is traceable to a generative model, a prior specification, and a convergence diagnostic, and that the domain of interpretive validity is stated explicitly rather than assumed. This final chapter summarises the principal physical findings (Section 9.1), the methodological contributions (Section 9.2), the main limitations that bound the generalisability of the results (Section 9.3), the most relevant future applications (Section 9.4), and the released data products and reproduction strategy (Section 9.5).

9.1 Physical conclusions

The analysis of NGC 6383 through the COSMIC framework, operating on Gaia DR3 astrometry and 2MASS photometry under the inferential architecture of Chapters 4–6, yields a coherent physical picture of a dynamically young

cluster whose present-day properties retain substantial memory of its formation conditions. This section synthesises the principal findings of Chapter 7 and their astrophysical interpretation in Chapter 8, emphasising what the analysis resolves, what it constrains without closing, and what remains open.

9.1.1 A resolved distance and a revised age

The COSMIC pipeline identified 254 probable members at a pseudoprobability threshold of $\tilde{p}_i > 0.6$, drawn from a working sample of 15 276 Gaia DR3 sources within 40 arcmin of the cluster centre (Section 7.1). The threshold-sensitivity analysis demonstrated that the recovered population varies from 321 to 161 members across four probability thresholds, quantifying the population-level uncertainty that the continuous membership formalism of Section 4.2 is designed to capture.

The hierarchical distance model yields $d = 1.110 \pm 0.060$ kpc (Eq. 7.1), confirmed independently by the isochronal distance modulus $\mu_{\rm DM} = 10.3 \pm 0.26$ mag (Eq. 7.11). This convergence between two methodologically independent inference pathways—parallax-based hierarchical modelling and CMD forward modelling—collapses the factor-of-three scatter in the historical distance literature (0.76–2.13 kpc; Table 1.1) to a $\sim 5\%$ consensus. The residual uncertainty is dominated by the parallax zero-point systematic at $b \approx 0.05^\circ$ (Section 5.2.1) rather than by reddening ambiguity, representing a qualitative shift in the error budget from photometric to astrometric systematics.

The modal isochronal age is $\tau = 3.53^{+1.40}_{-1.00}$ Myr (Eq. 7.8), with metallicity $Z = 0.024 \pm 0.008$ and visual extinction $A_V = 1.24 \pm 0.26$ mag. The convergence of three independent age indicators—the isochronal posterior, the Sagitta PMS age distribution ($\sim$1–6 Myr), and the elevated YSO fraction $Y_{\rm frac} = 0.28$ (95% CI

$[0.22, 0.34]$)—provides robust support for a cluster age of ~ 3.5 Myr with evidence for an extended star-formation episode with a duration of ~ 5 Myr (from ~ 1.6 to ~ 6.3 Myr ago). This age is consistent with the recent Gaia-era estimates of Hunt and Reffert (2024) and Kalari (2019), and excludes the older estimates of > 10 Myr. The 6–10 Myr estimate of Aidelman et al. (2018) was explicitly conditioned on the blue-straggler membership of HD 159176, which Section 7.7 rules out; the older estimates of ~ 20 Myr (Lindoff, 1968; Battinelli and Capuzzo-Dolcetta, 1991) are excluded independently by the CMD morphology of the bona fide membership.

The first central question of the thesis—whether and how strongly the adopted inferential construction of membership reshapes the physical interpretation of the cluster (Section 1.6)—receives a clear affirmative answer. The threshold-sensitivity analysis demonstrates that the recovered population varies from 321 to 161 members across four pseudoprobability thresholds, with each threshold producing quantitatively different structural extents, mass distributions, and outer-population counts. The HD 159176 non-membership result (Section 7.7) provides the most dramatic single illustration: a change in membership status for one star revises the cluster age by a factor of two and eliminates a formation narrative that persisted for four decades. More broadly, every physical quantity reported in Table 7.3 is conditioned on the membership construction documented in Section 6.2, and the threshold-dependent variation quantified in Table 7.1 shows that this conditioning is not merely formal but physically consequential.

9.1.2 HD 159176: non-membership and the collapse of an inherited narrative

The non-membership of HD 159176 (Section 7.7) is among the cleanest results of the thesis. The O6.5 V + O7 V binary's proper motion deviates from the cluster

mean by 6.6σ in declination, and its parallax places it $\sim$250 pc closer than the cluster—deviations that remain decisive even under the most conservative treatment of bright-star astrometric systematics. This result resolves a four-decade ambiguity: it removes the observational basis for the blue-straggler hypothesis of Aidelman et al. (2018), eliminates the triggered-formation scenario of Fitzgerald et al. (1978), and frees the cluster age from the evolutionary constraints of a single dominant system.

More broadly, the HD 159176 case illustrates a general principle: in young clusters observed through dense Galactic fields, the few brightest stars exert disproportionate leverage on the isochronal fit. A single unrecognised non-member at the luminous end of the CMD can bias the inferred age by a factor of two or more. Astrometric membership verification for the brightest candidates is therefore not an optional refinement but a prerequisite for reliable age determination.

9.1.3 Structural morphology: a compact core in a diffuse envelope

The King profile yields a compact core ($R_c = 1.95 \pm 0.19$ arcmin ≈ 0.63 pc) embedded in an extended envelope ($R_t = 40.4 \pm 14.3$ arcmin ≈ 13.1 pc), producing a concentration parameter $C = 1.316$ (Section 7.3). The core radius is well constrained ($\sim$10% relative uncertainty), whereas the tidal radius carries $\sim$35% uncertainty, reflecting the fundamental difficulty of constraining an outer boundary in a dense Galactic field where the cluster density approaches the background level. As established in Sections 2.3.4 and 2.4.1, fitted structural extent and physical tidal boundary are not equivalent; the tidal radius should be interpreted as a model-dependent estimate rather than a sharply defined edge.

The 39 kinematically coherent sources between the Hill radius ($R_{\rm Hill} = 28.3$ ar-

cmin) and the tidal radius are consistent with either potential escapers or field contaminants. Their approximately isotropic distribution argues against organised tidal tails but does not discriminate between the two interpretations. What can be stated is that the morphology—a dense, concentrated core surrounded by a progressively sparser halo extending well beyond the half-mass radius—is naturally predicted by post-gas-expulsion models in which a bound remnant retains its pre-expulsion concentration while the surrounding material expands or becomes unbound (Baumgardt and Kroupa, 2007).

9.1.4 Dynamical youth: memory of formation

The half-mass relaxation time $t_{\rm rh} = 18.4 \pm 2.1$ Myr exceeds the cluster age by a factor of ~ 5, placing NGC 6383 firmly in the pre-relaxation regime at $\tau/t_{\rm rh} \approx 0.19$ (Eq. 7.16). This ratio is the single most important diagnostic in the thesis: it establishes that two-body relaxation has not had sufficient time to redistribute stellar masses globally, erase primordial spatial structure, or drive the system toward energy equipartition. The observable properties of NGC 6383 are therefore expected to carry substantial information about the conditions under which the cluster formed.

This expectation is borne out by the mass segregation analysis. The central concentration of binary and high-mass stars persists when the sample is restricted to stars whose dynamical friction timescales exceed the cluster age ($t_{\rm seg} > \tau$; Section 7.5.4). Stars with masses below $\sim 8.29\ M_\odot$ cannot have migrated to their present radial positions through two-body relaxation within 3.5 Myr, yet they show the same central concentration pattern as the most massive objects. The binary population is systematically more centrally concentrated than the single-star population across all mass ranges, with the distributions diverging beyond ~ 5 arcmin from the centre (Section 7.5.3).

The appropriate caution—developed in Sections 8.2 and 1.2.3—is that "primordial" segregation need not mean primordial in the strict sense of spatially segregated birth. The rapid-collapse pathway of Allison et al. (2009) and the sub-clump merging scenario of McMillan, Vesperini, and Portegies Zwart (2007) can both produce centrally concentrated massive-star distributions on timescales shorter than $t_{\rm rh}$, without requiring that massive stars formed preferentially at the centre of the natal cloud. The binary–single radial difference provides the strongest discriminant: it requires an explanation independent of total-mass-dependent dynamical friction, since the total system mass of a binary is only modestly larger than that of its primary. Our best reading of the evidence is that the central concentration of binaries in NGC 6383 reflects a formation channel—whether competitive accretion in the densest natal regions or rapid segregation during sub-clump coalescence—rather than secular post-formation evolution. No single statistical test is individually decisive (the binary–single K–S p-value is 0.07), but the convergence of the mass-quartile analysis, the $t_{\rm seg}$-filtered test, the binary–single comparison, and the consistency with N-body predictions collectively supports this interpretation.

9.1.5 The virial state: what can and cannot be said

The virial state of NGC 6383 was identified in Section 1.6 as one of the three central scientific questions of the thesis. The honest conclusion is that the available data constrain but do not close this question. The proper-motion dispersion ($\sigma_\mu \sim 0.15$ mas yr^{-1} ≈ 0.8 km s^{-1} at $d = 1.1$ kpc) is consistent with the virial expectation for a cluster of this mass and size, but the radial velocity coverage (16 stars with $\sigma_{v_r} = 31.1$ km s^{-1}) is too sparse and too contaminated by binary orbital motions to provide a meaningful third kinematic dimension. A definitive virial ratio cannot be computed.

What can be stated within the claim-status discipline is the following. First, the structural and kinematic indicators are consistent with a post-gas-expulsion state: the morphology of a concentrated core within an extended envelope, the kinematic coherence of sources beyond the Hill radius, the elevated YSO fraction, and the extended PMS age distribution all fit the qualitative predictions of gas-expulsion models. Second, the recent simulations of Ćalović et al. (2025)—in which stars inherit realistic velocities from hydrodynamic formation rather than idealised virial equilibrium—suggest that the post-gas-removal dynamical response may be gentler than the classical Baumgardt and Kroupa (2007) framework predicts, potentially explaining why the proper-motion dispersion appears roughly virial despite the cluster's extreme youth. Third, a quantitative virial state assessment requires radial velocities for a mass-representative subsample of at least $\sim$50–100 members, a dataset that Gaia DR4 and forthcoming multi-object spectrographs (4MOST, MOONS) are expected to provide.

The virial question is therefore constrained to a physically informative envelope—the cluster is young, structurally complex, and plausibly supervirial—without being closed by a single diagnostic number. We regard this as an appropriate conclusion for a system at $\tau/t_{\rm rh} \approx 0.19$, where the formal machinery developed in Section 2.1.2 explicitly warns against collapsing a non-equilibrium dynamical state into a scalar classificatory label.

9.1.6 Synthesis: what NGC 6383 teaches about young cluster interpretation

Taken together, the results of this thesis support a coherent physical picture of NGC 6383. It is a $\sim$3.5 Myr old system at $d \approx 1.1$ kpc, solar to moderately super-solar in metallicity ($Z = 0.024 \pm 0.008$), that formed through a non-instantaneous

episode of star formation lasting ~ 5 Myr within the Sh 2-012 region. Its stellar content retains the spatial imprint of formation: massive stars and binaries are centrally concentrated in a pattern that two-body relaxation cannot have produced at the current cluster age. The cluster has survived early gas loss with a bound core intact, surrounded by an extended envelope that may include material in the process of escaping the system. The historically dominant O-type binary HD 159176 is a foreground interloper whose assumed membership distorted four decades of published ages and formation narratives.

But the thesis also demonstrates that what NGC 6383 teaches is inseparable from how it is analysed. The recovered population is not a fixed property of the cluster but a function of the membership threshold, the clustering algorithm, and the quality filters applied to the catalogue. The inferred age depends on the adopted isochrone grid, the treatment of differential reddening, and the multiplicity model. The structural parameters depend on the background subtraction and the profile family. The segregation signal depends on the mass estimates, which are themselves uncertain for PMS stars. None of these dependencies invalidates the physical conclusions; all of them bound their interpretive scope.

The broader lesson is therefore methodological as much as astrophysical. Young open clusters in the Gaia era are not objects whose properties can be read transparently from the data. They are inference-sensitive systems in which membership construction, model choice, and uncertainty propagation condition the physical conclusions that may legitimately be drawn. The claim-status discipline developed in Chapter 3 and enforced throughout this thesis is not an exercise in epistemological caution for its own sake; it is the minimum requirement for making defensible physical claims about systems whose observable properties remain entangled with their inferential construction.

9.2 Methodological contributions

The principal methodological contribution of this thesis is the COSMIC framework itself: an open-source, modular pipeline that integrates density-based clustering with Bayesian hierarchical modelling for the complete physical characterisation of open clusters from Gaia-era data (Pulgar-Escobar et al., 2024). Five specific advances are identified below; their significance extends beyond the particular case of NGC 6383 to the broader practice of Gaia-era cluster science.

First, the hybrid membership architecture of Section 6.2 combines the non-parametric discovery power of HDBSCAN with iterative pseudoprobability refinement to suppress the false-positive rate inherent to density-based clustering in dense Galactic fields. The pseudoprobability metric $\tilde{p}_i$ of Eq. (4.16) provides a practical ranking that captures both spatial centrality and recovery robustness, while the tiered structure (members at $\tilde{p}_i \geq 0.8$, probable members at $0.6 \leq \tilde{p}_i < 0.8$) preserves the continuous character of the membership information. The architecture addresses a practical reality of Gaia-era cluster work: no single algorithm captures both the non-parametric topology of the density field and the calibrated posterior probabilities required for downstream physical inference (Section 8.4.1).

Second, the replacement of ASteCA's default sampling backend with NUTS via PyMC 5 (Section 6.6.5) provides access to the exact posterior and to the geometric convergence diagnostics—divergent transitions, E-BFMI, $\hat{R}$—that constitute formal guarantees on sampling validity unavailable in non-gradient-based samplers. This substitution is computationally inexpensive for the four-dimensional isochronal parameter space and should be adopted as the default wherever gradient-based sampling is feasible. The practical consequence is that every posterior reported in this thesis is accompanied by a verifiable convergence certificate, rather

than an unauditable claim of convergence from a non-diagnostic sampler.

Third, the formal separation between the HDBSCAN λ-persistence ratio and the posterior marginal membership probability of Eq. (4.7) is enforced as a doctrinal distinction throughout the analysis (Section 4.2.2). This prevents the conflation of two fundamentally different statistical objects—a geometric centrality metric and a calibrated conditional probability—in the interpretation of membership-dependent results. The distinction is not merely taxonomic: it determines whether downstream physical summaries are computed under calibrated probabilistic weights or under uncalibrated geometric rankings, with material consequences for the uncertainty attached to every derived quantity.

Fourth, the threshold-sensitivity analysis (Section 7.1.2) demonstrates that physical summaries derived from hard membership lists are threshold-conditioned quantities whose values change materially as the threshold moves, quantifying the practical extent of the population-level indeterminacy formalised in Section 4.2.3. This practice—reporting the dependence of results on the membership boundary rather than presenting a single threshold as definitive—should become standard in Gaia-era cluster work, where the transition between cluster and field is typically diffuse rather than sharp.

Fifth, the integrated claim-status discipline of Section 3.5 governs every physical claim in Chapters 7–8, preventing the silent universalisation of model-specific results by requiring that each result is stated together with its conditioning variables. The tripartite distinction among data-driven, model-conditioned, and interpretive claims provides a practical labelling scheme that any multi-method cluster analysis can adopt without requiring the full philosophical apparatus from which it derives.

Collectively, these contributions address a gap in the Gaia-era open cluster literature: the need for pipelines that are not only automated and reproducible,

but formally transparent about the assumptions on which their outputs depend. COSMIC does not resolve every limitation identified in the preceding chapters, but it makes those limitations visible and traceable—a prerequisite for improving them.

9.3 Limitations

The physical conclusions and methodological contributions of this thesis are bounded by operational limitations that constrain the generalisability and precision of the results. These are stated here as binding conditions on the inference; each connects to a specific component of the COSMIC pipeline and is documented in technical detail in Section 6.9. The purpose of this section is not to repeat that documentation but to assess the scientific consequences of these limitations for the principal claims of the thesis.

9.3.1 Membership construction

The membership pipeline operates exclusively in two-dimensional proper-motion space, with parallax entering only through a deterministic sigma-clipping step and photometric information absent from the membership construction entirely (Section 6.9.1). This design choice was deliberate—proper motions provide the strongest cluster–field contrast for NGC 6383—but it means that the membership inference does not exploit the full information content of the Gaia five-parameter solution. A photometrically discrepant field star whose proper motion coincidentally matches the cluster passes the membership test undetected. The pseudoprobability $\tilde{p}_i$ lacks the formal calibration of a posterior marginal probability (Section 6.9.2), so the membership tiers used throughout Chapters 7–8 are operationally convenient but inferentially informal. We use $\tilde{p}_i$ as a practical membership proxy

despite its lack of formal calibration, and the threshold-sensitivity analysis of Table 7.1 quantifies the resulting uncertainty; but full membership-marginalised physical summaries in the sense of Eq. (4.8) remain an aspiration rather than an achievement of the current pipeline.

The scientific consequence is that the recovered population is more uncertain than the formal error bars on derived quantities suggest. The spread from 321 to 161 members across four thresholds (Section 7.1.2) indicates that the cluster periphery, the low-mass end of the luminosity function, and the outer structural profile are all threshold-conditioned. Any downstream quantity that depends on these features—the tidal radius, the mass function slope, the outer segregation signal—inherits this conditioning.

9.3.2 Isochronal systematics

The isochronal fit adopts a single isochrone grid (MIST) under zero differential reddening (Sections 6.9.5 and 6.9.4). No alternative PMS evolutionary models—such as PARSEC (Bressan et al., 2012), Baraffe et al. (Baraffe et al., 2015), Siess et al. (Siess, Dufour, and Forestini, 2000), or SPOTS (Somers, Cao, and Pinsonneault, 2020)—were explored. Fitting the same photometric data with different grids can produce systematic age offsets of 0.1–0.3 dex for PMS populations, arising from differences in the treatment of convective boundary mixing, initial helium abundance, atmospheric boundary conditions, and magnetic activity (Bell et al., 2013). This grid-to-grid systematic variance is not propagated into the reported uncertainties and constitutes an unquantified systematic floor on the age accuracy. The model-discrepancy formalism developed in Section 4.5.6 provides the framework for incorporating this variance, but its application was deferred to future work.

