Dissertation

submitted to the

Combined Faculty of Natural Sciences and Mathematics

of the Ruperto-Carola-University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Put forward by

Victor Francisco Ksoll

born in: Diez, Germany

Oral examination: 7 July 2021

Characterising Pre-Main-Sequence Stars in the Large Magellanic Cloud with Machine and Deep Learning Techniques

Referees: Prof. Dr. Ralf Klessen
           Prof. Dr. Cornelis Dullemond

Abstract

The Large Magellanic Cloud (LMC) exhibits an extraordinary star-forming activity, providing excellent targets for star formation research. Photometric observations with the Hubble Space Telescope (HST) allow for deep, high-resolution studies of young stellar clusters and still-forming pre-main-sequence (PMS) stars in the LMC. In this thesis we study two LMC star-forming complexes, the Tarantula Nebula and N44. Using HST photometry of the Tarantula Nebula from the "Hubble Tarantula Treasury Project" (HTTP), we devise a machine-learning (ML) classification procedure to identify PMS stars from photometry and recover the PMS population captured by the HTTP survey. We introduce new HST observations of N44, the "Measuring Young Stars in Space and Time" (MYSST) survey, identify N44's PMS content with our ML classification procedure, and conduct a clustering analysis of the identified PMS stars. Additionally, we develop a conditional invertible neural network approach to predict stellar physical parameters from photometric observations, based on the PARSEC stellar evolution models. We perform a test on HST observations of the Milky Way clusters Westerlund 2 and NGC 6397, and successfully confirm previous findings on e.g. the age of Westerlund 2. For NGC 6397, however, we identify discrepancies between the PARSEC stellar evolution models and HST observations that prevent accurate predictions.

Zusammenfassung

Die Große Magellansche Wolke (LMC) besitzt eine außergewöhnliche Sternentstehungsaktivität und bietet hervorragende Ziele für Sternentstehungsforschung. Photometrische Beobachtungen mit dem Hubble-Weltraumteleskop (HST) erlauben tiefe, hochauflösende Studien von Sternhaufen und von in der Entstehung befindlichen Vorhauptreihensternen in der LMC. In dieser Dissertation untersuchen wir zwei Sternentstehungskomplexe der LMC, den Tarantelnebel und N44. Mit Hilfe der HST-Photometrie des Tarantelnebels vom „Hubble Tarantula Treasury Project" (HTTP) entwickeln wir eine Klassifikationsprozedur mit maschinellem Lernen (ML), um Vorhauptreihensterne anhand von Photometrie zu identifizieren und ermitteln die Vorhauptreihenpopulation des Tarantelnebels. Wir stellen außerdem neue HST-Beobachtungen von N44 vor, das „Measuring Young Stars in Space and Time" (MYSST) Projekt, identifizieren die Vorhauptreihensterne in N44 mit unserer ML-Klassifikationsmethode und führen eine Clustering-Analyse der identifizierten Vorhauptreihensterne durch. Zusätzlich entwickeln wir ein Conditional Invertible Neural Network mit Hilfe der PARSEC Sternentwicklungsmodelle, um physikalische Eigenschaften von Sternen anhand photometrischer Beobachtungen vorherzusagen. Wir testen unsere Methode auf HST-Beobachtungen der Sternhaufen Westerlund 2 und NGC 6397 in der Milchstraße und können erfolgreich bekannte Ergebnisse für z.B. das Alter von Westerlund 2 bestätigen. Für NGC 6397 jedoch identifizieren wir eine Diskrepanz zwischen den PARSEC Sternentwicklungsmodellen und HST-Beobachtungen, welche präzise Vorhersagen verhindert.

Contents

List of Figures

List of Tables

List of Publications

First Author Papers (included in this thesis)

1. **Ksoll V. F.**, Gouliermis D. A., Klessen R. S., Grebel E. K., Sabbi E., Anderson J., Lennon D. J., Cignoni M., de Marchi G., Smith L. J., Tosi M., van der Marel R. P. *Hubble Tarantula Treasury Project - VI. Identification of pre-main-sequence stars using machine-learning techniques.* MNRAS, 479(2):2389–2414, 2018. doi:10.1093/mnras/sty1317.

2. **Ksoll V. F.**, Ardizzone L., Klessen R. S., Koethe U., Sabbi E., Robberto M., Gouliermis D. A., Rother C., Zeidler P., Gennaro M. *Stellar parameter determination from photometry using invertible neural networks.* MNRAS, 499(4):5447–5485, 2020. doi:10.1093/mnras/staa2931.

3. **Ksoll V. F.**, Gouliermis D. A, Sabbi E., Ryon J. E., Robberto M., Gennaro M., Klessen R. S., Koethe U., de Marchi G., Chen C.-H. R., Cignoni M., Dolphin A. E. *Measuring Young Stars in Space and Time – I. The Photometric Catalog and Extinction Properties of N44.* Astronomical Journal, in production, tbd., 2021. preprint: arXiv:2012.00521.

4. **Ksoll V. F.**, Gouliermis D. A, Sabbi E., Ryon J. E., Robberto M., Gennaro M., Klessen R. S., Koethe U., de Marchi G., Chen C.-H. R., Cignoni M., Dolphin A. E. *Measuring Young Stars in Space and Time – II. The Pre-Main-Sequence Stellar Content of N44.* Astronomical Journal, in production, tbd., 2021. preprint: arXiv:2012.00524

Co-Authored Papers (not in this thesis)

5. Haldemann J., **Ksoll V. F.**, Walter D., Alibert Y., Klessen R. S. and Benz W. *Planet Characterization using invertible Neural Networks*, in prep.

6. Rousseau-Nepton L., et. al (including **Ksoll V. F.**), *SIGNALS: I. Survey description*, MNRAS, 489(4):5530–5546, 2019.

Introduction

Stars and the night sky have fascinated humankind for centuries and have influenced our history in various ways, be it through e.g. astronomical calendars, celestial navigation or science. Not without reason astronomy is believed to be one of the oldest occupations in the world. For instance, Hipparchus already compiled a catalogue of over a thousand stars in the second century BC (Leverington, 2003). Studying stars is a fundamental part of understanding our universe. After all, stars are the very entities that bring light into the galaxies and in the case of our Sun are the direct reason that life on Earth continues to exist. With the discovery of nuclear fusion as the central energy source within stars and the subsequent realisation that (almost) all elements heavier than hydrogen are synthesised within the hot, dense furnaces in the centres of stars, as well as during the explosive demise of the most massive stars, trying to understand stars has turned into a quest for the very early origins of human life itself (Beech, 2019).

In the study of stars one of the central questions is how the stars themselves form. Going from the diffuse interstellar medium to the hot and dense environments necessary to ignite and sustain nuclear fusion, star formation is a complex process that covers many orders of magnitude in length scales, temperature and density (Schulz, 2012; Bodenheimer, 2011; Stahler and Palla, 2004). By now we have a fairly decent conceptual understanding of stellar birth, starting from giant molecular clouds, enormous reservoirs of molecular hydrogen, that form dense cores through density fluctuations driven by large scale turbulence, which then in turn begin to gravitationally collapse. Continuously accreting more material from their natal environment, these protostellar cores then contract under their own self-gravity until their interiors become hot and dense enough to finally ignite hydrogen fusion and become a star (Klessen and Glover, 2016). Many details regarding this process are, however, still not yet fully understood. This includes e.g. the

processes that dictate how much gas is actually converted to stars, how feedback of massive stars affects and perhaps even triggers star formation, or whether the initial mass function – the distribution of stellar masses at birth – is truly a universal relation or not (e.g. Bodenheimer, 2011). To understand these phenomena we need to observe star formation as it is happening.

This is not an easy undertaking, as there are many factors that make observing star formation difficult. One of the primary difficulties arises from the timescales on which star formation takes place. Ranging for instance into the tens of millions of years for the formation of a solar mass star, we cannot simply observe a star forming from start to finish (Stahler and Palla, 2004). Instead, we have to capture the process of stellar birth at different stages, in varying environments and with large number statistics in order to properly piece this puzzle together. Other challenges arise e.g. from the fact that young forming stars are often heavily embedded within their natal environments and are, thus, subject to large amounts of extinction from the surrounding gas and dust (Schulz, 2012; Bodenheimer, 2011).

One of the main observational techniques to capture information about star formation are large photometric surveys of active star-forming regions and young stellar clusters. Although photometry in contrast to spectroscopy provides only glimpses into the total spectrum of a star, measuring a star's brightness in a series of broadband filters, its lower observational time requirements make photometry an integral approach to observe large amounts of individual stars, in particular in extragalactic environments where spectroscopy becomes unfeasible. Among the prime targets to study star formation are the Magellanic Clouds, the two companion dwarf galaxies of our Milky Way, as they exhibit an extraordinary star-forming activity with many large and impressive centres of stellar birth, which are observable at low extinction due to the position of the Clouds above and away from the Milky Way's dusty disk (Gouliermis, 2012).

Over the last decades astronomy has become an exceptionally data-rich environment. With the constant development of more precise instrumentation and the construction of larger and better telescopes, both ground-based – e.g. the Very Large Telescope (VLT), Extremely Large Telescope (ELT), Vera C. Rubin Observatory (VRO) – and in space – e.g. Hubble Space Telescope,

James Webb Space Telescope (JWST), Gaia – large surveys, in particular the all-sky variants – e.g. Sloan Digital Sky Survey (SDSS), Gaia, Panoramic Survey Telescope And Rapid Response System (Pan-STARRS) – deliver data sets now that may contain billions of stars. With astronomy arriving in the big data era the adequate analysis of these enormous data sets requires efficient automatic algorithms. Machine and deep learning, revolving around data-driven algorithms that learn from data itself to make predictions, are the obvious choice to tackle these challenges. From galaxy morphology classification (e.g. Huertas-Company et al., 2015) to exoplanet identification (e.g. Armstrong et. al, 2018), machine learning approaches have successfully solved a variety of astronomical problems over the last decade (see e.g. Fluke and Jacobs, 2020, for a recent review). Although arguably now established in astronomy, machine learning still remains a relatively new addition to the tool set of the community. Therefore, continuous development and exploration of the steadily evolving machine learning methodologies for astronomical questions is necessary to fully use the wealth of data that is and will be available to astronomers.

1.0.1 Motivation and Outline

In photometric surveys of young stellar clusters and large star-forming re-gions one of the main difficulties in *identifying* young, still-forming pre-main-sequence (PMS) stars arises from the fact that old populations (from e.g. the host galaxy's field) often significantly contaminate the line of sight towards the target in the foreground or background. In this case, observational effects such as extinction or photometric uncertainties combined with properties intrinsic to stellar evolution (see Section 1.4) may render it difficult to disen-tangle PMS sources from e.g. lower main-sequence or red giant branch stars (e.g. Gouliermis, 2012). To subsequently investigate central open topics in star formation research, e.g. the initial mass function or primordial mass segregation in clusters, the observed PMS stars need to be *characterised* by recovering their fundamental physical properties, such as age and mass. This constitutes the inverse problem of predicting stellar physical parameters from observations. This task is subject to considerable degeneracy from

e.g. intrinsic properties of stellar evolution or, in the case of photometric surveys, the inherent information loss of photometry. To adequately solve this problem, therefore, methodologies are required that can take these degeneracies into account.