The zero-differential-reddening assumption is a further source of systematic concern. NGC 6383 lies within the H II region Sh 2-012, where spatially variable extinction is physically expected. Uniform-extinction models average over this structure, potentially biasing the age and extinction posteriors in ways that the statistical credible intervals do not capture. A spatially resolved extinction model would address this limitation but was not implemented in the current version of COSMIC.

The combined effect of these two systematics—single grid and zero differential reddening—means that the reported age uncertainty of $\tau = 3.53^{+1.40}_{-1.00}$ Myr represents the statistical precision under the MIST assumption with spatially uniform extinction, not the total uncertainty inclusive of theoretical model systematics and environmental reddening. A specific systematic of known direction operates through the treatment of magnetic activity: standard non-magnetic models such as MIST are known to underestimate ages for magnetically active PMS populations by $\sim$30–50% relative to lithium-depletion-boundary calibrations (Binks et al., 2021; Jeffries et al., 2023). If this correction applies to NGC 6383, the modal age may be closer to $\sim$4.5–5.3 Myr—still within the upper statistical uncertainty, but systematically higher than the MIST-based point estimate. The qualitative conclusions of the thesis—dynamical youth at $\tau/t_{\rm rh} < 0.3$ and primordial segregation—are robust to this shift, but the absolute age should be interpreted with this systematic in mind. The true uncertainty on the age is likely broader than the reported credible interval, though the convergence of three independent age indicators (Section 9.1.1) provides confidence that the modal age is not grossly in error.

9.3.3 Kinematic and completeness boundaries

The Gaia DR3 radial velocity coverage ($\sim$11.4% of the membership) is insufficient for a three-dimensional virial state assessment (Section 6.9.6). This limitation is

not a deficiency of the COSMIC pipeline but a boundary condition imposed by the survey: the Gaia RVS is magnitude-limited at $G_{\rm RVS} \lesssim 14$ mag, so radial velocities are available only for the brightest, most massive members. The consequence is that one of the three central scientific questions of the thesis—the dynamical-state interpretation of NGC 6383—is constrained to a qualitative envelope rather than a quantitative scalar. This was known at the outset of the project; the question was retained as a central theme because the formal framework for its eventual answer needed to be developed regardless of whether the current data could close it.

The Gaia photometric completeness truncates the recovered mass function at $\sim$0.3–0.5 $M_\odot$ (Section 5.5.1). The 254 members reported in Chapter 7 represent the recoverable membership above the Gaia detection threshold, not a complete census of the cluster population. The total cluster mass is therefore a lower bound, the mass function slope is constrained only above the completeness mass, and the spatial extent may underrepresent the true halo if faint, low-mass members preferentially populate the periphery. The 2MASS completeness limit further restricts the PMS census and YSO analysis to the brighter, infrared-detected subsample (Section 6.9.7).

9.3.4 Generalisability

Beyond these component-level limitations, the COSMIC pipeline has been validated on a single young, moderately rich cluster in the Galactic plane. Its performance on fundamentally different cluster environments—old globular clusters, sparse embedded protoclusters, rich clusters in extreme crowding—has not been tested. The priors, hyperparameter grids, and diagnostic thresholds documented in Chapter 6 are tuned to the NGC 6383 regime and may require revision for clusters with substantially different distances, ages, extinction environments,

or membership counts. Whether the primordial mass segregation pattern observed in NGC 6383 is a generic feature of young Galactic clusters or a property specific to this system cannot be determined from a single-object study. The systematic application of COSMIC to a sample of young cluster candidates (Section 9.4) is required to answer that question.

9.4 Future directions

The limitations documented in Section 9.3 define a concrete development programme for COSMIC, whose technical specification is given in Section 8.5. This section summarises the programme in terms of its expected scientific impact, distinguishing between methodological upgrades to the pipeline, observational advances that will become available in the near term, and the scientific questions that the combination of both will enable.

9.4.1 Pipeline upgrades

The most consequential upgrade replaces the two-stage HDBSCAN + clipping architecture with a single-stage generative mixture model operating jointly in astrometric, photometric, and spatial space through Eq. (4.6). This extension will produce calibrated posterior membership probabilities in a single inference pass, enabling the full membership-marginalised physical summaries formalised in Eq. (4.8) and eliminating the principal inferential gap in the current pipeline. The practical consequence will be that physical quantities—mean distance, velocity dispersion, mass function slope, structural radii—are computed as expectations over the uncertain membership rather than as fixed-sample statistics at an arbitrary threshold.

The integration of Gaia XP spectrophotometric metallicities as informative priors

on the isochronal fit will address the age–metallicity degeneracy of Section 4.5.3. Huang et al. (2025) and Nizovkina et al. (2025) demonstrated that XP-derived metallicities can collapse this covariance, enabling precise age determinations even for clusters where broadband photometry alone is insufficient. For NGC 6383, where the metallicity posterior is broad ($Z = 0.024 \pm 0.008$) and covariant with age, this constraint is expected to narrow the age posterior substantially.

The implementation of spatially resolved extinction through Gaussian process models over the cluster field (Edenhofer et al., 2024) will replace the zero-differential-reddening assumption of Section 6.9.4. For a cluster embedded within Sh 2-012, where spatially variable extinction is physically expected, this extension will provide per-star extinction estimates and reduce the systematic bias on the age posterior.

A multi-grid isochronal comparison, running the full fitting pipeline independently with MIST, PARSEC, Baraffe, and SPOTS grids, will quantify the theoretical systematic floor that the current single-grid analysis cannot capture. The model-discrepancy formalism of Eq. (4.18) provides the framework for incorporating this grid-to-grid variance into the posterior, transforming what is currently an unquantified systematic into a formally propagated uncertainty component.

9.4.2 Observational advances

Gaia DR4—expected in late 2026 with ~ 5.5 yr of data and proper-motion precisions $\sim \sqrt{2}$ better than DR3—will directly address the two most consequential data-quality limitations of the current analysis. The extended radial velocity sample will enable the first quantitative virial state assessment of NGC 6383 (Section 7.6.3), closing the gap identified in Section 9.3.3. The improved proper-motion precisions will enhance sensitivity to internal kinematic substructure—expansion

gradients, rotation, and sub-group kinematics—that the current data cannot resolve at the level required for a spatially resolved dynamical diagnosis.

Dedicated multi-object spectroscopic follow-up from instruments now entering commissioning will complement Gaia by providing radial velocities and chemical abundances for the faint, low-mass PMS population that is numerically dominant but absent from the current RVS sample. 4MOST, which achieved first light in October 2025 and is expected to begin science operations in 2026, will deliver $R \sim 20\,000$ spectra for millions of southern-hemisphere targets, with the young open cluster programme specifically designed to cover systems such as NGC 6383. MOONS, shipped to the VLT in December 2025 with first light anticipated in mid-2026, will provide near-infrared multi-object spectroscopy that is especially powerful for extincted PMS populations in star-forming environments. Together, these instruments will provide the three-dimensional kinematic and chemical data required for a definitive dynamical and chemical assessment.

On the photometric side, the VVV/VVVX near-infrared surveys reach $K_s \approx 18$ mag—approximately four magnitudes deeper than the 2MASS 10σ limit—and the VIRAC2 proper-motion catalogue (Smith et al., 2024) provides astrometry for > 545 million sources. Exploiting these data would extend the mass function below the current 2MASS completeness threshold and improve the census of disc-bearing PMS stars in the cluster outskirts, substantially strengthening the mass segregation and YSO analyses.

9.4.3 Scientific questions for the next stage

Three scientific questions emerge naturally from the present work and are addressable with the combination of pipeline upgrades and observational advances described above.

First, what is the quantitative virial state of NGC 6383, and does it support the post-gas-expulsion interpretation advanced in Section 8.3? This requires radial velocities for a mass-representative subsample and a three-dimensional kinematic model that accounts for expansion, rotation, and perspective effects simultaneously. The formal framework is already in place (Sections 2.1.2 and 6.4.1); the missing ingredient is the data.

Second, is the primordial mass segregation pattern observed in NGC 6383 a generic feature of young Galactic clusters, or a property specific to this system? The application of COSMIC to a systematic sample of young cluster candidates from the catalogues of Hunt and Reffert (2024) and Cantat-Gaudin et al. (2020) will address this question. The modular architecture of COSMIC enables this survey without modification of the core inference engine, and the threshold-sensitivity and claim-status protocols developed in this thesis provide the interpretive framework for comparing segregation signals across clusters with different richnesses, ages, and environments.

Third, can the apparent PMS age spread in NGC 6383 be decomposed into a genuine star-formation duration and a methodological component arising from accretion history heterogeneity, magnetic activity, differential reddening, and unresolved multiplicity? The multi-grid comparison and spatially resolved extinction model will remove two major confounders, while the deeper near-infrared photometry from VVV/VVVX will extend the PMS census to lower masses where the competing effects have different relative magnitudes. A definitive decomposition may remain out of reach from photometry alone, but the combination of narrower age posteriors, per-star extinction, and an extended mass baseline will substantially sharpen the constraints.

9.5 Data products and reproducibility

The scientific claims of this thesis are supported by a complete release of software, catalogues, and posterior samples, documented in technical detail in Section 6.10. This section describes the released products and the epistemological commitment that motivates their form.

9.5.1 Software

The COSMIC source code is publicly available under an open-source licence at `https://github.com/notluquis/COSMIC`. The repository contains the complete pipeline: data ingestion and quality filtering, HDBSCAN membership determination with pseudoprobability construction, the PyMC 5 implementations of the hierarchical distance model, the proper-motion model, the King profile model, and the modified ASteCA interface with the NUTS sampler. The Sagitta interface and the YSO classification module are included as auxiliary components. All software dependencies and their versions are specified, ensuring that the computational environment can be reconstructed exactly.

9.5.2 Catalogues and posterior samples

The membership catalogue, containing all 321 sources with $\tilde{p}_i > 0.5$, will be archived at the Centre de Données astronomiques de Strasbourg (CDS). For each source, the catalogue includes the Gaia DR3 source identifier, corrected astrometry, Gaia and 2MASS photometry, HDBSCAN persistence value, iteration fraction, pseudoprobability, membership tier, and all derived quantities: ASteCA mass and binary probability, Sagitta PMS probability and age, and Q-parameter YSO classification.

The posterior samples for all four Bayesian models—distance, proper motion, King profile, and isochronal fit—are released as serialised ArviZ InferenceData objects containing the full MCMC traces, prior and posterior predictive distributions, and convergence diagnostics. These objects enable the community to recompute any summary statistic, test any hypothesis, and condition on any alternative prior without re-running the sampler—operations that are impossible when only a point estimate and a symmetric error bar are reported.

9.5.3 Epistemological commitment

The release of full posterior samples as first-class data products reflects a position that the thesis has developed throughout: the posterior distribution, not a point estimate, is the scientific output of a Bayesian analysis. A single best-fit value with symmetric error bars discards the shape, asymmetry, multimodality, and covariance structure of the posterior—precisely the features that encode the degeneracies, prior dependencies, and identifiability limitations that this thesis has argued are central to the interpretation of young open clusters. The posterior is the complete answer to the question "what does the data, combined with the model, say about the parameters?" and any summary of it is a lossy compression. We recommend the release of full posteriors as a standard practice for Bayesian astrophysical analyses, and we note that the computational cost of archiving the posterior samples (~ 100 MB for the four models in this thesis) is negligible relative to the cost of the observations and the analysis that produced them.

References

Abril-Pla, O. et al. (Sept. 1, 2023). "PyMC: a modern, and comprehensive probabilistic programming framework in Python". en. In: *PeerJ. Computer Science* 9, e1516. ISSN: 2376-5992. DOI: 10.7717/peerj-cs.1516. URL: http://dx.doi.org/10.7717/peerj-cs.1516.

Adamo, A. et al. (June 2020). "Star clusters near and far: Tracing star formation across cosmic time". en. In: *Space Science Reviews* 216. ISSN: 0038-6308,1572-9672. DOI: 10.1007/s11214-020-00690-x. URL: http://dx.doi.org/10.1007/s11214-020-00690-x.

Aerts, C. et al. (July 19, 2018). "Forward asteroseismic modeling of stars with a convective core from gravity-mode oscillations: Parameter estimation and stellar model selection". In: *The Astrophysical Journal. Supplement Series* 237, p. 15. ISSN: 0067-0049,1538-4365. DOI: 10.3847/1538-4365/aaccfb. URL: http://dx.doi.org/10.3847/1538-4365/aaccfb.

Aerts, C., S. Mathis, and T. M. Rogers (Aug. 18, 2019). "Angular momentum transport in Stellar interiors". en. In: *Annual Review of Astronomy and Astrophysics* 57, pp. 35–78. ISSN: 0066-4146,1545-4282. DOI: 10.1146/annurev-astro-091918-104359. URL: http://dx.doi.org/10.1146/annurev-astro-091918-104359.

Agarwal, M. et al. (Feb. 11, 2021). "ML-MOC: Machine Learning (kNN and GMM) based Membership determination for Open Clusters". en. In: *Monthly Notices*

of the Royal Astronomical Society 502, pp. 2582–2599. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stab118. URL: http://dx.doi.org/10.1093/mnras/stab118.

Aggarwal, C. C., A. Hinneburg, and D. A. Keim (Oct. 12, 2001). "On the surprising behavior of distance metrics in high dimensional space". In: *Database Theory — ICDT 2001*. Ed. by G. Goos et al. Vol. 1973. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 420–434. ISBN: 9783540414568,9783540445036. DOI: 10.1007/3-540-44503-x_27. URL: http://dx.doi.org/10.1007/3-540-44503-x_27.

Aidelman, Y. et al. (2018). "Open clusters: III. Fundamental parameters of B stars in NGC 6087, NGC 6250, NGC 6383 and NGC 6530. B-type stars with circumstellar envelopes". In: *åp* 610, A30. URL: https://www.aanda.org/articles/aa/pdf/2018/02/aa30995-17.pdf.

Albrow, M. D. (Feb. 15, 2024). "The frequency and mass-ratio distribution of binaries in clusters II: radial segregation in the nearby dissolving open clusters Hyades and Praesepe". en. In: *Monthly Notices of the Royal Astronomical Society* 528, pp. 6211–6220. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stae425. URL: http://dx.doi.org/10.1093/mnras/stae425.

Alexander, J. S. and M. D. Albrow (Nov. 26, 2024). "The frequency and mass-ratio distribution of binaries in clusters – III: Probabilistic generative modelling of six young open clusters". en. In: *Monthly Notices of the Royal Astronomical Society*. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stae2636. URL: http://dx.doi.org/10.1093/mnras/stae2636.

Alfonso, J., A. García-Varela, and K. Vieira (Sept. 2024). "Exploring Galactic open clusters with *Gaia*: I. An examination in the first kiloparsec". In: *Astronomy and Astrophysics* 689, A18. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202450901. URL: http://dx.doi.org/10.1051/0004-6361/202450901.

Allison, R. J. et al. (Aug. 1, 2009). "Dynamical mass segregation on a very short timescale". In: *The Astrophysical Journal* 700, pp. L99–L103. ISSN: 0004-637X,1538-4357. DOI: 10.1088/0004-637x/700/2/l99. URL: http://dx.doi.org/10.1088/0004-637x/700/2/l99.

Anderson, L. et al. (Sept. 10, 2018). "Improving Gaia parallax precision with a data-driven model of stars". In: *The Astronomical Journal* 156, p. 145. ISSN: 0004-6256,1538-3881. DOI: 10.3847/1538-3881/aad7bf. URL: http://dx.doi.org/10.3847/1538-3881/aad7bf.

Andreon, S. and M. Hurn (Jan. 2013). "Measurement errors and scaling relations in astrophysics: a review". en. In: *Statistical Analysis and Data Mining* 6, pp. 15–33. ISSN: 1932-1864,1932-1872. DOI: 10.1002/sam.11173. URL: http://dx.doi.org/10.1002/sam.11173.

Arnold, B. and N. J. Wright (May 13, 2024). "Kinematic substructure in star clusters constrains star cluster formation". en. In: *Monthly Notices of the Royal Astronomical Society* 531, pp. 1191–1202. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stae1236. URL: http://dx.doi.org/10.1093/mnras/stae1236.

Asplund, M. et al. (2009). "The Chemical Composition of the Sun". In: *Annu. Rev. Astron. Astrophys.* 47, pp. 481–522. ISSN: 0066-4146,1545-4282. DOI: 10.1146/annurev.astro.46.060407.145222.

Astraatmadja, T. L. and C. A. L. Bailer-Jones (Nov. 23, 2016). "Estimating distances from parallaxes. Ii. Performance of Bayesian distance estimators on agaia-like catalogue". In: *The Astrophysical Journal* 832, p. 137. ISSN: 0004-637X,1538-4357. DOI: 10.3847/0004-637x/832/2/137. URL: http://dx.doi.org/10.3847/0004-637x/832/2/137.

El-Badry, K. (June 2024). "Gaia's binary star renaissance". en. In: *New Astronomy Reviews* 98, p. 101694. ISSN: 1387-6473,1872-9630. DOI: 10.1016/j.newar.2024.101694. URL: http://dx.doi.org/10.1016/j.newar.2024.101694.

Bailer-Jones, C. A. L. (Dec. 1, 2023). "Estimating distances from parallaxes. VI. A method for inferring distances and transverse velocities from parallaxes and proper motions demonstrated on Gaia data release 3". In: *The Astronomical Journal* 166, p. 269. ISSN: 0004-6256,1538-3881. DOI: 10.3847/1538-3881/ad08bb. URL: http://dx.doi.org/10.3847/1538-3881/ad08bb.

Bailer-Jones, C. A. L. et al. (Aug. 1, 2018). "Estimating distance from parallaxes. IV. Distances to 1.33 billion stars in Gaia data release 2". In: *The Astronomical Journal* 156, p. 58. ISSN: 0004-6256,1538-3881. DOI: 10.3847/1538-3881/aacb21. URL: http://dx.doi.org/10.3847/1538-3881/aacb21.

Bailer-Jones, C. A. L. et al. (Mar. 1, 2021). "Estimating distances from parallaxes. V. geometric and photogeometric distances to 1.47 billion stars in Gaia early data release 3". In: *The Astronomical Journal* 161, p. 147. ISSN: 0004-6256,1538-3881. DOI: 10.3847/1538-3881/abd806. URL: http://dx.doi.org/10.3847/1538-3881/abd806.

Bailer-Jones, C. A. L. (Oct. 2015). "Estimating Distances from Parallaxes". In: *Publications of the Astronomical Society of the Pacific. Astronomical Society of the Pacific* 127, pp. 994–1009. ISSN: 0004-6280,1538-3873. DOI: 10.1086/683116. URL: http://dx.doi.org/10.1086/683116.

Balbinot, E. and M. Gieles (Feb. 21, 2018). "The devil is in the tails: the role of globular cluster mass evolution on stream properties". In: *Monthly Notices of the Royal Astronomical Society* 474, pp. 2479–2492. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stx2708. URL: http://dx.doi.org/10.1093/mnras/stx2708.

Ballesteros-Paredes, J. (Oct. 11, 2006). "Six myths on the virial theorem for interstellar clouds". en. In: *Monthly Notices of the Royal Astronomical Society* 372, pp. 443–449. ISSN: 0035-8711,1365-2966. DOI: 10.1111/j.1365-2966.2006.10880.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2006.10880.x.

Banfield, J. D. and A. E. Raftery (Sept. 1993). "Model-based Gaussian and non-Gaussian clustering". In: *Biometrics* 49, p. 803. ISSN: 0006-341X,1541-0420. DOI: 10.2307/2532201. URL: http://dx.doi.org/10.2307/2532201.

Baraffe, I., G. Chabrier, and J. Gallardo (Sept. 1, 2009). "Episodic accretion at early stages of evolution of low-mass stars and brown dwarfs: A solution for the observed luminosity spread in h-r diagrams?" In: *The Astrophysical Journal* 702, pp. L27–L31. ISSN: 0004-637X,1538-4357. DOI: 10.1088/0004-637x/702/1/l27. URL: http://dx.doi.org/10.1088/0004-637x/702/1/l27.

Baraffe, I. et al. (May 2015). "New evolutionary models for pre-main sequence and main sequence low-mass stars down to the hydrogen-burning limit". In: *Astronomy and Astrophysics* 577, A42. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/201425481. URL: http://dx.doi.org/10.1051/0004-6361/201425481.

Baribault, B. and A. G. E. Collins (Feb. 2025). "Troubleshooting Bayesian cognitive models". en. In: *Psychological Methods* 30, pp. 128–154. ISSN: 1082-989X,1939-1463. DOI: 10.1037/met0000554. URL: http://dx.doi.org/10.1037/met0000554.

Barmby, P. et al. (June 2007). "Structural parameters for globular clusters in M31 and generalizations for the fundamental plane". en. In: *The Astronomical Journal* 133, pp. 2764–2786. ISSN: 0004-6256,1538-3881. DOI: 10.1086/516777. URL: http://dx.doi.org/10.1086/516777.

Bastian, N., K. R. Covey, and M. R. Meyer (Aug. 1, 2010). "A universal stellar initial mass function? A critical look at variations". en. In: *Annual Review of Astronomy and Astrophysics* 48, pp. 339–389. ISSN: 0066-4146,1545-4282. DOI: 10.1146/annurev-astro-082708-101642. URL: http://dx.doi.org/10.1146/annurev-astro-082708-101642.

Battinelli, P. and R. Capuzzo-Dolcetta (Mar. 1991). "Formation and evolutionary properties of the galactic open cluster system". In: *Monthly Notices of the*

Royal Astronomical Society 249, pp. 76–83. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/249.1.76. URL: http://dx.doi.org/10.1093/mnras/249.1.76.

Baumgardt, H. (Aug. 21, 2001). "Scaling of N-body calculations". In: *Monthly Notices of the Royal Astronomical Society* 325, pp. 1323–1331. ISSN: 0035-8711,1365-2966. DOI: 10.1046/j.1365-8711.2001.04272.x. URL: http://dx.doi.org/10.1046/j.1365-8711.2001.04272.x.

Baumgardt, H. and M. Hilker (Aug. 1, 2018). "A catalogue of masses, structural parameters, and velocity dispersion profiles of 112 Milky Way globular clusters". en. In: *Monthly Notices of the Royal Astronomical Society* 478, pp. 1520–1557. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/sty1057. URL: http://dx.doi.org/10.1093/mnras/sty1057.

Baumgardt, H. and P. Kroupa (Oct. 13, 2007). "A comprehensive set of simulations studying the influence of gas expulsion on star cluster evolution". en. In: *Monthly Notices of the Royal Astronomical Society* 380, pp. 1589–1598. ISSN: 0035-8711,1365-2966. DOI: 10.1111/j.1365-2966.2007.12209.x. arXiv: 0707.1944 [astro-ph]. URL: http://dx.doi.org/10.1111/j.1365-2966.2007.12209.x.

Baumgardt, H. and J. Makino (Mar. 21, 2003). "Dynamical evolution of star clusters in tidal fields". en. In: *Monthly Notices of the Royal Astronomical Society* 340, pp. 227–246. ISSN: 0035-8711,1365-2966. DOI: 10.1046/j.1365-8711.2003.06286.x. URL: http://dx.doi.org/10.1046/j.1365-8711.2003.06286.x.

Becker, W. and R. Fenkart (Oct. 1971). "A catalogue of galactic star clusters observed in three colours". In: *åps* 4, p. 241.

Bell, C. P. M. et al. (July 11, 2013). "Pre-main-sequence isochrones – II. Revising star and planet formation time-scales". In: *Monthly Notices of the Royal*

Astronomical Society 434, pp. 806–831. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stt1075. URL: http://dx.doi.org/10.1093/mnras/stt1075.

Belokurov, V. et al. (Aug. 1, 2020). "Unresolved stellar companions with Gaia DR2 astrometry". en. In: *Monthly Notices of the Royal Astronomical Society* 496, pp. 1922–1940. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/staa1522. URL: http://dx.doi.org/10.1093/mnras/staa1522.

Bernardo, J. M. and A. F. M. Smith (Mar. 28, 2000). *Bayesian Theory*. en. Wiley Series in Probability and Statistics. John Wiley & Sons. ISBN: 9780471494645.

Betancourt, M. J. and M. Girolami (2013). "Hamiltonian Monte Carlo for hierarchical models". In: *arXiv [stat.ME]*. DOI: 10.48550/ARXIV.1312.0906. eprint: 1312.0906 (stat.ME). URL: http://dx.doi.org/10.48550/ARXIV.1312.0906.

Betancourt, M. (Apr. 3, 2016). "Diagnosing Suboptimal Cotangent Disintegrations in Hamiltonian Monte Carlo". In: *arXiv [stat.ME]*. arXiv: 1604.00695 [stat.ME]. URL: http://arxiv.org/abs/1604.00695.

— (Jan. 9, 2017). "A conceptual introduction to Hamiltonian Monte Carlo". In: *arXiv [stat.ME]*. arXiv: 1701.02434 [stat.ME]. URL: http://arxiv.org/abs/1701.02434.

— (Mar. 22, 2018). "Calibrating model-based inferences and decisions". In: *arXiv [stat.ME]*. arXiv: 1803.08393 [stat.ME]. URL: http://arxiv.org/abs/1803.08393.

Beyer, K. et al. (1999). "When is "nearest neighbor" meaningful?" In: *Lecture Notes in Computer Science*. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 217–235. ISBN: 9783540654520,9783540492573. DOI: 10.1007/3-540-49257-7_15. URL: http://dx.doi.org/10.1007/3-540-49257-7_15.

Bianchini, P. et al. (Mar. 2, 2016a). "A novel look at energy equipartition in globular clusters". In: *arXiv [astro-ph.GA]*. DOI: 10.48550/ARXIV.1603.00878. arXiv: 1603.00878 [astro-ph.GA]. URL: http://arxiv.org/abs/1603.00878.

Bianchini, P. et al. (June 1, 2016b). "A novel look at energy equipartition in globular clusters". en. In: *Monthly Notices of the Royal Astronomical Society* 458, pp. 3644–3654. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stw552. URL: http://dx.doi.org/10.1093/mnras/stw552.

Bianchini, P. et al. (June 7, 2018). "The internal rotation of globular clusters revealed by Gaia DR2". In: *arXiv [astro-ph.GA]*. arXiv: 1806.02580 [astro-ph.GA]. URL: http://arxiv.org/abs/1806.02580.

Binks, A. S. et al. (June 1, 2021). "The Gaia-ESO survey: a lithium depletion boundary age for NGC 2232". en. In: *Monthly Notices of the Royal Astronomical Society* 505, pp. 1280–1292. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stab1351. URL: http://dx.doi.org/10.1093/mnras/stab1351.

Binney, J. and S. Tremaine (Dec. 31, 2008). *Galactic dynamics: Second edition*. en. 2nd ed. Princeton Series in Astrophysics. Princeton University Press. ISBN: 9781400828722. DOI: 10.1515/9781400828722. URL: http://dx.doi.org/10.1515/9781400828722.

Bodensteiner, J., T. Shenar, and H. Sana (Sept. 2020). "Investigating the lack of main-sequence companions to massive Be stars". In: *Astronomy and Astrophysics* 641, A42. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202037640. URL: http://dx.doi.org/10.1051/0004-6361/202037640.

Boer, T. J. L. de et al. (June 1, 2019). "Globular cluster number density profiles using Gaia DR2". en. In: *Monthly Notices of the Royal Astronomical Society* 485, pp. 4906–4935. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stz651. URL: http://dx.doi.org/10.1093/mnras/stz651.

Bokulich, A. (Dec. 2020). "Towards a taxonomy of the model-ladenness of data". en. In: *Philosophy of Science* 87, pp. 793–806. ISSN: 0031-8248,1539-767X. DOI: 10.1086/710516. URL: http://dx.doi.org/10.1086/710516.

Bonaca, A. and A. M. Price-Whelan (June 2025). "Stellar streams in the Gaia era". en. In: *New Astronomy Reviews* 100, p. 101713. ISSN: 1387-6473,1872-9630.

DOI: 10.1016/j.newar.2024.101713. URL: http://dx.doi.org/10.1016/j.newar.2024.101713.

Bonnell, I. A. and M. R. Bate (July 21, 2006). "Star formation through gravitational collapse and competitive accretion". en. In: *Monthly Notices of the Royal Astronomical Society* 370, pp. 488–494. ISSN: 0035-8711,1365-2966. DOI: 10.1111/j.1365-2966.2006.10495.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2006.10495.x.

Bonnell, I. A. and M. B. Davies (Apr. 1998). "Mass segregation in young stellar clusters". In: *Monthly Notices of the Royal Astronomical Society* 295, pp. 691–698. ISSN: 0035-8711,1365-2966. DOI: 10.1046/j.1365-8711.1998.01372.x. URL: http://dx.doi.org/10.1046/j.1365-8711.1998.01372.x.

Bossini, D. et al. (Mar. 2019). "Age determination for 269 Gaia DR2 open clusters". In: *Astronomy and Astrophysics* 623, A108. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/201834693. URL: http://dx.doi.org/10.1051/0004-6361/201834693.

Boubert, D. and A. Everall (Oct. 1, 2020). "Completeness of the Gaia verse II: what are the odds that a star is missing from Gaia DR2?" en. In: *Monthly Notices of the Royal Astronomical Society* 497, pp. 4246–4261. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/staa2305. arXiv: 2005.08983 [astro-ph.GA]. URL: http://dx.doi.org/10.1093/mnras/staa2305.

— (Nov. 7, 2021). "A selection function toolbox for subsets of astronomical catalogues". In: *arXiv [astro-ph.IM]*. arXiv: 2111.04126 [astro-ph.IM]. URL: http://arxiv.org/abs/2111.04126.

Bovy, J. (Feb. 3, 2015). "galpy: A python LIBRARY FOR GALACTIC DYNAMICS". In: *The Astrophysical Journal. Supplement Series* 216, p. 29. ISSN: 0067-0049,1538-4365. DOI: 10.1088/0067-0049/216/2/29. URL: http://dx.doi.org/10.1088/0067-0049/216/2/29.

Bovy, J., D. W. Hogg, and S. T. Roweis (June 1, 2011). "Extreme deconvolution: Inferring complete distribution functions from noisy, heterogeneous and incomplete observations". In: *The Annals of Applied Statistics* 5, pp. 1657–1677. ISSN: 1932-6157,1941-7330. DOI: 10.1214/10-aoas439. URL: http://dx.doi.org/10.1214/10-aoas439.

Box, G. E. P. (Dec. 1976). "Science and statistics". en. In: *Journal of the American Statistical Association* 71, pp. 791–799. ISSN: 0162-1459,1537-274X. DOI: 10.1080/01621459.1976.10480949. URL: http://dx.doi.org/10.1080/01621459.1976.10480949.

Breen, P. G., A. L. Varri, and D. C. Heggie (Nov. 1, 2017). "The kinematic richness of star clusters – I. Isolated spherical models with primordial anisotropy". en. In: *Monthly Notices of the Royal Astronomical Society* 471, pp. 2778–2789. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stx1750. URL: http://dx.doi.org/10.1093/mnras/stx1750.

Bressan, A. et al. (Nov. 21, 2012). "parsec: stellar tracks and isochrones with the PAdova and TRieste Stellar Evolution Code: parsec:tracks and isochrones". en. In: *Monthly Notices of the Royal Astronomical Society* 427, pp. 127–145. ISSN: 0035-8711,1365-2966. DOI: 10.1111/j.1365-2966.2012.21948.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2012.21948.x.

Brinkmann, N. et al. (Apr. 2017). "The bound fraction of young star clusters". In: *Astronomy and Astrophysics* 600, A49. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/201629312. URL: http://dx.doi.org/10.1051/0004-6361/201629312.

Brynjarsdóttir, J. and A. O'Hagan (Nov. 1, 2014). "Learning about physical parameters: the importance of model discrepancy". In: *Inverse Problems* 30, p. 114007. ISSN: 0266-5611,1361-6420. DOI: 10.1088/0266-5611/30/11/114007. URL: http://dx.doi.org/10.1088/0266-5611/30/11/114007.

Buckner, A. S. M. and D. Froebrich (Dec. 1, 2013). "Properties of star clusters – I. Automatic distance and extinction estimates". In: *Monthly Notices of the Royal Astronomical Society* 436, pp. 1465–1478. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stt1665. URL: http://dx.doi.org/10.1093/mnras/stt1665.

Buckner, A. S. M. et al. (Feb. 2019). "The spatial evolution of young massive clusters". In: *Astronomy and Astrophysics* 622, A184. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/201832936. URL: http://dx.doi.org/10.1051/0004-6361/201832936.

Buckner, A. S. M. et al. (June 29, 2022). "Observational bias and young massive cluster characterization – I. 2D perspective effects". en. In: *Monthly Notices of the Royal Astronomical Society* 514, pp. 4087–4101. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stac1297. URL: http://dx.doi.org/10.1093/mnras/stac1297.

Ćalović, A. et al. (Oct. 28, 2025). "The long-term effects of gas removal from hydrodynamic simulations of star formation". en. In: *Monthly Notices of the Royal Astronomical Society* 544, pp. 1109–1123. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/staf1795. URL: http://dx.doi.org/10.1093/mnras/staf1795.

Cameron, E. (2011). "On the estimation of confidence intervals for binomial population proportions in astronomy: The simplicity and superiority of the Bayesian approach". en. In: *Publications of the Astronomical Society of Australia* 28, pp. 128–139. ISSN: 1323-3580,1448-6083. DOI: 10.1071/as10046. URL: http://dx.doi.org/10.1071/AS10046.

Campello, R. J. G. B., D. Moulavi, and J. Sander (2013). "Density-based clustering based on hierarchical density estimates". In: *Advances in Knowledge Discovery and Data Mining*. Ed. by J. Pei et al. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 160–172. ISBN: 9783642374555,9783642374562.

DOI: 10.1007/978-3-642-37456-2_14. URL: http://dx.doi.org/10.1007/978-3-642-37456-2_14.

Campello, R. J. G. B. et al. (July 27, 2015). "Hierarchical density estimates for data clustering, visualization, and outlier detection". en. In: *ACM Transactions on Knowledge Discovery From Data* 10, pp. 1–51. ISSN: 1556-4681,1556-472X. DOI: 10.1145/2733381. URL: http://dx.doi.org/10.1145/2733381.

Cantat-Gaudin, T. and L. Casamiquela (Dec. 2024). "How Gaia sheds light on the Milky Way star cluster population". en. In: *New Astronomy Reviews* 99, p. 101696. ISSN: 1387-6473,1872-9630. DOI: 10.1016/j.newar.2024.101696. URL: http://dx.doi.org/10.1016/j.newar.2024.101696.

Cantat-Gaudin, T. et al. (Oct. 2018). "A Gaia DR2 view of the open cluster population in the Milky Way". In: *Astronomy and Astrophysics* 618, A93. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/201833476. URL: http://dx.doi.org/10.1051/0004-6361/201833476.

Cantat-Gaudin, T. et al. (Aug. 2020). "Painting a portrait of the Galactic disc with its stellar clusters". In: *Astronomy and Astrophysics* 640, A1. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202038192. URL: http://dx.doi.org/10.1051/0004-6361/202038192.

Cantat-Gaudin, T. (Feb. 9, 2022). "Milky Way star clusters and Gaia: A review of the ongoing revolution". en. In: *Universe* 8, p. 111. ISSN: 2218-1997. DOI: 10.3390/universe8020111. URL: http://dx.doi.org/10.3390/universe8020111.

Cantat-Gaudin, T. and T. D. Brandt (May 2021). "Characterizing and correcting the proper motion bias of the bright Gaia EDR3 sources". In: *Astronomy and Astrophysics* 649, A124. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202140807. URL: http://dx.doi.org/10.1051/0004-6361/202140807.

Cantat-Gaudin, T. et al. (Jan. 2023). "An empirical model of the *Gaia* DR3 selection function". In: *Astronomy and Astrophysics* 669, A55. ISSN: 0004-

6361,1432-0746. DOI: 10.1051/0004-6361/202244784. URL: http://dx.doi.org/10.1051/0004-6361/202244784.