With this thesis we aim to contribute towards solving these challenges in *identifying* and *characterising* PMS stars by *exploring* the application of different machine learning methodologies. In particular, we develop these approaches to gain insights into the environments of two of the most prominent star-forming complexes in the Large Magellanic Cloud, the Tarantula Nebula and LHα 120–N44, and to ultimately advance the recovery of the star formation histories of these regions. Towards this goal we analyse data from two large photometric surveys with the Hubble Space Telescope (HST) called "Hubble Tarantula Treasury Project" (HTTP, see Section 2.4) and "Measuring Young Stars in Space and Time" (MYSST, see Section 2.5).

Following the outlined directive, we present four studies in this cumulative thesis. In our first study (Section 3.1) we employ the HTTP photometric catalogue to establish a machine learning classification approach to *identify* and distinguish the young PMS sources from old contaminating populations. Subsequently, we recover a census of the PMS population of the Tarantula Nebula. In our second study (Section 3.2) we develop a deep learning regression approach, based on the conditional neural network architecture, to *characterise* PMS stars by solving the inverse problem of predicting stellar physical parameters from photometry. In this work we detail the proof-of-concept implementation and perform a real-data benchmark on HST observations of the two well-studied Milky Way clusters, Westerlund 2 and NGC 6397. In the third study (Section 3.3) we introduce the MYSST survey of LHα 120–N44, presenting its observing strategy and photometric catalogue, and devise an application of the RANSAC machine learning algorithm to constrain reddening properties of the region. In our fourth and final study (Section 3.4) we conduct the first in-depth analysis of the MYSST data and apply the PMS identification procedure from our first study to recover the PMS census of N44. Additionally, we perform a clustering analysis of the identified PMS population and quantify the spatial structure of star-forming centres across N44.

In total this thesis consists of four chapters and is structured as follows. Chapter 1 provides a general introduction to the relevant topics that are subject of this thesis. Section 1.1 gives an overview of the star formation process and the natal environments of stars, the giant molecular clouds. Section 1.2 briefly summarises the post-main-sequence evolution of stars and Section 1.3 discusses observation of star formation. Sections 1.4 and 1.5 introduce the methodological background of the machine and deep learning approaches applied throughout this thesis. Lastly, Section 1.6 gives a short summary of related studies. In Chapter 2 we provide an overview of the target regions and associated data sets that are analysed throughout this thesis. Chapter 3 presents the four studies, Ksoll et al. (2018, 2020, 2021a,b), that make up the core of this thesis as they appear in the Monthly Notices of the Royal Astronomical Society and The Astronomical Journal. The final Chapter 4 gives a summary and discussion of the work done in this thesis, as well as an outlook on potential follow-up studies.

1.1 Star Formation

1.1.1 The Interstellar Medium

Properties of the ISM

Giant molecular clouds – enormous and dense accumulations of molecular hydrogen – are the birth places of stars (Kennicutt and Evans, 2012). Before we describe the properties of molecular clouds and the processes that lead to the formation of stars within them, however, we have to ask from where these clouds themselves originate. The answer is the interstellar medium (ISM), i.e. the baryonic matter that permeates and fills a galaxy in between stars (not to be confused with the *intergalactic* medium between galaxies; Schulz, 2012). The ISM consists primarily of hydrogen ($\sim 70\%$) and helium ($\sim 28\%$), with heavier elements accounting for the remaining 2%. The former two are mainly found in the gas phase, whereas a substantial fraction of the heavier elements may be bound in *dust* grains (Girichidis et al., 2020).

Besides the molecular clouds the gas component of the ISM is often further distinguished into four main phases that coexist in an approximate pressure equilibrium (Beech, 2019; Stahler and Palla, 2004). These are the cold neutral medium (CNM), the warm neutral medium (WNM), the warm ionised medium (WIM) and the hot ionised medium (HIM), whose characteristic temperatures, densities and length scales are summarised in Table 1.1.

Tab. 1.1.: Properties of the components of the ISM.

Component	Temperature (K)	Density (cm^{-3})	Length Scales (pc)
HIM	10^6	0.01	Galactic (50%)
WIM	8000	$0.03 - 0.5$	Galactic ($15 - 25\%$)
WNM	$6000 - 10000$	$0.1 - 0.5$	Galactic ($15 - 30\%$)
CNM	$50 - 100$	$20 - 50$	Galactic ($\sim 5\%$)
GMC	15	> 100	~ 50 ($< 1\%$)
MC	10	> 300	~ 10
MC clump	10	10^3	~ 5
MC core	10	10^5	$\sim 0.1 - 0.3$

The term in parenthesis in the length scale column refers to the volume filling factor of each component. The 'Galactic' keyword in the same column indicates that a given phase is widespread through the Galactic disc. Compiled from Girichidis et al. (2020); Beech (2019); Schulz (2012); Bodenheimer (2011); Stahler and Palla (2004).

Most of the volume of the ISM is occupied by ionised gas, but this component accounts only for a small fraction ($\sim 25\%$) of the total gas mass of a galaxy ($M_{\mathrm{gas,tot}} \sim 10^{10}\,M_\odot$ for e.g. the Milky Way; Kalberla and Kerp, 2009). Most of the mass is located in regions of neutral atomic (H, He) or molecular gas ($\mathrm{H_2}$), which are primarily found in dense clouds that occupy only a minute fraction ($\sim 1\%$) of the total ISM volume (Girichidis et al., 2020). This picture of distinct ISM phases is, of course, not absolute and there is considerable mixing between the various components, in particular because the ISM is subject to significant amounts of turbulence, driven by effects such as thermal instability, supernova feedback and gas inflow on the Galactic disc (Girichidis et al., 2020; Klessen and Glover, 2016).