Cardelli, J. A., G. C. Clayton, and J. S. Mathis (Oct. 1989). "The relationship between infrared, optical, and ultraviolet extinction". en. In: *The Astrophysical Journal* 345, p. 245. ISSN: 0004-637X,1538-4357. DOI: 10.1086/167900. URL: http://dx.doi.org/10.1086/167900.

Carrera, R. et al. (July 2019a). "Extended halo of NGC 2682 (M 67) from Gaia DR2". In: *Astronomy and Astrophysics* 627, A119. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/201935599. URL: http://dx.doi.org/10.1051/0004-6361/201935599.

Carrera, R. et al. (Mar. 2019b). "Open clusters in APOGEE and GALAH: Combining Gaia and ground-based spectroscopic surveys". In: *Astronomy and Astrophysics* 623, A80. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/201834546. URL: http://dx.doi.org/10.1051/0004-6361/201834546.

Carroll, B. W. and D. A. Ostlie (July 18, 2006). *An introduction to modern astrophysics*. en. 2nd ed. Pearson Education. ISBN: 9780805304022.

Cartwright, A. and A. P. Whitworth (Feb. 2004). "The statistical analysis of star clusters". In: *Monthly Notices of the Royal Astronomical Society* 348, pp. 589–598. ISSN: 0035-8711,1365-2966. DOI: 10.1111/j.1365-2966.2004.07360.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2004.07360.x.

Castro-Ginard, A. et al. (May 2022). "Hunting for open clusters in *Gaia* EDR3: 628 new open clusters found with OCfinder". In: *Astronomy and Astrophysics* 661, A118. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202142568. URL: http://dx.doi.org/10.1051/0004-6361/202142568.

Castro-Ginard, A. et al. (Aug. 2024). "*Gaia* DR3 detectability of unresolved binary systems". In: *Astronomy and Astrophysics* 688, A1. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202450172. URL: http://dx.doi.org/10.1051/0004-6361/202450172.

Cerviño, M. and V. Luridiana (Jan. 2004). "Physical limits to the validity of synthesis models". In: *Astronomy and Astrophysics* 413, pp. 145–157. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361:20031454. URL: http://dx.doi.org/10.1051/0004-6361:20031454.

Chabrier, G. (2003). "Galactic Stellar and Substellar Initial Mass Function". In: *Publ. Astron. Soc. Pac.* 115.809, pp. 763–795. ISSN: 0004-6280,1538-3873. DOI: 10.1086/376392.

Chabrier, G., P. Hennebelle, and S. Charlot (Nov. 6, 2014). "Variations of the stellar initial mass function in the progenitors of massive early-type galaxies and in extreme starburst environments". In: *The Astrophysical Journal* 796, p. 75. ISSN: 0004-637X,1538-4357. DOI: 10.1088/0004-637x/796/2/75. URL: http://dx.doi.org/10.1088/0004-637X/796/2/75.

Chandrasekhar, S. (Mar. 19, 1970). *Ellipsoidal figures of equilibrium*. en. Mrs. H. E. Silliman Memorial Lecture. Yale University Press. ISBN: 9780300011166.

Chazal, F. et al. (Nov. 2013). "Persistence-based clustering in Riemannian manifolds". en. In: *Journal of the ACM* 60, pp. 1–38. ISSN: 0004-5411,1557-735X. DOI: 10.1145/2535927. URL: http://dx.doi.org/10.1145/2535927.

Chen, L., R. de Grijs, and J. L. Zhao (Oct. 2007). "Mass segregation in very young open clusters: A case study of NGC 2244 and NGC 6530". en. In: *The Astronomical Journal* 134, pp. 1368–1379. ISSN: 0004-6256,1538-3881. DOI: 10.1086/521022. URL: http://dx.doi.org/10.1086/521022.

Choi, J. et al. (May 26, 2016). "Mesa isochrones and stellar tracks (mist). I. Solar-scaled models". In: *The Astrophysical Journal* 823, p. 102. ISSN: 0004-637X,1538-4357. DOI: 10.3847/0004-637x/823/2/102. URL: http://dx.doi.org/10.3847/0004-637x/823/2/102.

Claydon, I., M. Gieles, and A. Zocchi (May 1, 2017). "The properties of energetically unbound stars in stellar clusters". en. In: *Monthly Notices of the Royal*

Astronomical Society 466, pp. 3937–3950. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stw3309. URL: http://dx.doi.org/10.1093/mnras/stw3309.

Cottaar, M., M. R. Meyer, and R. J. Parker (Nov. 2012). "Characterizing a cluster's dynamic state using a single epoch of radial velocities". In: *Astronomy and Astrophysics* 547, A35. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/201219673. URL: http://dx.doi.org/10.1051/0004-6361/201219673.

Cuevas-Otahola, B. et al. (Feb. 11, 2020). "Structural analysis of disc super star clusters of M82: size and profile shape at intermediate ages". en. In: *Monthly Notices of the Royal Astronomical Society* 492, pp. 993–1007. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stz3524. URL: http://dx.doi.org/10.1093/mnras/stz3524.

Cummings, J. D. et al. (Oct. 5, 2018). "The white dwarf initial–final mass relation for progenitor stars from 0.85 to 7.5 M ⊙". In: *The Astrophysical Journal* 866, p. 21. ISSN: 0004-637X,1538-4357. DOI: 10.3847/1538-4357/aadfd6. URL: http://dx.doi.org/10.3847/1538-4357/aadfd6.

Daffern-Powell, E. C. and R. J. Parker (Apr. 21, 2020). "Dynamical evolution of fractal structures in star-forming regions". en. In: *Monthly Notices of the Royal Astronomical Society* 493, pp. 4925–4935. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/staa575. URL: http://dx.doi.org/10.1093/mnras/staa575.

Dalessandro, E. et al. (Aug. 27, 2015). "No evidence of mass segregation in the low-mass galactic globular cluster ngc 6101". In: *The Astrophysical Journal* 810, p. 40. ISSN: 0004-637X,1538-4357. DOI: 10.1088/0004-637x/810/1/40. URL: http://dx.doi.org/10.1088/0004-637x/810/1/40.

Daniel, K. J., D. C. Heggie, and A. L. Varri (June 2017). "An approximate analytic model of a star cluster with potential escapers". en. In: *Monthly Notices of the Royal Astronomical Society* 468, pp. 1453–1473. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stx571. URL: http://dx.doi.org/10.1093/mnras/stx571.

Dasgupta, A. and A. E. Raftery (Mar. 1998). "Detecting features in spatial point processes with clutter via model-based clustering". en. In: *Journal of the American Statistical Association* 93, pp. 294–302. ISSN: 0162-1459,1537-274X. DOI: 10.1080/01621459.1998.10474110. URL: http://dx.doi.org/10.1080/01621459.1998.10474110.

De Becker, M. et al. (Mar. 2004). "AnXMM-Newtonobservation of the massive binary HD 159176". In: *Astronomy and Astrophysics* 416, pp. 221–233. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361:20031710. URL: http://dx.doi.org/10.1051/0004-6361:20031710.

Deb, S., A. Baruah, and S. Kumar (Aug. 11, 2022). "Ensemble-based unsupervised machine learning method for membership determination of open clusters using Mahalanobis distance". en. In: *Monthly Notices of the Royal Astronomical Society* 515, pp. 4685–4701. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stac2116. URL: http://dx.doi.org/10.1093/mnras/stac2116.

Dias, W. S. et al. (Apr. 17, 2021). "Updated parameters of 1743 open clusters based on Gaia DR2". en. In: *Monthly Notices of the Royal Astronomical Society* 504, pp. 356–371. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stab770. URL: http://dx.doi.org/10.1093/mnras/stab770.

Dib, S., J. Kim, and M. Shadmehri (Oct. 1, 2007). "The origin of the Arches stellar cluster mass function". In: *Monthly Notices of the Royal Astronomical Society. Letters* 381, pp. L40–L44. ISSN: 1745-3925,1745-3933. DOI: 10.1111/j.1745-3933.2007.00362.x. URL: http://dx.doi.org/10.1111/j.1745-3933.2007.00362.x.

Dib, S., S. Schmeja, and R. J. Parker (Jan. 1, 2018). "Structure and mass segregation in Galactic stellar clusters". en. In: *Monthly Notices of the Royal Astronomical Society* 473, pp. 849–859. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stx2413. URL: http://dx.doi.org/10.1093/mnras/stx2413.

Dickson, N. et al. (May 11, 2023). "Multimass modelling of Milky Way globular clusters – I. Implications on their stellar initial mass function above 1 M⊙". en. In: *Monthly Notices of the Royal Astronomical Society* 522, pp. 5320–5339. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stad1254. URL: http://dx.doi.org/10.1093/mnras/stad1254.

Dickson, N. et al. (Feb. 27, 2024). "Multimass modelling of milky way globular clusters – II. Present-day black hole populations". en. In: *Monthly Notices of the Royal Astronomical Society* 529, pp. 331–347. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stae470. URL: http://dx.doi.org/10.1093/mnras/stae470.

Ding, Y. et al. (Nov. 2024). "Analysis of the *Gaia* Data Release 3 parallax bias in the Galactic plane". In: *Astronomy and Astrophysics* 691, A81. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202450967. URL: http://dx.doi.org/10.1051/0004-6361/202450967.

Ding, Y. et al. (Apr. 1, 2025). "Analysis of the Gaia data release 3 parallax bias at bright magnitudes". In: *The Astronomical Journal* 169, p. 211. ISSN: 0004-6256,1538-3881. DOI: 10.3847/1538-3881/adba44. URL: http://dx.doi.org/10.3847/1538-3881/adba44.

Dinnbier, F. and P. Kroupa (Aug. 2020). "Tidal tails of open star clusters as probes of early gas expulsion". In: *Astronomy and Astrophysics* 640, A84. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/201936570. URL: http://dx.doi.org/10.1051/0004-6361/201936570.

Dinnbier, F. and S. Walch (Oct. 10, 2020). "How fast do young star clusters expel their natal gas? Estimating the upper limit of the gas expulsion timescale". en. In: *Monthly Notices of the Royal Astronomical Society* 499, pp. 748–767. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/staa2560. URL: http://dx.doi.org/10.1093/mnras/staa2560.

Dominguez, R. et al. (July 24, 2017). "How fast is mass-segregation happening in hierarchical formed embedded star clusters?" In: *arXiv [astro-ph.GA]*. arXiv: 1707.07710 [astro-ph.GA]. URL: http://arxiv.org/abs/1707.07710.

Domínguez, R. et al. (Oct. 29, 2021). "Are hierarchically formed embedded star clusters surviving gas expulsion depending on their initial conditions?" en. In: *Monthly Notices of the Royal Astronomical Society* 508, pp. 5410–5424. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stab2644. URL: http://dx.doi.org/10.1093/mnras/stab2644.

Donada, J. et al. (July 2023). "The multiplicity fraction in 202 open clusters from*Gaia*". In: *Astronomy and Astrophysics* 675, A89. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202245219. URL: http://dx.doi.org/10.1051/0004-6361/202245219.

Dotter, A. (Jan. 14, 2016). "Mesa isochrones and stellar tracks (mist) 0: Methods for the construction of stellar isochrones". In: *The Astrophysical Journal. Supplement Series* 222, p. 8. ISSN: 0067-0049,1538-4365. DOI: 10.3847/0067-0049/222/1/8. URL: http://dx.doi.org/10.3847/0067-0049/222/1/8.

Duchêne, G. and A. Kraus (Aug. 18, 2013). "Stellar multiplicity". en. In: *Annual Review of Astronomy and Astrophysics* 51, pp. 269–310. ISSN: 0066-4146,1545-4282. DOI: 10.1146/annurev-astro-081710-102602. URL: http://dx.doi.org/10.1146/annurev-astro-081710-102602.

Eadie, G. and M. Jurić (Apr. 20, 2019). "The cumulative mass profile of the Milky Way as determined by globular cluster kinematics from Gaia DR2". In: *The Astrophysical Journal* 875, p. 159. ISSN: 0004-637X,1538-4357. DOI: 10.3847/1538-4357/ab0f97. URL: http://dx.doi.org/10.3847/1538-4357/ab0f97.

Eadie, G. M., J. J. Webb, and J. S. Rosenthal (Feb. 1, 2022). "Bayesian inference of globular cluster properties using distribution functions". In: *The Astrophysical*

Journal 926, p. 211. ISSN: 0004-637X,1538-4357. DOI: 10.3847/1538-4357/ac4494. URL: http://dx.doi.org/10.3847/1538-4357/ac4494.

Edenhofer, G. et al. (May 2024). "A parsec-scale Galactic 3D dust map out to 1.25 kpc from the Sun". In: *Astronomy and Astrophysics* 685, A82. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202347628. URL: http://dx.doi.org/10.1051/0004-6361/202347628.

Eggen, O. (Jan. 1961). "Three-colour photometry in the southern hemisphere : NGC 6383, NGC 6405 and standard stars". In: *Royal Greenwich Observatory Bulletins* 27, pp. 61–69.

Elson, R. A. W., S. M. Fall, and K. C. Freeman (Dec. 1987). "The structure of young star clusters in the Large Magellanic Cloud". en. In: *The Astrophysical Journal* 323, p. 54. ISSN: 0004-637X,1538-4357. DOI: 10.1086/165807. URL: http://dx.doi.org/10.1086/165807.

Ernst, A. et al. (May 11, 2007). "N-body models of rotating globular clusters". en. In: *Monthly Notices of the Royal Astronomical Society* 377, pp. 465–479. ISSN: 0035-8711,1365-2966. DOI: 10.1111/j.1365-2966.2007.11602.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2007.11602.x.

Evans, M. and H. Moshonov (Dec. 1, 2006). "Checking for prior-data conflict". In: *Bayesian Analysis* 1, pp. 893–914. ISSN: 1936-0975,1931-6690. DOI: 10.1214/06-ba129. URL: http://dx.doi.org/10.1214/06-ba129.

Evans, T. H. H. L. (Jan. 1968). "Photometry of star clusters". Saint Andrews University, UK.

Everall, A. and D. Boubert (Dec. 17, 2021). "Completeness of the Gaia verse – V. Astrometry and radial velocity sample selection functions in Gaia EDR3". en. In: *Monthly Notices of the Royal Astronomical Society* 509, pp. 6205–6224. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stab3262. URL: http://dx.doi.org/10.1093/mnras/stab3262.

Fabricius, C. et al. (May 2021). "*Gaia*Early Data Release 3: Catalogue validation". In: *Astronomy and Astrophysics* 649, A5. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202039834. URL: http://dx.doi.org/10.1051/0004-6361/202039834.

Farias, J. P. et al. (July 1, 2015). "The difficult early stages of embedded star clusters and the importance of the pre-gas expulsion virial ratio". en. In: *Monthly Notices of the Royal Astronomical Society* 450, pp. 2451–2458. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stv790. URL: http://dx.doi.org/10.1093/mnras/stv790.

Feiden, G. A. (Sept. 2016). "Magnetic inhibition of convection and the fundamental properties of low-mass stars". In: *Astronomy and Astrophysics* 593, A99. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/201527613. URL: http://dx.doi.org/10.1051/0004-6361/201527613.

Feinstein, A. (July 1994). "The young open clusters in the Galaxy". In: *Rev. Mex. Astron. Astrofis.* 29, pp. 141–147.

Fitzgerald, M. P. et al. (Apr. 1, 1978). "NGC 6383 - I. The central core". In: *Monthly Notices of the Royal Astronomical Society* 182, pp. 607–616. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/182.4.607. URL: http://dx.doi.org/10.1093/mnras/182.4.607.

Fitzpatrick, E. L. et al. (Nov. 26, 2019). "An analysis of the shapes of interstellar extinction curves. VII. Milky way spectrophotometric optical-through-ultraviolet extinction and its R-dependence". In: *The Astrophysical Journal* 886, p. 108. ISSN: 0004-637X,1538-4357. DOI: 10.3847/1538-4357/ab4c3a. URL: http://dx.doi.org/10.3847/1538-4357/ab4c3a.

Fleck, J. .-. J. et al. (July 1, 2006). "On the mass of dense star clusters in starburst galaxies from spectrophotometry". en. In: *Monthly Notices of the Royal Astronomical Society* 369, pp. 1392–1406. ISSN: 0035-8711,1365-2966. DOI:

10.1111/j.1365-2966.2006.10390.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2006.10390.x.

Fukushige, T. and D. C. Heggie (Nov. 2000). "The time-scale of escape from star clusters". In: *Monthly Notices of the Royal Astronomical Society* 318, pp. 753–761. ISSN: 0035-8711,1365-2966. DOI: 10.1046/j.1365-8711.2000.03811.x. URL: http://dx.doi.org/10.1046/j.1365-8711.2000.03811.x.

Fuller, W. A. (June 30, 1987). *Measurement Error Models*. Ed. by W. A. Fuller. Wiley. ISBN: 9780471861874,9780470316665. DOI: 10.1002/9780470316665. URL: http://dx.doi.org/10.1002/9780470316665.

G. van de Ven et al. (Jan. 2006). "The dynamical distance and intrinsic structure of the globular cluster ω Centauri". In: *Astronomy and Astrophysics* 445, pp. 513–543. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361:20053061. URL: http://dx.doi.org/10.1051/0004-6361:20053061.

Gabry, J. et al. (Feb. 1, 2019). "Visualization in Bayesian workflow". en. In: *Journal of the Royal Statistical Society. Series A, (Statistics in Society)* 182, pp. 389–402. ISSN: 0964-1998,1467-985X. DOI: 10.1111/rssa.12378. URL: http://dx.doi.org/10.1111/rssa.12378.

Gaia Collaboration et al. (June 1, 2023). "*Gaia* Data Release 3: Summary of the content and survey properties". In: *Astronomy and Astrophysics* 674, A1. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202243940. URL: http://dx.doi.org/10.1051/0004-6361/202243940.

Galli, P. A. B. et al. (Feb. 2021). "Chamaeleon DANCe". In: *Astronomy and Astrophysics* 646, A46. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202039395. URL: http://dx.doi.org/10.1051/0004-6361/202039395.

Garrison, R. F. (Mar. 1967). "Some characteristics of the B and a stars in the upper Scorpius complex". en. In: *The Astrophysical Journal* 147, p. 1003. ISSN: 0004-637X,1538-4357. DOI: 10.1086/149090. URL: http://dx.doi.org/10.1086/149090.

Gavagnin, E. et al. (Dec. 21, 2017). "Star cluster formation in a turbulent molecular cloud self-regulated by photoionization feedback". en. In: *Monthly Notices of the Royal Astronomical Society* 472, pp. 4155–4172. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stx2222. URL: http://dx.doi.org/10.1093/mnras/stx2222.

Gelman, A. (Sept. 1, 2006). "Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)". In: *Bayesian Analysis* 1, pp. 515–534. ISSN: 1936-0975,1931-6690. DOI: 10.1214/06-ba117a. URL: http://dx.doi.org/10.1214/06-ba117a.

Gelman, A., D. Simpson, and M. Betancourt (Oct. 19, 2017). "The prior can often only be understood in the context of the likelihood". en. In: *Entropy (Basel, Switzerland)* 19, p. 555. ISSN: 1099-4300. DOI: 10.3390/e19100555. URL: http://dx.doi.org/10.3390/e19100555.

Gelman, A. et al. (Nov. 27, 2013). *Bayesian Data Analysis*. 3rd ed. Chapman & Hall/CRC Texts in Statistical Science. Chapman and Hall/CRC. ISBN: 9780429113079. DOI: 10.1201/b16018. URL: http://dx.doi.org/10.1201/b16018.

Gelman, A. et al. (2020). "Bayesian Workflow". In: *arXiv [stat.ME]*. DOI: 10.48550/ARXIV.2011.01808. eprint: 2011.01808 (stat.ME). URL: http://dx.doi.org/10.48550/ARXIV.2011.01808.

Getman, K. V. et al. (June 11, 2018). "Young star clusters in nearby molecular clouds". en. In: *Monthly Notices of the Royal Astronomical Society* 477, pp. 298–324. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/sty473. URL: http://dx.doi.org/10.1093/mnras/sty473.

Getman, K. V. et al. (Aug. 1, 2022). "Evolution of X-ray activity in <25 Myr old pre-main sequence stars". In: *The Astrophysical Journal* 935, p. 43. ISSN: 0004-637X,1538-4357. DOI: 10.3847/1538-4357/ac7c69. URL: http://dx.doi.org/10.3847/1538-4357/ac7c69.

Gieles, M. and S. F. Portegies Zwart (Jan. 1, 2011). "The distinction between star clusters and associations". en. In: *Monthly Notices of the Royal Astronomical Society. Letters* 410, pp. L6–L7. ISSN: 1745-3925,1745-3933. DOI: 10.1111/j.1745-3933.2010.00967.x. URL: http://dx.doi.org/10.1111/j.1745-3933.2010.00967.x.

Gieles, M. and F. Renaud (July 24, 2016). "If it does not kill them, it makes them stronger: collisional evolution of star clusters with tidal shocks". en. In: *Monthly Notices of the Royal Astronomical Society. Letters* 463, pp. L103–L107. ISSN: 1745-3925,1745-3933. DOI: 10.1093/mnrasl/slw163. URL: http://dx.doi.org/10.1093/mnrasl/slw163.

Gieles, M. and A. Zocchi (Nov. 21, 2015). "A family of lowered isothermal models". In: *Monthly Notices of the Royal Astronomical Society* 454, pp. 576–592. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stv1848. URL: http://dx.doi.org/10.1093/mnras/stv1848.

Gieles, M. et al. (Oct. 11, 2010). "On the mass-radius relation of hot stellar systems: The mass-radius relation of stellar systems". en. In: *Monthly Notices of the Royal Astronomical Society. Letters* 408, pp. L16–L20. ISSN: 1745-3925,1745-3933. DOI: 10.1111/j.1745-3933.2010.00919.x. URL: http://dx.doi.org/10.1111/j.1745-3933.2010.00919.x.

Gieles, M. et al. (July 5, 2021). "A supra-massive population of stellar-mass black holes in the globular cluster Palomar 5". en. In: *Nature Astronomy* 5, pp. 957–966. ISSN: 2397-3366. DOI: 10.1038/s41550-021-01392-2. URL: http://dx.doi.org/10.1038/s41550-021-01392-2.

Giersz, M. and D. C. Heggie (May 1, 1994). "Statistics of N-body simulations - I. Equal masses before core collapse". In: *Monthly Notices of the Royal Astronomical Society* 268, pp. 257–275. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/268.1.257. eprint: astro-ph/9305008 (astro-ph). URL: http://dx.doi.org/10.1093/mnras/268.1.257.

Giersz, M. et al. (May 21, 2013). "MOCCA code for star cluster simulations – II. Comparison with N-body simulations". en. In: *Monthly Notices of the Royal Astronomical Society* 431, pp. 2184–2199. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stt307. URL: http://dx.doi.org/10.1093/mnras/stt307.

Gnedin, O. Y. and J. P. Ostriker (Jan. 1997). "Destruction of the galactic globular cluster system". en. In: *The Astrophysical Journal* 474, pp. 223–255. ISSN: 0004-637X,1538-4357. DOI: 10.1086/303441. URL: http://dx.doi.org/10.1086/303441.

Goodwin, S. P. and N. Bastian (Dec. 1, 2006). "Gas expulsion and the destruction of massive young clusters". In: *Monthly Notices of the Royal Astronomical Society* 373, pp. 752–758. ISSN: 0035-8711,1365-2966. DOI: 10.1111/j.1365-2966.2006.11078.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2006.11078.x.

Goodwin, S. P. (Dec. 2009). "The effect of the dynamical state of clusters on gas expulsion and infant mortality". en. In: *Astrophysics and Space Science* 324, pp. 259–263. ISSN: 0004-640X,1572-946X. DOI: 10.1007/s10509-009-0116-5. URL: http://dx.doi.org/10.1007/s10509-009-0116-5.

Groenewegen, M. A. T. (Oct. 2021). "The parallax zero-point offset from Gaia EDR3 data". In: *Astronomy and Astrophysics* 654, A20. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202140862. URL: http://dx.doi.org/10.1051/0004-6361/202140862.

Grudić, M. Y. et al. (July 20, 2021). "STARFORGE: Towards a comprehensive numerical model of star cluster formation and feedback". en. In: *Monthly Notices of the Royal Astronomical Society* 506, pp. 2199–2231. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stab1347. arXiv: 2101.10326 [cs.NE]. URL: http://dx.doi.org/10.1093/mnras/stab1347.

Guglielmetti, F., R. Fischer, and V. Dose (June 11, 2009). "Background-source separation in astronomical images with Bayesian probability theory - I. The

method". en. In: *Monthly Notices of the Royal Astronomical Society* 396, pp. 165–190. ISSN: 0035-8711,1365-2966. DOI: 10.1111/j.1365-2966.2009.14739.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2009.14739.x.

Gustafson, P. (May 1, 2005). "On model expansion, model contraction, identifiability and prior information: Two illustrative scenarios involving mismeasured variables". In: *Statistical Science: A Review Journal of the Institute of Mathematical Statistics* 20, pp. 111–140. ISSN: 0883-4237,2168-8745. DOI: 10.1214/088342305000000098. URL: http://dx.doi.org/10.1214/088342305000000098.

Guszejnov, D. et al. (July 14, 2022a). "Cluster assembly and the origin of mass segregation in the STARFORGE simulations". en. In: *Monthly Notices of the Royal Astronomical Society* 515, pp. 167–184. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stac1737. URL: http://dx.doi.org/10.1093/mnras/stac1737.

Guszejnov, D. et al. (Aug. 17, 2022b). "Effects of the environment and feedback physics on the initial mass function of stars in the STARFORGE simulations". en. In: *Monthly Notices of the Royal Astronomical Society* 515, pp. 4929–4952. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stac2060. URL: http://dx.doi.org/10.1093/mnras/stac2060.

Haghi, H., A. H. Zonoozi, and S. Taghavi (July 1, 2015). "Galactic orbital motions of star clusters: static versus semicosmological time-dependent Galactic potentials". en. In: *Monthly Notices of the Royal Astronomical Society* 450, pp. 2812–2821. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stv827. URL: http://dx.doi.org/10.1093/mnras/stv827.

Haisch Jr., K. E., L. E. A, and C. J. Lada (June 1, 2001). "Disk frequencies and lifetimes in young clusters". In: *The Astrophysical Journal* 553, pp. L153–L156. ISSN: 0004-637X,1538-4357. DOI: 10.1086/320685. URL: http://dx.doi.org/10.1086/320685.

Hao, C. J. et al. (Apr. 2022). "Newly detected open clusters in the Galactic disk using *Gaia* EDR3". In: *Astronomy and Astrophysics* 660, A4. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202243091. URL: http://dx.doi.org/10.1051/0004-6361/202243091.

Hartigan, J. A. (Apr. 1975). *Clustering Algorithms*. en. Probability & Mathematical Statistics S. John Wiley & Sons. ISBN: 9780471356455.

Hathaway, R. J. (June 1, 1985). "A constrained formulation of maximum-likelihood estimation for normal mixture distributions". In: *Annals of Statistics* 13, pp. 795–800. ISSN: 0090-5364,2168-8966. DOI: 10.1214/aos/1176349557. URL: http://dx.doi.org/10.1214/aos/1176349557.

He, Z. et al. (Sept. 1, 2022). "A blind all-sky search for star clusters in Gaia EDR3: 886 clusters within 1.2 kpc of the sun". In: *The Astrophysical Journal. Supplement Series* 262, p. 7. ISSN: 0067-0049,1538-4365. DOI: 10.3847/1538-4365/ac7c17. URL: http://dx.doi.org/10.3847/1538-4365/ac7c17.

Heggie, D. and P. Hut (June 5, 2012). *The gravitational million-body problem: A multidisciplinary approach to star cluster dynamics*. en. Cambridge University Press. ISBN: 9781139164535,9780521774864. DOI: 10.1017/cbo9781139164535. URL: http://dx.doi.org/10.1017/cbo9781139164535.

Helmi, A. (Aug. 18, 2020). "Streams, substructures, and the early history of the Milky Way". en. In: *Annual Review of Astronomy and Astrophysics* 58, pp. 205–256. ISSN: 0066-4146,1545-4282. DOI: 10.1146/annurev-astro-032620-021917. URL: http://dx.doi.org/10.1146/annurev-astro-032620-021917.

Hénault-Brunet, V. et al. (Feb. 11, 2019). "Mass modelling globular clusters in theGaiaera: a method comparison using mock data from anN-body simulation of M 4". en. In: *Monthly Notices of the Royal Astronomical Society* 483, pp. 1400–1425. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/sty3187. URL: http://dx.doi.org/10.1093/mnras/sty3187.

Hetem, A. and J. Gregorio-Hetem (Dec. 1, 2019). "Fractal statistics in young star clusters: structural parameters and dynamical evolution". en. In: *Monthly Notices of the Royal Astronomical Society* 490, pp. 2521–2541. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stz2698. URL: http://dx.doi.org/10.1093/mnras/stz2698.

Heyl, J., I. Caiazzo, and H. B. Richer (Feb. 1, 2022). "Reconstructing the Pleiades with Gaia EDR3". In: *The Astrophysical Journal* 926, p. 132. ISSN: 0004-637X,1538-4357. DOI: 10.3847/1538-4357/ac45fc. URL: http://dx.doi.org/10.3847/1538-4357/ac45fc.

Hills, J. G. (Feb. 1980). "The effect of mass loss on the dynamical evolution of a stellar system - Analytic approximations". en. In: *The Astrophysical Journal* 235, p. 986. ISSN: 0004-637X,1538-4357. DOI: 10.1086/157703. URL: http://dx.doi.org/10.1086/157703.

Hippel, T. von et al. (July 10, 2006). "Inverting color-magnitude diagrams to access precise star cluster parameters: A Bayesian approach". en. In: *The Astrophysical Journal* 645, pp. 1436–1447. ISSN: 0004-637X,1538-4357. DOI: 10.1086/504369. URL: http://dx.doi.org/10.1086/504369.

Hoerner, S. von (Mar. 1957). "Internal structure of globular clusters". en. In: *The Astrophysical Journal* 125, p. 451. ISSN: 0004-637X,1538-4357. DOI: 10.1086/146321. URL: http://dx.doi.org/10.1086/146321.

Hoffman, M. D. and A. Gelman (Nov. 17, 2011). "The no-U-Turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo". In: *arXiv [stat.CO]*. arXiv: 1111.4246 [stat.CO]. URL: http://arxiv.org/abs/1111.4246.

Hogg, D. W., A. D. Myers, and J. Bovy (Dec. 20, 2010). "Inferring the eccentricity distribution". In: *The Astrophysical Journal* 725, pp. 2166–2175. ISSN: 0004-637X,1538-4357. DOI: 10.1088/0004-637x/725/2/2166. URL: http://dx.doi.org/10.1088/0004-637x/725/2/2166.

Holl, B. et al. (June 2023). "*Gaia* Data Release 3". In: *Astronomy and Astrophysics* 674, A10. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202244161. URL: http://dx.doi.org/10.1051/0004-6361/202244161.

Hon, M., Y. Li, and J. Ong (Oct. 1, 2024). "Flow-based generative emulation of grids of stellar evolutionary models". In: *The Astrophysical Journal* 973, p. 154. ISSN: 0004-637X,1538-4357. DOI: 10.3847/1538-4357/ad6320. URL: http://dx.doi.org/10.3847/1538-4357/ad6320.

Hopkins, A. M. (July 26, 2018). "The Dawes review 8: Measuring the stellar initial mass function". en. In: *Publications of the Astronomical Society of Australia* 35, e039. ISSN: 1323-3580,1448-6083. DOI: 10.1017/pasa.2018.29. arXiv: 1807.09949 [astro-ph.GA]. URL: http://dx.doi.org/10.1017/pasa.2018.29.

Hosek Jr., M. W. et al. (Jan. 4, 2019). "The unusual initial mass function of the arches cluster". In: *The Astrophysical Journal* 870, p. 44. ISSN: 0004-637X,1538-4357. DOI: 10.3847/1538-4357/aaef90. arXiv: 1811.05513 [cond-mat.quant-gas]. URL: http://dx.doi.org/10.3847/1538-4357/aaef90.

Hota, S., R. de Grijs, and A. Subramaniam (Aug. 20, 2025). "UVIT study of the Magellanic clouds (U-SMAC). III. Hierarchical star formation in the Small Magellanic Cloud regulated by turbulence". In: *The Astrophysical Journal* 989, p. 216. ISSN: 0004-637X,1538-4357. DOI: 10.3847/1538-4357/adec84. URL: http://dx.doi.org/10.3847/1538-4357/adec84.

Huang, B. et al. (Mar. 1, 2025). "Stellar loci. VIII. Photometric metallicities for 100 million stars based on synthetic Gaia colors". In: *The Astrophysical Journal. Supplement Series* 277, p. 7. ISSN: 0067-0049,1538-4365. DOI: 10.3847/1538-4365/ada9e6. URL: http://dx.doi.org/10.3847/1538-4365/ada9e6.

Hunt, E. L. and S. Reffert (Dec. 8, 2020). "Improving the open cluster census. I. Comparison of clustering algorithms applied to Gaia DR2 data". In: *arXiv [astro-ph.GA]*. arXiv: 2012.04267 [astro-ph.GA]. URL: http://arxiv.org/abs/2012.04267.

— (Mar. 23, 2023). "Improving the open cluster census. II. An all-sky cluster catalogue with Gaia DR3". In: *arXiv [astro-ph.GA]*. arXiv: 2303.13424 [astro-ph.GA]. URL: http://arxiv.org/abs/2303.13424.

— (Mar. 8, 2024). "Improving the open cluster census. III. Using cluster masses, radii, and dynamics to create a cleaned open cluster catalogue". In: *arXiv [astro-ph.GA]*, arXiv:2403.05143. DOI: 10.48550/ARXIV.2403.05143. arXiv: 2403.05143 [astro-ph.GA]. URL: http://dx.doi.org/10.48550/arXiv.2403.05143.

Hunt, E. L. et al. (June 23, 2025). "The completeness of the open cluster census towards the Galactic anticentre". In: *arXiv [astro-ph.GA]*. DOI: 10.48550/ARXIV.2506.18708. arXiv: 2506.18708 [astro-ph.GA]. URL: http://arxiv.org/abs/2506.18708.

Hunt, E. L. et al. (Feb. 20, 2026). "The selection function of the *Gaia* DR3 open cluster census". In: *Astronomy and Astrophysics* 706, A341. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202557781. arXiv: 2510.18343 [astro-ph.GA]. URL: http://dx.doi.org/10.1051/0004-6361/202557781.

Jackson, R. J. et al. (Nov. 16, 2021). "The *Gaia*–ESO Survey: Membership probabilities for stars in 63 open and 7 globular clusters from 3D kinematics". en. In: *Monthly Notices of the Royal Astronomical Society* 509, pp. 1664–1680. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stab3032. URL: http://dx.doi.org/10.1093/mnras/stab3032.

Jaehnig, K., J. Bird, and K. Holley-Bockelmann (Dec. 1, 2021). "Membership lists for 431 open clusters in Gaia DR2 using Extreme Deconvolution Gaussian Mixture Models". In: *The Astrophysical Journal* 923, p. 129. ISSN: 0004-637X,1538-4357. DOI: 10.3847/1538-4357/ac1d51. URL: http://dx.doi.org/10.3847/1538-4357/ac1d51.

Jaynes, E. T. (Apr. 10, 2003). *Probability theory: The logic of science*. Ed. by G. L. Bretthorst. Cambridge University Press. ISBN: 9780521592710,9780511790423.

DOI: 10.1017/cbo9780511790423. URL: http://dx.doi.org/10.1017/cbo9780511790423.

Jeans, J. H. (Dec. 10, 1915). "On the theory of star-streaming and the structure of the universe". In: *Monthly Notices of the Royal Astronomical Society* 76, pp. 70–84. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/76.2.70. URL: http://dx.doi.org/10.1093/mnras/76.2.70.