As Table 1.1 demonstrates, there are quite significant differences in density and temperature between the material that occupies most of the volume of a galaxy, but is too hot and diffuse to form stars, and the molecular clouds

that stars are born in. This begs the question how the ISM coagulates into molecular clouds to provide the initial conditions for star formation.

Formation of Molecular Hydrogen

Considering only ISM chemistry, this breaks down to the formation of molecular hydrogen H_2. The most basic formation pathways for molecular hydrogen through radiative association

$$H + H \rightarrow H_2 + \gamma \tag{1.1}$$

and ion-neutral reactions, e.g.

$$H + e^- \rightarrow H^- + \gamma$$
$$H^- + H \rightarrow H_2 + e^-, \tag{1.2}$$

are not very efficient in the ISM. Instead H_2 forms primarily not in the gas phase, but via association reactions of H atoms adsorbed on *dust grains* (Klessen and Glover, 2016). A viable H_2 formation mechanism alone is, however, not sufficient to form a molecular cloud as effects that destroy H_2 have to be suppressed at the same time in order for H_2 to accumulate. The main relevant H_2 destruction mechanism is *photodissociation*, where an H_2 molecule is excited by a UV photon with energy $E > 11.2\,\mathrm{eV}$ from the interstellar radiation field and in the subsequent radiative de-excitation transition returns to a ground state in which the molecule dissociates (*spontaneous radiative dissociation*; Stecher and Williams, 1967; van Dishoeck and Black, 1988). Being line-based, photodissociation of H_2 is subject to the effect of *self-shielding*. This means in regions of high H_2 column density photons with the energies of the main absorption lines are primarily absorbed by H_2 in the outer parts, drastically reducing the amount of photons reaching the centre and subsequently decreasing the central dissociation rate significantly (Klessen and Glover, 2016). Following these considerations of the ISM chemistry, a molecular cloud, thus, forms simply when enough gas is accumulated in one place to exceed the necessary column density for

effective self-shielding from the interstellar radiation field for as long as H_2 needs to form (Klessen and Glover, 2016).

Cloud Formation Scenarios

Although the main process for molecular cloud formation appears fairly clear from an ISM chemistry point of view, we are yet to describe how the necessary column densities are reached within the ISM. This topic is in fact still subject to ongoing research and a definite formation scenario for molecular clouds is currently debated. We shall briefly outline three potential molecular cloud formation scenarios.

In the *coagulation model*, originally proposed by Oort (1954) (see also e.g. Field and Saslaw, 1965; Tasker and Tan, 2009), the ISM is considered to consist of cold atomic and molecular clouds (of various size and mass), where small atomic clouds initially form via thermal instabilities of warmer atomic gas. Larger clouds are then formed via cloud-cloud collisions, efficiently dissipating energy and coagulating the interacting clouds. At some point the growing atomic clouds reach the necessary column densities for effective self-shielding and become predominantly molecular. In this picture the now-molecular cloud may continue to grow by accreting even more mass via collisions, potentially becoming very massive. This cycle of continuous cloud growth is expected to end only once star formation sets in and stellar feedback disrupts the cloud (Klessen and Glover, 2016).

An alternative proposed mechanism is the *converging flow model*. It postulates that molecular gas forms in the dense, post-shock regions created where flows of low-density gas converge, collide and interact. If the flows originally consist of warm atomic hydrogen then thermal instabilities can be triggered by these collisions, rapidly producing cold clouds, which are dense enough to enable self-shielding from the interstellar radiation field (see e.g. Hennebelle and Pérault, 1999; Koyama and Inutsuka, 2002; Heitsch and Hartmann, 2008; Klessen and Glover, 2016). Both the coagulation and the converging flow models provide potential pathways for the formation of small and intermediate molecular clouds, but are found to struggle with reproducing the most massive observed clouds (Klessen and Glover, 2016).

A third proposed picture assumes a more top-down formation scenario, where molecular clouds are created by larger scale gravitational instabilities in the Galactic disc (Chevance et al., 2020b). This process of self-gravitating fragmentation favours higher density environments, e.g. spiral arms, and tends to form fairly massive molecular clouds (up to $10^6\,M_\odot$). In this picture low mass clouds can only form when stellar feedback disrupts the gravitationally unstable material before it has collapsed completely. For a more thorough discussion of the potential molecular cloud formation scenarios, their advantages and shortcomings, we shall refer to more dedicated reviews, e.g. Chevance et al. (2020b); Klessen and Glover (2016); Dobbs et al. (2014); Hennebelle and Falgarone (2012).

1.1.2 Giant Molecular Clouds

Cloud properties

With the formation scenarios for giant molecular clouds (GMC) outlined, we can now proceed to describe GMC properties and the processes that initiate star formation within them. Table 1.1 provides an overview of temperatures, densities and length scales of GMCs and their substructures. Molecular clouds (MC) are found in sizes between a few pc and up to 200 pc for the very enormous GMCs, covering a mass range of several 10s to $10^6\,M_\odot$ (Schulz, 2012). GMCs consist primarily of molecular hydrogen, but also entail other molecular species, e.g. CO, NH_3 or HCN (Rosen et al., 2020), and are often surrounded by an envelope of atomic gas (Klessen and Glover, 2016). GMCs and MCs are very cold with temperatures between $10 - 20$ K (Girichidis et al., 2020). They reach and maintain these low temperatures through a complex balance of different heating and cooling mechanisms (Schulz, 2012).

The primary heating processes are *photodissociation* of H_2 by interstellar photons, *photoionisation* of atomic carbon by the interstellar radiation field, *cosmic ray ionisation* of H and H_2, *photoelectrons* from dust grains produced by interstellar photons and (during collapse) *compressional heating*. Cooling is mainly provided by *collisional excitation* (of e.g. C^+, C, O or CO) and the

subsequent radiative decay and escape of photons, as well as *dust grain cooling*. In the latter gas molecules collide with dust grains, heating the dust and cooling the gas, followed by the dust radiating away the excess energy in the infrared, escaping the cloud as long as it is optically thin to this radiation (Bodenheimer, 2011). A more thorough discussion of the intricate interplay of cooling and heating is beyond the scope of this introduction and we shall refer to e.g. Girichidis et al. (2020); Klessen and Glover (2016); Schulz (2012); Bodenheimer (2011) for more details.

Like the ISM, MCs are turbulent environments (Rosen et al., 2020). Although the extend is still debated, it is argued that this turbulence at least partially provides support for MCs against collapse (Krause et al., 2020). Also characteristic for GMCs and MCs is their highly hierarchical structure, exhibiting complexes of filaments, dense clumps and even denser gas accumulations that are commonly referred to as cores (Rosen et al., 2020; Krause et al., 2020; Klessen and Glover, 2016). It is these dense ($n > 10^5 \, \mathrm{cm}^{-3}$), small ($< 0.1$ pc) cores that are the progenitors of stars. Dense cores can be further distinguished into *prestellar* and *protostellar* cores. A *protostellar* core is a dense core that is already so far into the star-formation process that it harbours a central luminosity source, contrary to a prestellar core. A *prestellar* core may have already initiated star formation, i.e. it is collapsing or about to (bound prestellar core), but may also not be forming a star or ever do so (unbound prestellar core, Rosen et al., 2020).

The Jeans criterion

So far we have mentioned that stars form when MCs and MC cores collapse, but we have not yet discussed the forces that drive or oppose this collapse. The primary driving force of collapse is, of course, gravity. The forces that oppose collapse (to a varying and not yet fully established degree) are thermal gas pressure, rotation, turbulent pressure and magnetic fields (Bodenheimer, 2011). For a dense core, clump, filament or whole MC to collapse its (self) gravitational energy must, therefore, exceed that of all opposing forces, i.e.

$$|E_{\mathrm{grav}}| > E_{\mathrm{th}} + E_{\mathrm{rot}} + E_{\mathrm{turb}} + E_{\mathrm{mag}}. \tag{1.3}$$

To analyse whether a certain object is about to collapse, common helpful concepts are the *Jeans length* and *mass*, originally proposed by James Jeans (Beech, 2019). Considering a uniform density/temperature sphere and taking the thermal pressure as the dominant opposing force to collapse, we can determine a collapse criterion by equating the gravitational energy (of a uniform density sphere)

$$E_{\mathrm{grav}} = -\frac{3}{5}\frac{\mathrm{G}M^2}{R}, \tag{1.4}$$

where G is the gravitational constant, and M and R the mass and radius of the sphere, to the thermal energy of an isothermal, ideal gas

$$E_{\mathrm{th}} = \frac{3}{2}\frac{k_{\mathrm{b}}TM}{\mu m_{\mathrm{u}}}, \tag{1.5}$$

with Boltzmann constant k_{b}, gas temperature T, atomic mass unit m_{u} and gas molecular weight μ (Bodenheimer, 2011). This returns the following characteristic radius R_{J}, the (thermal) Jeans length,

$$R_{\mathrm{J}} = \frac{2}{5}\frac{\mathrm{G}M\mu m_{\mathrm{u}}}{k_{\mathrm{b}}T} \tag{1.6}$$

and characteristic mass M_{J}, the (thermal) Jeans mass,

$$M_{\mathrm{J}} = \left(\frac{5}{2}\frac{k_{\mathrm{b}}T}{\mu m_{\mathrm{u}}\mathrm{G}}\right)^{\frac{3}{2}}\left(\frac{4}{3}\pi\rho\right)^{-\frac{1}{2}}, \tag{1.7}$$

where ρ is the gas density. A sphere of isothermal, ideal gas with constant density is unstable and will collapse under its own gravity if its radius is smaller than the Jeans length $R < R_{\mathrm{J}}$ or, alternatively formulated, if its mass exceeds the Jeans mass $M > M_{\mathrm{J}}$ (Bodenheimer, 2011). Analogous Jeans criteria (i.e. masses and lengths) can be derived for the other possible opposing forces. An object freely collapsing under self-gravity will do so on the *free-fall timescale* t_{ff}

$$\mathrm{t}_{ff} = \sqrt{\frac{3\pi}{32\mathrm{G}\rho}} \sim 3\,\mathrm{Myr}\left(\frac{n}{10^2\,\mathrm{cm}^{-3}}\right)^{-\frac{1}{2}}, \tag{1.8}$$

where n is the gas number density.