Jeffreson, S. M. R. et al. (Aug. 2017). "The Gaia–ESO Survey: dynamical models of flattened, rotating globular clusters". en. In: *Monthly Notices of the Royal Astronomical Society* 469, pp. 4740–4762. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stx1152. URL: http://dx.doi.org/10.1093/mnras/stx1152.

Jeffries, R. D. et al. (May 23, 2023). "The *Gaia*-ESO Survey: empirical estimates of stellar ages from lithium equivalent widths (eagles)". en. In: *Monthly Notices of the Royal Astronomical Society* 523, pp. 802–824. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stad1293. URL: http://dx.doi.org/10.1093/mnras/stad1293.

Just, A. et al. (Jan. 21, 2009). "Quantitative analysis of clumps in the tidal tails of star clusters". en. In: *Monthly Notices of the Royal Astronomical Society* 392, pp. 969–981. ISSN: 0035-8711,1365-2966. DOI: 10.1111/j.1365-2966.2008.14099.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2008.14099.x.

Kalari, V. M. (Apr. 21, 2019). "Classical T-Tauri stars with VPHAS+: II: NGC 6383 in Sh 2-012". en. In: *Monthly Notices of the Royal Astronomical Society* 484, pp. 5102–5112. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stz250. URL: http://dx.doi.org/10.1093/mnras/stz250.

Kamann, S. et al. (Feb. 1, 2018). "A stellar census in globular clusters with MUSE: The contribution of rotation to cluster dynamics studied with 200 000 stars". In: *Monthly Notices of the Royal Astronomical Society* 473, pp. 5591–5616. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stx2719. URL: http://dx.doi.org/10.1093/mnras/stx2719.

Karam, J. and A. Sills (July 2022). "Modelling star cluster formation: mergers". In: *Monthly Notices of the Royal Astronomical Society* 513.4, pp. 6095–6104. DOI: 10.1093/mnras/stac1298. arXiv: 2205.03265 [astro-ph.GA].

Katz, D. et al. (June 2023). "*Gaia* Data Release 3: Properties and validation of the radial velocities". In: *Astronomy and Astrophysics* 674, A5. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202244220. URL: http://dx.doi.org/10.1051/0004-6361/202244220.

Kelly, B. C. (Aug. 20, 2007). "Some aspects of measurement error in linear regression of astronomical data". en. In: *The Astrophysical Journal* 665, pp. 1489–1506. ISSN: 0004-637X,1538-4357. DOI: 10.1086/519947. URL: http://dx.doi.org/10.1086/519947.

Kennedy, M. C. and A. O'Hagan (Sept. 1, 2001). "Bayesian calibration of computer models". en. In: *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 63, pp. 425–464. ISSN: 1369-7412,1467-9868. DOI: 10.1111/1467-9868.00294. URL: http://dx.doi.org/10.1111/1467-9868.00294.

Kerr, R., A. L. Kraus, and A. C. Rizzuto (Sept. 1, 2023). "SPYGLASS. IV. New stellar survey of recent star formation within 1 kpc". In: *The Astrophysical Journal* 954, p. 134. ISSN: 0004-637X,1538-4357. DOI: 10.3847/1538-4357/ace5b3. URL: http://dx.doi.org/10.3847/1538-4357/ace5b3.

Kharchenko, N. V. et al. (Aug. 2005). "Astrophysical parameters of Galactic open clusters". In: *Astronomy and Astrophysics* 438, pp. 1163–1173. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361:20042523. URL: http://dx.doi.org/10.1051/0004-6361:20042523.

King, I. (Oct. 1962). "The structure of star clusters. I. an empirical density law". In: *The Astronomical Journal* 67, p. 471. ISSN: 0004-6256,1538-3881. DOI: 10.1086/108756. URL: http://dx.doi.org/10.1086/108756.

King, I. R. (Feb. 1966). "The structure of star clusters. III. Some simple dvriamical models". In: *The Astronomical Journal* 71, p. 64. ISSN: 0004-6256,1538-3881. DOI: 10.1086/109857. URL: http://dx.doi.org/10.1086/109857.

Kounkel, M. and K. Covey (July 17, 2019). "Untangling the galaxy I: Local structure and star formation history of the milky way". In: *arXiv [astro-ph.GA]*. arXiv: 1907.07709 [astro-ph.GA]. URL: http://arxiv.org/abs/1907.07709.

Kounkel, M., K. Covey, and K. G. Stassun (Nov. 23, 2020). "Untangling the galaxy. II. Structure within 3 kpc". In: *The Astronomical Journal* 160, p. 279. ISSN: 0004-6256,1538-3881. DOI: 10.3847/1538-3881/abc0e6. URL: http://dx.doi.org/10.3847/1538-3881/abc0e6.

Kouwenhoven, M. B. N. and R. de Grijs (Oct. 5, 2007). "Binaries and the dynamical mass of star clusters". In: *arXiv [astro-ph]*. arXiv: 0710.1207 [astro-ph]. URL: http://arxiv.org/abs/0710.1207.

Kouwenhoven, M. B. N. et al. (Jan. 2009). "Exploring the consequences of pairing algorithms for binary stars". In: *Astronomy and Astrophysics* 493, pp. 979–1016. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361:200810234. URL: http://dx.doi.org/10.1051/0004-6361:200810234.

Krause, M. G. H. et al. (June 2020). "The physics of star cluster formation and evolution". en. In: *Space Science Reviews* 216. ISSN: 0038-6308,1572-9672. DOI: 10.1007/s11214-020-00689-4. URL: http://dx.doi.org/10.1007/s11214-020-00689-4.

Krone-Martins, A. and A. Moitinho (Jan. 2014). "UPMASK: unsupervised photometric membership assignment in stellar clusters". In: *Astronomy and Astrophysics* 561, A57. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/201321143. URL: http://dx.doi.org/10.1051/0004-6361/201321143.

Kroupa, P. (Apr. 1, 2001). "On the variation of the initial mass function". In: *Monthly Notices of the Royal Astronomical Society* 322, pp. 231–246. ISSN:

0035-8711,1365-2966. DOI: 10.1046/j.1365-8711.2001.04022.x. URL: http://dx.doi.org/10.1046/j.1365-8711.2001.04022.x.

Kruijssen, J. M. D. et al. (June 21, 2011). "Modelling the formation and evolution of star cluster populations in galaxy simulations: Star cluster populations in galaxy simulations". en. In: *Monthly Notices of the Royal Astronomical Society* 414, pp. 1339–1364. ISSN: 0035-8711,1365-2966. DOI: 10.1111/j.1365-2966.2011.18467.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2011.18467.x.

Krumholz, M. R., C. F. McKee, and J. Bland-Hawthorn (Aug. 18, 2019). "Star clusters across cosmic time". en. In: *Annual Review of Astronomy and Astrophysics* 57, pp. 227–303. ISSN: 0066-4146,1545-4282. DOI: 10.1146/annurev-astro-091918-104430. arXiv: 1812.01615 [astro-ph.GA]. URL: http://dx.doi.org/10.1146/annurev-astro-091918-104430.

Kuhn, M. A. et al. (Jan. 2, 2019). "Kinematics in young star clusters and associations withGaiaDR2". In: *The Astrophysical Journal* 870, p. 32. ISSN: 0004-637X,1538-4357. DOI: 10.3847/1538-4357/aaef8c. arXiv: 1810.08631 [hep-ph]. URL: http://dx.doi.org/10.3847/1538-4357/aaef8c.

Küpper, A. H. W. et al. (July 5, 2010). "Peculiarities in velocity dispersion and surface density profiles of star clusters: Peculiarities in VDP and SDP of star clusters". en. In: *Monthly Notices of the Royal Astronomical Society* 407, pp. 2241–2260. ISSN: 0035-8711,1365-2966. DOI: 10.1111/j.1365-2966.2010.17084.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2010.17084.x.

Lee, Y.-N. and P. Hennebelle (July 2016). "Formation of a protocluster: A virialized structure from gravoturbulent collapse: II. A two-dimensional analytical model for a rotating and accreting system". In: *Astronomy and Astrophysics* 591, A31. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/201527982. URL: http://dx.doi.org/10.1051/0004-6361/201527982.

Li, L., Z. Li, and Z. Shao (Oct. 27, 2025). "Validating open cluster candidates with photometric Bayesian evidence". In: *arXiv [astro-ph.GA]*. DOI: 10.48550/arXiv.2510.23375. arXiv: 2510.23375 [astro-ph.GA]. URL: http://dx.doi.org/10.48550/arXiv.2510.23375.

Lim, B. et al. (Jan. 15, 2013). "The starburst cluster Westerlund 1: The initial mass function and mass segregation". In: *The Astronomical Journal* 145, p. 46. ISSN: 0004-6256,1538-3881. DOI: 10.1088/0004-6256/145/2/46. arXiv: 1211.5832 [astro-ph.SR]. URL: http://dx.doi.org/10.1088/0004-6256/145/2/46.

Lindegren, L. et al. (Aug. 2018). "Gaia Data Release 2: The astrometric solution". In: *Astronomy and Astrophysics* 616, A2. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/201832727. URL: http://dx.doi.org/10.1051/0004-6361/201832727.

Lindegren, L. et al. (May 2021a). "*Gaia*Early Data Release 3: Parallax bias versus magnitude, colour, and position". In: *Astronomy and Astrophysics* 649, A4. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202039653. URL: http://dx.doi.org/10.1051/0004-6361/202039653.

Lindegren, L. et al. (May 2021b). "*Gaia*Early Data Release 3: The astrometric solution". In: *Astronomy and Astrophysics* 649, A2. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202039709. URL: http://dx.doi.org/10.1051/0004-6361/202039709.

Linder, N. et al. (Oct. 2007). "The Struve-Sahade effect in the optical spectra of O-type binaries". In: *Astronomy and Astrophysics* 474, pp. 193–204. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361:20077902. URL: http://dx.doi.org/10.1051/0004-6361:20077902.

Lindoff, U. (Jan. 1968). "The ages of open clusters". In: *Arkiv for Astronomi* 5, pp. 1–21.

Livernois, A. et al. (Aug. 12, 2021). "Early dynamics and violent relaxation of multimass rotating star clusters". en. In: *Monthly Notices of the Royal Astro-*

nomical Society 506, pp. 5781–5801. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stab2119. URL: http://dx.doi.org/10.1093/mnras/stab2119.

Lloyd Evans, T. (Oct. 1, 1978). "The open cluster NGC 6383". In: *Monthly Notices of the Royal Astronomical Society* 184, pp. 661–676. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/184.4.661. URL: http://dx.doi.org/10.1093/mnras/184.4.661.

Loredo, T. J. (2004). "Accounting for source uncertainties in analyses of astronomical survey data". In: *AIP Conference Proceedings*. AIP. DOI: 10.1063/1.1835214. URL: http://dx.doi.org/10.1063/1.1835214.

Loredo, T. J. and R. L. Wolpert (June 27, 2024). "Bayesian inference: More than Bayes's theorem". In: *arXiv [stat.ME]*. arXiv: 2406.18905 [stat.ME]. URL: http://arxiv.org/abs/2406.18905.

Luri, X. et al. (Aug. 2018). "*Gaia*Data Release 2: Using*Gaia*parallaxes". In: *Astronomy and Astrophysics* 616, A9. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/201832964. URL: http://dx.doi.org/10.1051/0004-6361/201832964.

Maíz Apellániz, J. (Jan. 2022). "An estimation of the*Gaia*EDR3 parallax bias from stellar clusters and Magellanic Clouds data". In: *Astronomy and Astrophysics* 657, A130. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202142365. URL: http://dx.doi.org/10.1051/0004-6361/202142365.

Malzer, C. and M. Baum (Sept. 14, 2020). "A hybrid approach to hierarchical density-based cluster selection". In: *2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. IEEE. ISBN: 9781728164229. DOI: 10.1109/mfi49285.2020.9235263. URL: http://dx.doi.org/10.1109/mfi49285.2020.9235263.

Marrese, P. M. et al. (June 2022). "Gaia DR3 documentation Chapter 15: Cross-match with external catalogues". en. In: *Gaia DR3 documentation*, p. 15. URL: https://ui.adsabs.harvard.edu/abs/2022gdr3.reptE..15M/abstract?.

Martens, S. et al. (Jan. 27, 2023). "Kinematic differences between multiple populations in Galactic globular clusters". In: *Astronomy and Astrophysics*. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202244787. URL: http://dx.doi.org/10.1051/0004-6361/202244787.

Martin, W. and D. Mortlock (Jan. 3, 2025). "An approach to robust Bayesian regression in astronomy". en. In: *RAS Techniques and Instruments* 4. ISSN: 2752-8200. DOI: 10.1093/rasti/rzaf035. URL: http://dx.doi.org/10.1093/rasti/rzaf035.

Martinez-Medina, L. A. et al. (Feb. 11, 2018). "New insights into the origin and evolution of the old, metal-rich open cluster NGC 6791". en. In: *Monthly Notices of the Royal Astronomical Society* 474, pp. 32–44. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stx2739. URL: http://dx.doi.org/10.1093/mnras/stx2739.

McBride, A. et al. (Dec. 1, 2021). "Untangling the galaxy. III. Photometric search for pre-main-sequence stars with deep learning". In: *The Astronomical Journal* 162, p. 282. ISSN: 0004-6256,1538-3881. DOI: 10.3847/1538-3881/ac2432. URL: http://dx.doi.org/10.3847/1538-3881/ac2432.

McInnes, L., J. Healy, and S. Astels (Mar. 21, 2017). "hdbscan: Hierarchical density based clustering". In: *Journal of Open Source Software* 2, p. 205. ISSN: 2475-9066. DOI: 10.21105/joss.00205. URL: http://dx.doi.org/10.21105/joss.00205.

McKee, C. F. and E. G. Zweibel (Nov. 1992). "On the virial theorem for turbulent molecular clouds". In: *The Astrophysical Journal* 399, p. 551. ISSN: 0004-637X,1538-4357. DOI: 10.1086/171946. URL: http://dx.doi.org/10.1086/171946.

McLachlan, G. J. and D. Peel (Sept. 18, 2000). *Finite Mixture Models*. en. Wiley Series in Probability and Statistics. John Wiley & Sons. ISBN:

9780471006268,9780471721185. DOI: 10.1002/0471721182. URL: http://dx.doi.org/10.1002/0471721182.

McMillan, S. L. W., E. Vesperini, and S. F. Portegies Zwart (Jan. 20, 2007). "A dynamical origin for early mass segregation in young star clusters". en. In: *The Astrophysical Journal* 655, pp. L45–L49. ISSN: 0004-637X,1538-4357. DOI: 10.1086/511763. URL: http://dx.doi.org/10.1086/511763.

Meingast, S., J. Alves, and A. Rottensteiner (Jan. 2021). "Extended stellar systems in the solar neighborhood: V. Discovery of coronae of nearby star clusters". In: *Astronomy and Astrophysics* 645, A84. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202038610. URL: http://dx.doi.org/10.1051/0004-6361/202038610.

Menon, S. H. et al. (Sept. 17, 2021). "The dependence of the hierarchical distribution of star clusters on galactic environment". en. In: *Monthly Notices of the Royal Astronomical Society* 507, pp. 5542–5566. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stab2413. arXiv: 2108.04387 [astro-ph.GA]. URL: http://dx.doi.org/10.1093/mnras/stab2413.

Meyer, M. R., N. Calvet, and L. A. Hillenbrand (July 1997). "Intrinsic near-infrared excesses of T Tauri stars: Understanding the classical T Tauri star locus". In: *The Astronomical Journal* 114, p. 288. ISSN: 0004-6256,1538-3881. DOI: 10.1086/118474. URL: http://dx.doi.org/10.1086/118474.

Minniti, D. et al. (July 2010). "VISTA Variables in the Via Lactea (VVV): The public ESO near-IR variability survey of the Milky Way". en. In: *New Astronomy* 15, pp. 433–443. ISSN: 1384-1076,1384-1092. DOI: 10.1016/j.newast.2009.12.002. URL: http://dx.doi.org/10.1016/j.newast.2009.12.002.

Moeckel, N. and I. A. Bonnell (Dec. 1, 2009). "Does subcluster merging accelerate mass segregation in local clusters?" en. In: *Monthly Notices of the Royal Astronomical Society* 400, pp. 657–664. ISSN: 0035-8711,1365-2966. DOI: 10.

1111/j.1365-2966.2009.15499.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2009.15499.x.

Monteiro, H. et al. (Sept. 28, 2020). "Fundamental parameters for 45 open clusters with Gaia DR2, an improved extinction correction and a metallicity gradient prior". en. In: *Monthly Notices of the Royal Astronomical Society*. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/staa2983. URL: http://dx.doi.org/10.1093/mnras/staa2983.

Moran, G. E., J. P. Cunningham, and D. M. Blei (Jan. 1, 2022). "The posterior predictive null". In: *Bayesian Analysis* -1. ISSN: 1936-0975,1931-6690. DOI: 10.1214/22-ba1313. URL: http://dx.doi.org/10.1214/22-ba1313.

Morris, C. N. and M. Lysy (Feb. 1, 2012). "Shrinkage estimation in multilevel normal models". In: *Statistical Science: A Review Journal of the Institute of Mathematical Statistics* 27, pp. 115–134. ISSN: 0883-4237,2168-8745. DOI: 10.1214/11-sts363. URL: http://dx.doi.org/10.1214/11-sts363.

Motherway, E. et al. (Feb. 1, 2024). "Tracing the origins of mass segregation in M35: Evidence for primordially segregated binaries". In: *The Astrophysical Journal. Letters* 962, p. L9. ISSN: 2041-8205,2041-8213. DOI: 10.3847/2041-8213/ad18bf. URL: http://dx.doi.org/10.3847/2041-8213/ad18bf.

Moya, A. et al. (Jan. 13, 2022). "Stellar dating using chemical clocks and Bayesian inference". In: *arXiv [astro-ph.SR]*. DOI: 10.48550/ARXIV.2201.05228. arXiv: 2201.05228 [astro-ph.SR]. URL: http://arxiv.org/abs/2201.05228.

Murphy, S. J. et al. (Oct. 22, 2020). "The discovery of lambda Bootis stars - the Southern Survey II". en. In: *Monthly Notices of the Royal Astronomical Society* 499, pp. 2701–2713. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/staa2347. URL: http://dx.doi.org/10.1093/mnras/staa2347.

Naylor, T. and R. D. Jeffries (Dec. 2006). "A maximum-likelihood method for fitting colour–magnitude diagrams". en. In: *Monthly Notices of the Royal Astronomical Society* 373, pp. 1251–1263. ISSN: 0035-8711,1365-2966. DOI: 10.

1111/j.1365-2966.2006.11099.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2006.11099.x.

Neal, R. M. (June 8, 2012). "MCMC using Hamiltonian dynamics". In: *arXiv [stat.CO]*. arXiv: 1206.1901 [stat.CO]. URL: http://arxiv.org/abs/1206.1901.

Nizovkina, M. et al. (Oct. 12, 2025). "Refining open cluster parameters with Gaia XP metallicities". In: *arXiv [astro-ph.GA]*. DOI: 10.48550/arXiv.2510.10385. arXiv: 2510.10385 [astro-ph.GA]. URL: http://dx.doi.org/10.48550/arXiv.2510.10385.

Noormohammadi, M., M. Khakian Ghomi, and A. Javadi (June 26, 2024). "Detection of open cluster members inside and beyond tidal radius by machine learning methods based on *Gaia* DR3". en. In: *Monthly Notices of the Royal Astronomical Society* 532, pp. 622–642. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stae1448. URL: http://dx.doi.org/10.1093/mnras/stae1448.

O'Donnell, J. E. (Feb. 1994). "R[SUB]nu[/SUB]-dependent optical and near-ultraviolet extinction". en. In: *The Astrophysical Journal* 422, p. 158. ISSN: 0004-637X,1538-4357. DOI: 10.1086/173713. URL: http://dx.doi.org/10.1086/173713.

Offner, S. S. R. et al. (2022). "The origin and evolution of multiple star systems". In: *arXiv [astro-ph.SR]*. DOI: 10.48550/ARXIV.2203.10066. eprint: 2203.10066 (astro-ph.SR). URL: http://dx.doi.org/10.48550/ARXIV.2203.10066.

Olivares, J. et al. (Dec. 2020). "Kalkayotl: A cluster distance inference code". In: *Astronomy and Astrophysics* 644, A7. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202037846. URL: http://dx.doi.org/10.1051/0004-6361/202037846.

Padoan, P. et al. (Sept. 3, 2020). "The origin of massive stars: The inertial-inflow model". In: *The Astrophysical Journal* 900, p. 82. ISSN: 0004-637X,1538-4357.

DOI: 10.3847/1538-4357/abaa47. URL: http://dx.doi.org/10.3847/1538-4357/abaa47.

Pancino, E. et al. (June 2024). "Differential reddening in 48 globular clusters: An end to the quest for the intracluster medium". In: *Astronomy and Astrophysics* 686, A283. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202449462. URL: http://dx.doi.org/10.1051/0004-6361/202449462.

Pandey, A. K. et al. (Jan. 1, 1989). "Integrated parameters of open clusters". In: *Monthly Notices of the Royal Astronomical Society* 236, pp. 263–276. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/236.2.263. URL: http://dx.doi.org/10.1093/mnras/236.2.263.

Pang, X. et al. (Aug. 6, 2020). "Different fates of young star clusters after gas expulsion". In: *arXiv [astro-ph.GA]*. arXiv: 2008.02803 [astro-ph.GA]. URL: http://arxiv.org/abs/2008.02803.

Pang, X. et al. (Sept. 1, 2023). "Binary star evolution in different environments: Filamentary, fractal, halo, and tidal tail clusters". In: *The Astronomical Journal* 166, p. 110. ISSN: 0004-6256,1538-3881. DOI: 10.3847/1538-3881/ace76c. URL: http://dx.doi.org/10.3847/1538-3881/ace76c.

Parker, R. J. and S. P. Goodwin (June 1, 2015). "Comparisons between different techniques for measuring mass segregation". In: *Monthly Notices of the Royal Astronomical Society* 449, pp. 3381–3392. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stv539. URL: http://dx.doi.org/10.1093/mnras/stv539.

Parker, R. J. et al. (Feb. 11, 2014). "Dynamical evolution of star-forming regions". In: *Monthly Notices of the Royal Astronomical Society* 438, pp. 620–638. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stt2231. URL: http://dx.doi.org/10.1093/mnras/stt2231.

Paunzen, E., M. Netopil, and K. Zwintz (Jan. 2007). "Investigating star formation in the young open cluster NGC 6383". In: *Astronomy and Astrophysics* 462,

pp. 157–162. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361:20065513. URL: http://dx.doi.org/10.1051/0004-6361:20065513.

Pedregosa, F. et al. (Jan. 2, 2012). "Scikit-learn: Machine Learning in Python". In: *arXiv [cs.LG]*, pp. 2825–2830. arXiv: 1201.0490 [cs.LG]. URL: http://arxiv.org/abs/1201.0490.

Penny, L. R., J. G. Epps, and J. D. Snyder (Dec. 1, 2016). "Tomographic separation of composite spectra. Xii. The physical properties and spectral phase variability of the massive close binary hd 159176". In: *The Astrophysical Journal* 832, p. 211. ISSN: 0004-637X,1538-4357. DOI: 10.3847/0004-637x/832/2/211. URL: http://dx.doi.org/10.3847/0004-637X/832/2/211.

Penoyre, Z., V. Belokurov, and N. W. Evans (May 5, 2022a). "Astrometric identification of nearby binary stars – I. Predicted astrometric signals". en. In: *Monthly Notices of the Royal Astronomical Society* 513, pp. 2437–2456. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stac959. URL: http://dx.doi.org/10.1093/mnras/stac959.

— (2022b). "Astrometric identification of nearby binary stars II: Astrometric binaries in the Gaia Catalogue of Nearby Stars". In: *arXiv [astro-ph.SR]*. DOI: 10.48550/ARXIV.2202.06963. eprint: 2202.06963 (astro-ph.SR). URL: http://dx.doi.org/10.48550/ARXIV.2202.06963.

Penoyre, Z. et al. (June 11, 2020). "Binary deviations from single object astrometry". en. In: *Monthly Notices of the Royal Astronomical Society* 495, pp. 321–337. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/staa1148. URL: http://dx.doi.org/10.1093/mnras/staa1148.

Perren, G. I., R. A. Vázquez, and A. E. Piatti (Apr. 2015). "ASteCA: Automated Stellar Cluster Analysis". In: *Astronomy and Astrophysics* 576, A6. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/201424946. URL: http://dx.doi.org/10.1051/0004-6361/201424946.

Peterson, C. J. and I. R. King (June 1975). "The structure of stars. VI. Observed radii and structural parameters in globular clusters". In: *The Astronomical Journal* 80, p. 427. ISSN: 0004-6256,1538-3881. DOI: 10.1086/111759. URL: http://dx.doi.org/10.1086/111759.

Pfalzner, S., S. Dehghani, and A. Michel (Nov. 1, 2022). "Most planets might have more than 5 Myr of time to form". In: *The Astrophysical Journal. Letters* 939, p. L10. ISSN: 2041-8205,2041-8213. DOI: 10.3847/2041-8213/ac9839. URL: http://dx.doi.org/10.3847/2041-8213/ac9839.

Pfalzner, S. and F. Dincer (Mar. 1, 2024). "Low-mass stars: Their protoplanetary disk lifetime distribution". In: *The Astrophysical Journal* 963, p. 122. ISSN: 0004-637X,1538-4357. DOI: 10.3847/1538-4357/ad1bef. URL: http://dx.doi.org/10.3847/1538-4357/ad1bef.

Pfeffer, J. et al. (Apr. 21, 2018). "The E-MOSAICS project: simulating the formation and co-evolution of galaxies and their star cluster populations". en. In: *Monthly Notices of the Royal Astronomical Society* 475, pp. 4309–4346. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stx3124. URL: http://dx.doi.org/10.1093/mnras/stx3124.

Pianta, C., R. Capuzzo-Dolcetta, and G. Carraro (Nov. 1, 2022). "The impact of binaries on the dynamical mass estimate of dwarf galaxies". In: *The Astrophysical Journal* 939, p. 3. ISSN: 0004-637X,1538-4357. DOI: 10.3847/1538-4357/ac9303. URL: http://dx.doi.org/10.3847/1538-4357/ac9303.

Piironen, J. and A. Vehtari (Jan. 1, 2017). "Sparsity information and regularization in the horseshoe and other shrinkage priors". In: *Electronic Journal of Statistics* 11, pp. 5018–5051. ISSN: 1935-7524. DOI: 10.1214/17-ejs1337si. URL: http://dx.doi.org/10.1214/17-ejs1337si.

Piskunov, A. E. et al. (Jan. 2008). "Tidal radii and masses of open clusters". In: *Astronomy and Astrophysics* 477, pp. 165–172. ISSN: 0004-6361,1432-0746. DOI:

10.1051/0004-6361:20078525. URL: http://dx.doi.org/10.1051/0004-6361:20078525.

Plummer, H. C. (Mar. 10, 1911). "On the problem of distribution in globular star clusters: (plate 8.)" In: *Monthly Notices of the Royal Astronomical Society* 71, pp. 460–470. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/71.5.460. URL: http://dx.doi.org/10.1093/mnras/71.5.460.

Polak, B. et al. (Sept. 16, 2025). "Massive star cluster formation III. Early mass segregation during cluster assembly". In: *arXiv [astro-ph.GA]*. arXiv: 2408.14592 [astro-ph.GA]. URL: http://arxiv.org/abs/2408.14592.

Polson, N. G. and J. G. Scott (Dec. 1, 2012). "On the half-Cauchy prior for a global scale parameter". In: *Bayesian Analysis* 7, pp. 887–902. ISSN: 1936-0975,1931-6690. DOI: 10.1214/12-ba730. URL: http://dx.doi.org/10.1214/12-ba730.

Portegies Zwart, S. F., S. L. W. McMillan, and M. Gieles (Aug. 1, 2010). "Young massive star clusters". en. In: *Annual Review of Astronomy and Astrophysics* 48, pp. 431–493. ISSN: 0066-4146,1545-4282. DOI: 10.1146/annurev-astro-081309-130834. URL: http://dx.doi.org/10.1146/annurev-astro-081309-130834.

Pulgar-Escobar, L. M. et al. (May 15, 2024). "Characterizing NGC 6383: A study of pre-main sequence stars, mass segregation, and age using Gaia DR3 and 2MASS". In: *arXiv [astro-ph.SR]*. arXiv: 2405.09145 [astro-ph.SR]. URL: http://arxiv.org/abs/2405.09145.

Queiroz, A. B. A. et al. (May 2023). "StarHorse results for spectroscopic surveys and *Gaia* DR3: Chrono-chemical populations in the solar vicinity, the genuine thick disk, and young alpha-rich stars". In: *Astronomy and Astrophysics* 673, A155. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202245399. URL: http://dx.doi.org/10.1051/0004-6361/202245399.

Randich, S. et al. (Apr. 2018). "The*Gaia*-ESO Survey: open clusters in*Gaia*-DR1". In: *Astronomy and Astrophysics* 612, A99. ISSN: 0004-6361,1432-0746. DOI:

10.1051/0004-6361/201731738. URL: http://dx.doi.org/10.1051/0004-6361/201731738.

Rauw, G., J. Manfroid, and M. De Becker (Feb. 2010). "A photometric and spectroscopic investigation of star formation in the very young open cluster NGC 6383". In: *Astronomy and Astrophysics* 511, A25. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/200912780. URL: http://dx.doi.org/10.1051/0004-6361/200912780.

Rauw, G. et al. (Sept. 2003). "AnXMM-Newtonobservation of the very young open cluster NGC 6383". In: *Astronomy and Astrophysics* 407, pp. 925–934. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361:20030910. URL: http://dx.doi.org/10.1051/0004-6361:20030910.

Rauw, G. and M. De Becker (2008). "The multiwavelength picture of star formation in the very young open cluster NGC6383". In: *arXiv [astro-ph]*. Ed. by B. Reipurth, p. 497. DOI: 10.48550/ARXIV.0808.3887. eprint: 0808.3887 (astro-ph). URL: http://dx.doi.org/10.48550/arXiv.0808.3887.

Read, J. I. and P. Steger (Nov. 11, 2017). "How to break the density-anisotropy degeneracy in spherical stellar systems". In: *Monthly Notices of the Royal Astronomical Society* 471, pp. 4541–4558. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stx1798. URL: http://dx.doi.org/10.1093/mnras/stx1798.

Read, J. I. et al. (Dec. 24, 2020). "Breaking beta: a comparison of mass modelling methods for spherical systems". en. In: *Monthly Notices of the Royal Astronomical Society* 501, pp. 978–993. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/staa3663. URL: http://dx.doi.org/10.1093/mnras/staa3663.

Renaud, F., M. Gieles, and C. M. Boily (Dec. 1, 2011). "Evolution of star clusters in arbitrary tidal fields: Star clusters in arbitrary tidal fields". en. In: *Monthly Notices of the Royal Astronomical Society* 418, pp. 759–769. ISSN: 0035-8711,1365-2966. DOI: 10.1111/j.1365-2966.2011.19531.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2011.19531.x.

Ribas, Á., H. Bouy, and B. Merín (Apr. 2015). "Protoplanetary disk lifetimes vs. stellar mass and possible implications for giant planet populations". In: *Astronomy and Astrophysics* 576, A52. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/201424846. URL: http://dx.doi.org/10.1051/0004-6361/201424846.

Riello, M. et al. (May 2021). "*Gaia*Early Data Release 3: Photometric content and validation". In: *Astronomy and Astrophysics* 649, A3. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202039587. URL: http://dx.doi.org/10.1051/0004-6361/202039587.

Risbud, D., V. V. Jadhav, and P. Kroupa (Feb. 2025). "Tidal tails of nearby open clusters: I. Mapping with *Gaia* DR3". In: *Astronomy and Astrophysics* 694, A258. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202453302. URL: http://dx.doi.org/10.1051/0004-6361/202453302.

Rix, H.-W. et al. (June 14, 2021). "Selection functions in astronomical data modeling, with the space density of white dwarfs as worked example". In: *arXiv [astro-ph.IM]*. arXiv: 2106.07653 [astro-ph.IM]. URL: http://arxiv.org/abs/2106.07653.

Robert, C. and G. Casella (July 26, 2005). *Monte Carlo Statistical Methods*. en. 2nd ed. Springer Texts in Statistics. Springer. ISBN: 9780387212395.

Robotham, A. S. G. and D. Obreschkow (2015). "Hyper-fit: Fitting linear models to multidimensional data with multivariate Gaussian uncertainties". en. In: *Publications of the Astronomical Society of Australia* 32. ISSN: 1323-3580,1448-6083. DOI: 10.1017/pasa.2015.33. URL: http://dx.doi.org/10.1017/pasa.2015.33.

Rodríguez-Casal, A. and P. Saavedra-Nieves (June 1, 2022). "A data-adaptive method for estimating density level sets under shape conditions". In: *Annals of Statistics* 50. ISSN: 0090-5364,2168-8966. DOI: 10.1214/21-aos2168. URL: http://dx.doi.org/10.1214/21-aos2168.

Rybizki, J. et al. (Jan. 5, 2022). "A classifier for spurious astrometric solutions in Gaia eDR3". en. In: *Monthly Notices of the Royal Astronomical Society* 510, pp. 2597–2616. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stab3588. URL: http://dx.doi.org/10.1093/mnras/stab3588.

Sabatini, G. et al. (Sept. 1, 2022). "The ALMA Survey of 70 μm Dark High-mass clumps in Early Stages (ASHES). VI. The core-scale CO depletion". In: *The Astrophysical Journal* 936, p. 80. ISSN: 0004-637X,1538-4357. DOI: 10.3847/1538-4357/ac83aa. URL: http://dx.doi.org/10.3847/1538-4357/ac83aa.

Sabbi, E. et al. (Jan. 1, 2008). "The stellar mass distribution in the giant star forming region ngc 346". In: *The Astronomical Journal* 135, pp. 173–181. ISSN: 0004-6256,1538-3881. DOI: 10.1088/0004-6256/135/1/173. URL: http://dx.doi.org/10.1088/0004-6256/135/1/173.

Sanders, W. L. (Sept. 1971). "An improved method for computing membership probabilities in open clusters". In: *å* 14, pp. 226–232.

Sandoval-Garrido, N. A. et al. (Apr. 2025). "ALMA-IMF: XVIII. The assembly of a star cluster: Dense N_2H^+ (1-0) kinematics in the massive G351.77 protocluster". In: *Astronomy and Astrophysics* 696, A202. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202452589. URL: http://dx.doi.org/10.1051/0004-6361/202452589.

Sanford, R. F. (Sept. 1949). "Interstellar calcium lines in the spectra of stars in open clusters". en. In: *The Astrophysical Journal* 110, p. 117. ISSN: 0004-637X,1538-4357. DOI: 10.1086/145191. URL: http://dx.doi.org/10.1086/145191.

Sanhueza, P. et al. (Nov. 25, 2019). "The ALMA survey of 70 μm dark high-mass clumps in early stages (ASHES). I. pilot survey: Clump fragmentation". In: *The Astrophysical Journal* 886, p. 102. ISSN: 0004-637X,1538-4357. DOI: 10.3847/1538-4357/ab45e9. arXiv: 1910.06338 [astro-ph.EP]. URL: http://dx.doi.org/10.3847/1538-4357/ab45e9.

Schad, D. J., M. Betancourt, and S. Vasishth (Feb. 2021). "Toward a principled Bayesian workflow in cognitive science". en. In: *Psychological Methods* 26, pp. 103–126. ISSN: 1082-989X,1939-1463. DOI: 10.1037/met0000275. URL: http://dx.doi.org/10.1037/met0000275.

Schneider, F. R. N. et al. (Dec. 16, 2013). "Ages of young star clusters, massive blue stragglers, and the upper mass limit of stars: Analyzing age-dependent stellar mass functions". In: *The Astrophysical Journal* 780, p. 117. ISSN: 0004-637X,1538-4357. DOI: 10.1088/0004-637x/780/2/117. URL: http://dx.doi.org/10.1088/0004-637x/780/2/117.