Cloud lifetimes

GMCs are relatively short lived objects. A recent census of the GMC population across several nearby star-forming galaxies puts their lifetimes on average on the order of a few 10^7 yr (Kruijssen et al., 2019; Chevance et al., 2020a). If only thermal pressure were to oppose gravity in a MC, it would collapse on timescales comparable to $t_{\rm ff}$. Comparing Eq. (1.8) with these observed average GMC lifetimes, thus, indicates that the other gravity opposing forces, e.g. turbulence, must play a notable role in supporting GMCs against gravity (Klessen and Glover, 2016). The measured GMC lifetimes also have an interesting implication when compared to the molecular *gas depletion* time of their host galaxies, i.e. the time required to turn a galaxy's entire reservoir of molecular gas into stars at its current star formation rate ($\sim 1 - 2$ Gyr, e.g. Bigiel et al., 2008; Leroy et al., 2008). From the fact that GMC lifetimes are considerably shorter than this, the integrated star formation efficiency (ratio of GMC lifetime to gas depletion time) of GMCs can be determined to be only about $2 - 10\%$ (Chevance et al., 2020b). In other words, star formation in GMCs is a fast and inefficient process, where only a small fraction of a GMCs total mass budget of gas is actually converted to stars before they are rapidly dispersed, likely by early stellar feedback (e.g. photoionisation or stellar winds, Chevance et al., 2020b)

The gravoturbulent and global hierarchical collapse scenarios

To describe the global dynamics of a GMC and the internal processes that create the dense filaments, clumps and cores, which ultimately may collapse to form stars, two general scenarios have been proposed. The first one is the *gravoturbulent* (GT) scenario (e.g. Klessen et al., 2000; Vázquez-Semadeni et al., 2003; Mac Low and Klessen, 2004). In this picture GMCs are globally supported againt collapse by the pressure exerted by continuously driven, supersonic, small-scale, isotropic turbulence. At the same time, this turbulence generates local density fluctuations by producing shocks, which manifest in filaments, clumps and cores that then may become Jeans unstable and collapse to form stars (Krause et al., 2020).

The second proposed scenario is the *global hierarchical collapse* (GHC; Vázquez-Semadeni et al., 2019). In this view GMCs quickly exceed their thermal Jeans mass, because a) it is low in cold, dense gas and b) GMCs actively accrete additional material from their diffuse environments (at least in simulations, e.g. Ballesteros-Paredes et al., 1999; Hartmann et al., 2001; Heiner et al., 2015; Wareing et al., 2019). As the clouds, thus, generally grow in mass due to this accretion, they eventually become both gravitationally unstable and magnetically supercritical (i.e. the magnetic field can no longer provide support against collapse). Initial cloud-internal (moderately supersonic) turbulence is generated during the cloud formation process and induces non-linear density fluctuations in the clouds (which have significantly shorter free-fall times than the cloud average; Krause et al., 2020). Unlike in the GT scenario, however, this turbulence cannot provide longtime support and is quickly overwhelmed by the cloud's self gravity, such that the cloud enters a global gravitational contraction (Krause et al., 2020). The resulting infall motions are highly chaotic so that this collapse maintains a turbulent component (e.g. Klessen and Hennebelle, 2010). Due to the turbulent density fluctuations (and the typically amorphous cloud shapes) the collapse is not homologous, such that the densest cloud regions collapse and terminate their collapse (by reaching protostellar densities) significantly faster than the remaining lower density material (Krause et al., 2020). Consequently, in the GHC scenario a cloud becomes a complex hierarchical system of *collapses within collapses* with different collapse scales, each of which continues to accrete material from the next larger one (Vázquez-Semadeni et al., 2019; Krause et al., 2020). In this picture for instance filaments effectively work as mass funnels from large to small scales (Gómez and Vázquez-Semadeni, 2014), which is consistent with the observation that MC cores and clumps often appear as intersection points of filaments (e.g. Myers, 2009). For more details on these two scenarios of the internal dynamics of GMCs we shall refer to e.g. Krause et al. (2020); Klessen and Glover (2016) at this point.

Initial core collapse and evolution

With the overall properties and larger-scale dynamics of GMCs established we shall now outline the evolution of a collapsing dense MC core towards

forming a star. Initially the MC core collapses almost in free-fall and isother-mally over several orders of magnitude in central density increase. Eventually, however, the densities of the collapsing core reach a point where the central regions start to become opaque to the core's own infrared cooling radiation, so that the energy released by the compression of the gas can no longer be effectively radiated away by the dust (Bodenheimer, 2011; Krumholz, 2020). As soon as this happens the core enters an adiabatic collapse phase, in which the central regions are now starting to heat up considerably. During this transition from isothermal to adiabatic collapse and in the early phases of the latter, collapse is actually temporarily halted in the very centre as a short-lived hydrostatic object, supported by its own internal pressure, is formed, upon which the remaining infalling material is then accreted. This objected is known as the *first core* of the protostar and is expected to be only a few AU in size, entailing a few hundredths of a Solar mass (Krumholz, 2020).

As this first core is further compressed while accumulating infalling gas, its central temperatures eventually reaches and exceeds $T \approx 1000 - 2000$ K. At this point the thermal kinetic energy of the gas becomes large enough to collisionally dissociate molecular hydrogen. Once this process initiates most of the gravitational energy released by the compression of the gas now goes into the dissociation of hydrogen instead of further heating the gas. Slowing down the internal temperature increase considerably, the thermal pressure can then no longer compensate gravity and the core enters a second (free-fall) collapse phase (Krumholz, 2020; Bodenheimer, 2011). This second collapse lasts until the molecular hydrogen is completely dissociated and the produced atomic hydrogen is subsequently ionised. At this point the cloud core reaches densities and temperatures ($\sim 10^4$ K) at which thermal pressure can again provide enough support to halt the free-fall collapse, forming a second hydrostatic core, a true *protostar* (Krumholz, 2020).

When the second hydrostatic core forms, it is still far away from having the necessary mass to ultimately ignite hydrogen fusion and become a star. Thus, the protostar enters the *main accretion phase*, in which it acquires most of its final stellar mass by accreting material from its still-collapsing envelope. As the gas from the envelope plummets onto the protostellar

surface, settling at an accretion shock front, all of the kinetic energy of the gas is converted into radiation, giving rise to significant accretion luminosity. Although the (slowly) gravitationally contracting protostar also provides a source of luminosity, in this phase the contribution from accretion completely dominates the protostar's total luminosity (at least for low- and intermediate-mass stars). Flowing back through the infalling material, gradually degraded to longer and longer wavelengths, with this escaping radiation the protostar now becomes visible in the far infrared (Krumholz, 2020; Bodenheimer, 2011; Stahler and Palla, 2004).