Scott, D. W. (1979). "On optimal and data-based histograms". In: *Biometrika* 66, pp. 605–610. ISSN: 0006-3444,1464-3510. DOI: 10.1093/biomet/66.3.605. URL: http://dx.doi.org/10.1093/biomet/66.3.605.

Shanahan, R. L. and M. Gieles (Mar. 21, 2015). "Biases in the inferred mass-to-light ratio of globular clusters: no need for variations in the stellar mass function". In: *Monthly Notices of the Royal Astronomical Society. Letters* 448, pp. L94–L98. ISSN: 1745-3925,1745-3933. DOI: 10.1093/mnrasl/slu205. URL: http://dx.doi.org/10.1093/mnrasl/slu205.

Shukirgaliyev, B. et al. (June 11, 2019). "The star cluster survivability after gas expulsion is independent of the impact of the Galactic tidal field". en. In: *Monthly Notices of the Royal Astronomical Society* 486, pp. 1045–1052. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stz876. arXiv: 1903.09560 [astro-ph.GA]. URL: http://dx.doi.org/10.1093/mnras/stz876.

Siess, L., E. Dufour, and M. Forestini (2000). "An internet server for update pre-main sequence tracks of low- and intermediate-mass stars". In: *arXiv [astro-ph]*. DOI: 10.48550/ARXIV.ASTRO-PH/0003477. eprint: astro-ph/0003477 (astro-ph). URL: http://dx.doi.org/10.48550/ARXIV.ASTRO-PH/0003477.

Simpson, D. et al. (Feb. 1, 2017). "Penalising model component complexity: A principled, practical approach to constructing priors". In: *Statistical Science:*

A Review Journal of the Institute of Mathematical Statistics 32, pp. 1–28. ISSN: 0883-4237,2168-8745. DOI: 10.1214/16-sts576. URL: http://dx.doi.org/10.1214/16-sts576.

Sisson, S. A., Y. Fan, and M. Beaumont, eds. (Dec. 18, 2020). *Handbook of approximate Bayesian computation*. en. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Chapman & Hall/CRC. ISBN: 9781439881507.

Skrutskie, M. F. et al. (Feb. 2006). "The two micron all sky survey (2MASS)". en. In: *The Astronomical Journal* 131, pp. 1163–1183. ISSN: 0004-6256,1538-3881. DOI: 10.1086/498708. URL: http://dx.doi.org/10.1086/498708.

Smith, L. C. et al. (Dec. 20, 2024). "VIRAC2: NIR astrometry and time series photometry for 500M+ stars from the VVV and VVVX surveys". en. In: *Monthly Notices of the Royal Astronomical Society*. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stae2797. URL: http://dx.doi.org/10.1093/mnras/stae2797.

Smith, R. et al. (July 11, 2011). "Surviving infant mortality in the hierarchical merging scenario: Surviving infant mortality". en. In: *Monthly Notices of the Royal Astronomical Society* 414, pp. 3036–3043. ISSN: 0035-8711,1365-2966. DOI: 10.1111/j.1365-2966.2011.18604.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2011.18604.x.

Sollima, A. (June 21, 2020). "The eye of Gaia on globular clusters structure: tidal tails". en. In: *Monthly Notices of the Royal Astronomical Society* 495, pp. 2222–2233. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/staa1209. URL: http://dx.doi.org/10.1093/mnras/staa1209.

Sollima, A., H. Baumgardt, and M. Hilker (May 1, 2019). "The eye of Gaia on globular clusters kinematics: internal rotation". en. In: *Monthly Notices of the Royal Astronomical Society* 485, pp. 1460–1476. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stz505. URL: http://dx.doi.org/10.1093/mnras/stz505.

Sollima, A. et al. (Aug. 1, 2015). "Biases in the determination of dynamical parameters of star clusters: today and in the Gaia era". In: *Monthly Notices of the Royal Astronomical Society* 451, pp. 2185–2197. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stv1079. URL: http://dx.doi.org/10.1093/mnras/stv1079.

Sollima, A. et al. (July 29, 2023). "Testing quasi-linear modified Newtonian dynamics theory with Galactic globular clusters in a weak external field". en. In: *Monthly Notices of the Royal Astronomical Society* 524, pp. 5291–5308. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stad2186. URL: http://dx.doi.org/10.1093/mnras/stad2186.

Somers, G., L. Cao, and M. H. Pinsonneault (Mar. 1, 2020). "The SPOTS models: A grid of theoretical stellar evolution tracks and isochrones for testing the effects of starspots on structure and colors". In: *The Astrophysical Journal* 891, p. 29. ISSN: 0004-637X,1538-4357. DOI: 10.3847/1538-4357/ab722e. URL: http://dx.doi.org/10.3847/1538-4357/ab722e.

Spera, M., M. Mapelli, and R. D. Jeffries (July 21, 2016). "Do open star clusters evolve towards energy equipartition?" In: *Monthly Notices of the Royal Astronomical Society* 460, pp. 317–328. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stw998. URL: http://dx.doi.org/10.1093/mnras/stw998.

Spitzer, L. S. (July 14, 2014). *Dynamical evolution of globular clusters*. en. Princeton Legacy Library. Princeton University Press. ISBN: 9781400858736. DOI: 10.1515/9781400858736. URL: http://dx.doi.org/10.1515/9781400858736.

Spitzer Jr., L. (Dec. 1969). "Equipartition and the formation of compact nuclei in spherical stellar systems". en. In: *The Astrophysical Journal* 158, p. L139. ISSN: 0004-637X,1538-4357. DOI: 10.1086/180451. URL: http://dx.doi.org/10.1086/180451.

Stein, M. L. (Dec. 6, 2012). *Interpolation of spatial data*. en. 1999th ed. Springer Series in Statistics. Springer. ISBN: 9781461214946.

Stenning, D. C. et al. (July 18, 2016). "Bayesian analysis of two stellar populations in galactic globular clusters. I. Statistical and computational methods". In: *The Astrophysical Journal* 826, p. 41. ISSN: 0004-637X,1538-4357. DOI: 10.3847/0004-637x/826/1/41. URL: http://dx.doi.org/10.3847/0004-637x/826/1/41.

Stepanishchev, A. S. and V. V. Bobylev (Mar. 2013). "Corrections for the Lutz-Kelker bias for Galactic masers". en. In: *Astronomy Letters* 39, pp. 185–191. ISSN: 1063-7737,1562-6873. DOI: 10.1134/s1063773713030067. URL: http://dx.doi.org/10.1134/s1063773713030067.

Suppes, P. (1969). "Models of Data". In: *Studies in the Methodology and Foundations of Science*. Springer Netherlands, pp. 24–35. ISBN: 9789048183203. DOI: 10.1007/978-94-017-3173-7_2. URL: http://dx.doi.org/10.1007/978-94-017-3173-7_2.

Tarricq, Y. et al. (Jan. 14, 2021). "3D kinematics and age distribution of the open cluster population". In: *Astronomy and Astrophysics*. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202039388. URL: http://dx.doi.org/10.1051/0004-6361/202039388.

Tarricq, Y. et al. (Mar. 2022). "Structural parameters of 389 local open clusters". In: *Astronomy and Astrophysics* 659, A59. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/202142186. URL: http://dx.doi.org/10.1051/0004-6361/202142186.

The, P. S. et al. (Oct. 1985). "The spectral energy distribution of stars above the ZAMS in the central part of the open cluster NGC 6383". In: *åp* 151, pp. 391–398.

The Astropy Collaboration et al. (Oct. 2013). "Astropy: A community Python package for astronomy". In: *Astronomy and Astrophysics* 558, A33. ISSN: 0004-6361,1432-0746. DOI: 10.1051/0004-6361/201322068. URL: http://dx.doi.org/10.1051/0004-6361/201322068.

The Astropy Collaboration et al. (Aug. 23, 2018). "The astropy project: Building an open-science project and status of the v2.0 core package". In: *The Astronomical Journal* 156, p. 123. ISSN: 0004-6256,1538-3881. DOI: 10.3847/1538-3881/aabc4f. URL: http://dx.doi.org/10.3847/1538-3881/aabc4f.

The Astropy Collaboration et al. (Aug. 1, 2022). "The Astropy Project: Sustaining and growing a community-oriented open-source Project and the latest major release (v5.0) of the core package". In: *The Astrophysical Journal* 935, p. 167. ISSN: 0004-637X,1538-4357. DOI: 10.3847/1538-4357/ac7c74. URL: http://dx.doi.org/10.3847/1538-4357/ac7c74.

Tiongco, M. A., E. Vesperini, and A. L. Varri (Apr. 8, 2017). "Kinematical evolution of tidally limited star clusters: rotational properties". In: *Monthly Notices of the Royal Astronomical Society*. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stx853. URL: http://dx.doi.org/10.1093/mnras/stx853.

Torniamenti, S. et al. (Aug. 31, 2021). "The impact of binaries on the evolution of star clusters from turbulent molecular clouds". en. In: *Monthly Notices of the Royal Astronomical Society* 507, pp. 2253–2266. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stab2238. URL: http://dx.doi.org/10.1093/mnras/stab2238.

Trenti, M. and R. van der Marel (Feb. 8, 2013). "No energy equipartition in globular clusters". In: *arXiv [astro-ph.GA]*. arXiv: 1302.2152 [astro-ph.GA]. URL: http://arxiv.org/abs/1302.2152.

Trenti, M. et al. (Jan. 21, 2007). "Star clusters with primordial binaries - III. Dynamical interaction between binaries and an intermediate-mass black hole". In: *Monthly Notices of the Royal Astronomical Society* 374, pp. 857–866. ISSN: 0035-8711,1365-2966. DOI: 10.1111/j.1365-2966.2006.11189.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2006.11189.x.

Trotta, R. (Mar. 2008). "Bayes in the sky: Bayesian inference and model selection in cosmology". en. In: *Contemporary Physics* 49, pp. 71–104. ISSN: 0010-7514,1366-

5812. DOI: 10.1080/00107510802066753. URL: http://dx.doi.org/10.1080/00107510802066753.

Trumpler, R. J. (Jan. 1930). "Preliminary results on the distances, dimensions and space distribution of open star clusters". In: *Lick Observatory Bulletins* 14, pp. 154–188. ISSN: 0075-9317. DOI: 10.5479/ads/bib/1930licob.14.154t. URL: http://dx.doi.org/10.5479/ADS/bib/1930LicOB.14.154T.

Vasiliev, E. (Jan. 11, 2019). "AGAMA: action-based galaxy modelling architecture". en. In: *Monthly Notices of the Royal Astronomical Society* 482, pp. 1525–1544. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/sty2672. URL: http://dx.doi.org/10.1093/mnras/sty2672.

Vasiliev, E. and H. Baumgardt (July 6, 2021). "Gaia EDR3 view on galactic globular clusters". en. In: *Monthly Notices of the Royal Astronomical Society* 505, pp. 5978–6002. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stab1475. URL: http://dx.doi.org/10.1093/mnras/stab1475.

Vázquez-Semadeni, E. et al. (Dec. 11, 2019). "Global hierarchical collapse in molecular clouds. Towards a comprehensive scenario". en. In: *Monthly Notices of the Royal Astronomical Society* 490, pp. 3061–3097. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stz2736. arXiv: 1903.11247 [astro-ph.GA]. URL: http://dx.doi.org/10.1093/mnras/stz2736.

Vehtari, A. et al. (June 1, 2021). "Rank-normalization, folding, and localization: An improved R^ for assessing convergence of MCMC (with discussion)". en. In: *Bayesian Analysis* 16. ISSN: 1936-0975,1931-6690. DOI: 10.1214/20-ba1221. URL: http://dx.doi.org/10.1214/20-ba1221.

Vesperini, E. et al. (Feb. 24, 2021). "Dynamical evolution of multiple-population globular clusters". en. In: *Monthly Notices of the Royal Astronomical Society* 502, pp. 4290–4304. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stab223. URL: http://dx.doi.org/10.1093/mnras/stab223.

Vishniac, E. T. (Aug. 1978). "A necessary condition for equilibrium in stellar systems with a continuous mass spectrum". en. In: *The Astrophysical Journal* 223, p. 986. ISSN: 0004-637X,1538-4357. DOI: 10.1086/156332. URL: http://dx.doi.org/10.1086/156332.

Vorobyov, E. I. and S. Basu (May 27, 2015). "Variable protostellar accretion with episodic bursts". In: *The Astrophysical Journal* 805, p. 115. ISSN: 0004-637X,1538-4357. DOI: 10.1088/0004-637x/805/2/115. URL: http://dx.doi.org/10.1088/0004-637x/805/2/115.

Wan, Z. et al. (Dec. 16, 2022). "Dynamics in the outskirts of four Milky Way globular clusters: it's the tides that dominate". en. In: *Monthly Notices of the Royal Astronomical Society* 519, pp. 192–207. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stac3566. URL: http://dx.doi.org/10.1093/mnras/stac3566.

Wasserman, L. (Sept. 15, 2004). *All of statistics*. en. 1st ed. Springer Texts in Statistics. Springer. ISBN: 9780387402727.

Watkins, L. L. et al. (Apr. 10, 2015). "Hubble space telescopeproper motion (hstpromo) catalogs of galactic globular clusters. Ii. Kinematic profiles and maps". In: *The Astrophysical Journal* 803, p. 29. ISSN: 0004-637X,1538-4357. DOI: 10.1088/0004-637x/803/1/29. URL: http://dx.doi.org/10.1088/0004-637x/803/1/29.

Webb, J. J., M. Reina-Campos, and J. M. D. Kruijssen (July 11, 2019). "A systematic analysis of star cluster disruption by tidal shocks – I. Controlled N-body simulations and a new theoretical model". en. In: *Monthly Notices of the Royal Astronomical Society* 486, pp. 5879–5894. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stz1264. URL: http://dx.doi.org/10.1093/mnras/stz1264.

Webb, J. J. et al. (Nov. 21, 2014). "The effects of orbital inclination on the scale size and evolution of tidally filling star clusters". en. In: *Monthly Notices of*

the Royal Astronomical Society 445, pp. 1048–1055. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stu1763. URL: http://dx.doi.org/10.1093/mnras/stu1763.

Wen, R. Y. et al. (Nov. 9, 2023). "Hierarchical Bayesian inference of globular cluster properties". en. In: *Monthly Notices of the Royal Astronomical Society* 527, pp. 4193–4208. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stad3536. URL: http://dx.doi.org/10.1093/mnras/stad3536.

White, H. (Jan. 1982). "Maximum likelihood estimation of misspecified models". In: *Econometrica: Journal of the Econometric Society* 50, p. 1. ISSN: 0012-9682,1468-0262. DOI: 10.2307/1912526. URL: http://dx.doi.org/10.2307/1912526.

Wieland, F.-G. et al. (Mar. 2021). "On structural and practical identifiability". en. In: *Current Opinion in Systems Biology* 25, pp. 60–69. ISSN: 2452-3100. DOI: 10.1016/j.coisb.2021.03.005. URL: http://dx.doi.org/10.1016/j.coisb.2021.03.005.

Wilson, C. P. (Mar. 1975). "Dynamical models of elliptical galaxies". In: *The Astronomical Journal* 80, p. 175. ISSN: 0004-6256,1538-3881. DOI: 10.1086/111729. URL: http://dx.doi.org/10.1086/111729.

Wolf, J. et al. (May 2010). "Accurate masses for dispersion-supported galaxies: Accurate masses for spheroidal galaxies". en. In: *Monthly Notices of the Royal Astronomical Society*, no–no. ISSN: 0035-8711,1365-2966. DOI: 10.1111/j.1365-2966.2010.16753.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2010.16753.x.

Wright, N. J. et al. (Aug. 7, 2024). "The *Gaia*-ESO Survey: 3D dynamics of young groups and clusters from GES and *Gaia* EDR3". en. In: *Monthly Notices of the Royal Astronomical Society* 533, pp. 705–728. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stae1806. URL: http://dx.doi.org/10.1093/mnras/stae1806.

Yalyalieva, L. et al. (Apr. 30, 2022). "The young Galactic cluster NGC 225: binary stars' content and total mass estimate". en. In: *Monthly Notices of the Royal Astronomical Society*. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stac1199. URL: http://dx.doi.org/10.1093/mnras/stac1199.

Zinn, J. C. (May 1, 2021). "Validation of the Gaia Early Data Release 3 parallax zero-point model with asteroseismology". In: *The Astronomical Journal* 161, p. 214. ISSN: 0004-6256,1538-3881. DOI: 10.3847/1538-3881/abe936. URL: http://dx.doi.org/10.3847/1538-3881/abe936.

Zocchi, A. et al. (Oct. 11, 2016). "Testing lowered isothermal models with directN-body simulations of globular clusters". en. In: *Monthly Notices of the Royal Astronomical Society* 462, pp. 696–714. ISSN: 0035-8711,1365-2966. DOI: 10.1093/mnras/stw1104. URL: http://dx.doi.org/10.1093/mnras/stw1104.

Zotos, E. E. (Oct. 2014). "Escapes in Hamiltonian systems with multiple exit channels: part I". en. In: *Nonlinear Dynamics* 78, pp. 1389–1420. ISSN: 0924-090X,1573-269X. DOI: 10.1007/s11071-014-1524-9. URL: http://dx.doi.org/10.1007/s11071-014-1524-9.

Zwart, S. P. et al. (1998). "On the dissolution of evolving star clusters". In: *arXiv [astro-ph]*. DOI: 10.48550/ARXIV.ASTRO-PH/9803084. eprint: astro-ph/9803084 (astro-ph). URL: http://dx.doi.org/10.48550/ARXIV.ASTRO-PH/9803084.

Zwintz, K. et al. (Jan. 12, 2005). "Search for pulsating pre-main-sequence stars in NGC 6383: Search for pulsating PMS stars in NGC 6383". en. In: *Monthly Notices of the Royal Astronomical Society* 357, pp. 345–353. ISSN: 0035-8711,1365-2966. DOI: 10.1111/j.1365-2966.2005.08655.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2005.08655.x.