During the accretion phase an important effect comes into play as a direct consequence of the fact that GMC, MCs and dense MC cores are *rotating*. Consequently, the infalling gas from the collapsing envelope of a protostar carries *angular momentum*, which it has to shed in order to be accreted onto the central protostar. Because of this, the accreting protostar forms an *accretion disc*, upon which the gas from the infalling envelope first settles, dissipates its angular momentum and then accretes onto the protostellar surface. Dissipation mechanisms include viscous shear within the disc, magnetohydrodynamically driven disc winds and highly-collimated bi-polar jets, which eject significant amounts of material from the accretion disc back into the surrounding envelope. These winds and jets can significantly disturb the infalling material and may even partially disrupt the collapsing envelope. In both cases they play a notable role in the total amounts of mass that the central protostar can accumulate during its main accretion phase (Beech, 2019; Bodenheimer, 2011; Stahler and Palla, 2004).

As the core accretes more and more material from its envelope and accretion disc, getting heated in the process, the core's internal temperature eventually rises above a value of 10^6 K. Upon reaching this temperature the core crosses the threshold to ignite a first nuclear fusion reaction, i.e. *deuterium burning* via

$$^2\mathrm{H} + {}^1\mathrm{H} \rightarrow {}^3\mathrm{He} + \gamma. \tag{1.9}$$
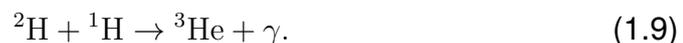
The onset of deuterium burning within the protostar temporarily halts the contraction of the core and also acts as a thermostat, keeping the internal temperature roughly at the deuterium ignition threshold by inducing an adiabatic swelling of the protostar's radius. Additionally, central deuterium
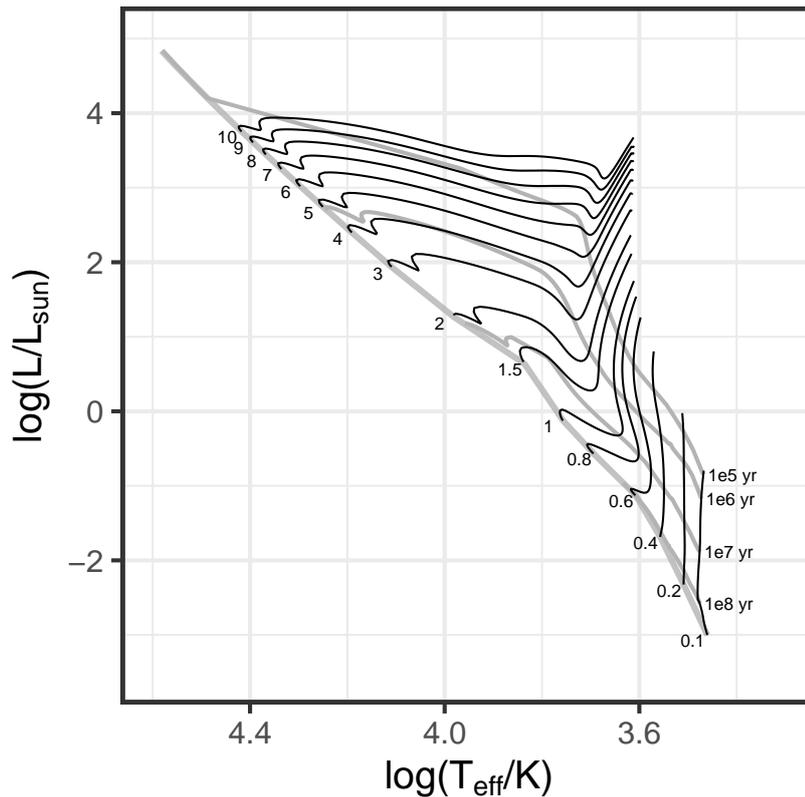
Fig. 1.1.: Hertzsprung-Russell diagram of pre-main-sequence evolutionary tracks of varying initial mass at solar metallicity from the MIST suite of stellar evolution models (Dotter, 2016; Choi et al., 2016). The thick grey line on the left indicates the position of the ZAMS, while the thinner grey lines mark loci of constant age across the different tracks as labelled on the right.

burning releases too much energy to be transported radiatively within the highly opaque interior. Consequently, energy transport within the star shifts to *convection*, i.e. the mechanical motion of buoyant discrete parcels of gas (Krumholz, 2020; Schulz, 2012; Stahler and Palla, 2004).

1.1.3 Pre-Main-Sequence Evolution

To outline the subsequent evolution of the protostellar progenitors of low- and intermediate-mass stars, it is instructive to introduce the Hertzsprung-Russell diagram (HRD). The HRD is a scatter plot that relates a star's luminosity L

to its effective surface temperature T_eff. In this diagram a star's position is directly dependent on its metallicity (chemical composition), mass and evolutionary state (i.e. age). Consequently, given the mass of a star (and its metallicity) its evolution from (the later stages of) formation over the *main-sequence* hydrogen burning phase to its post-main-sequence evolution (see Section 1.2) can be visualised in the HRD as a distinct *evolutionary track* (Schulz, 2012; Stahler and Palla, 2004).

A still-forming star can be placed in the HRD once it has ended its main accretion phase, i.e. accreted most of its final stellar mass from its infalling envelope, and has dispersed enough of its natal MC environment to become observable in the optical regime. This stage of the protostellar evolution is the so-called *pre-main-sequence* phase (PMS; Krumholz, 2020; Bodenheimer, 2011; Stahler and Palla, 2004). Note that the end of the main accretion phase does not mark the termination of all accretion and that, in particular, early-PMS stars can still accrete notable amounts of material from their circumstellar accretion disc, which in turn may also still be fed by the remnants of the natal envelope (Klessen and Glover, 2016).

Low-mass stars ($M \lesssim 0.4\,M_\odot$) enter the PMS phase fully convective and remain so for its entire duration. In more massive stars on the other hand central convection is eventually overturned by radiative transport again, in particular because the core cannot burn deuterium indefinitely. Although convection provides an efficient mixing agent in the protostellar interior to resupply the central deuterium-burning zone with new fuel from the outer layers of the protostar and the newly accreted material, the energy released by deuterium fusion is eventually not sufficient to prevent further contraction of the protostars. Subsequently, the interior temperature starts to rise again and the internal opacity begins to drop accordingly until energy can be effectively transported by radiation again. Consequently, an outwards growing radiative barrier forms within the protostar, ceasing convection within the core completely by cutting of the constant fuel resupply for the deuterium-burning zone, quickly exhausting the remaining deuterium and shutting off the fusion reaction (Krumholz, 2020).

Figure 1.1 shows example PMS evolutionary tracks for different stellar masses from 0.1 to $10\,M_\odot$, where the example PMS stars evolve from the

top right to the bottom left as indicated by the lines of constant age across the tracks. The initial path for the less massive examples ($M \lesssim 3~M_\odot$) is an almost vertical descent at constant effective temperature. This portion is called the *Hayashi track* and marks the convective contraction phase of the PMS evolution (Krumholz, 2020; Stahler and Palla, 2004).

As soon as the star has contracted enough to transition back to radiative energy transport, as described above, the evolutionary track takes a sharp turn in the HRD and subsequently follows an almost horizontal path. This portion is commonly referred to as the *Henyey track* and marks the final contraction phase of the PMS evolution, during which the PMS star continuously increases its internal temperature until it reaches about 10^7 K. At this point the conditions are met for the ignition of hydrogen fusion at the centre of the PMS object. Providing an enormous source of energy the fusion of hydrogen subsequently halts gravitational contraction and a stable hydrogen-burning main-sequence star is born (Krumholz, 2020; Stahler and Palla, 2004). In the HRD the locus, on which PMS stars of different masses appear as they start burning hydrogen, is called *zero-age main-sequence* (ZAMS). In Figure 1.1 it is marked as the thick grey line on the left of the diagram.

PMS stars that are less massive than about $0.072 - 0.075~M_\odot$ (Schulz, 2012; Krumholz, 2020) never reach central temperatures high enough to sustain hydrogen fusion as they contract. Although they may engage in a short phase of deuterium burning like their more massive siblings, their contraction is ultimately ended by electron degeneracy pressure rather than nuclear fusion. These objects are so-called *brown dwarves* (Krumholz, 2020; Schulz, 2012).

The time a forming-star spends in the PMS phase can be reasonably approximated by the *Kelvin-Helmholtz* (KH) timescale (Stahler and Palla, 2004), i.e. the time required for a PMS star to radiate away its gravitational bind-

ing energy to contract to the main sequence (Rosen et al., 2020). The KH timescale t_{KH} is given by

$$
\begin{aligned}
t_{\mathrm{KH}} &= \frac{G M_*^2}{R_* L_*} \\
&= 3 \times 10^7 \,\mathrm{yr} \left(\frac{M_*}{M_\odot} \right)^2 \left(\frac{R_*}{R_\odot} \right)^{-1} \left(\frac{L_*}{L_\odot} \right)^{-1},
\end{aligned} \tag{1.10}
$$

where M_*, R_* and L_* denote the mass, radius and luminosity of the star, respectively (Stahler and Palla, 2004). A solar mass star, thus, takes about 30 Myr to contract to the main sequence. Eq. 1.10 and Figure 1.1 also show that the contraction time is strongly anti-proportional to stellar mass, i.e. low-mass stars form very slowly, taking upwards of $10^8 - 10^9$ yr, whereas massive stars form rapidly (Bodenheimer, 2011).

1.1.4 Massive Star Formation

Very massive stars ($M \gtrsim 6 - 10 \, M_\odot$; Bodenheimer, 2011; Stahler and Palla, 2004) contract to the main-sequence so rapidly in fact that they begin to burn hydrogen before ending their main accretion phase (Krumholz, 2020; Bodenheimer, 2011; Stahler and Palla, 2004). Consequently, they become visible in the optical only after they have already joined the main-sequence. Because of this, observing massive star formation is difficult and many details of the process are not entirely clear yet.

One primary challenge in understanding massive star formation is the fact that these stars are very hot and upwards of $10 \, M_\odot$ emit significantly in the UV regime (Klessen and Glover, 2016). Consequently, they inject a significant amount of (ionising) radiation into their natal envelope, which could halt accretion all-together in a purely spherical accretion picture. As we know stars as massive as $200 - 300 \, M_\odot$ (Hainich et al., 2014; Bestenlehner et al., 2011; Crowther et al., 2010), mechanisms must exist to circumvent the significant radiation feedback to form such massive stars. Proposed scenarios for massive star formation are e.g. that massive stars actually form via continuous collisions of lower mass protostars in highly clustered environments (see also Section 1.1.5), although this requires extremely high

densities of protostars, which are not observed in but the most extreme star-forming environments (Klessen and Glover, 2016). Alternatively, a significantly scaled-up version of the low-mass protostellar disc accretion picture appears plausible, including even more extreme outflows and winds, as the strong UV radiation could be emitted effectively along the polar-axes of such systems while the accretion flows go from the envelope through the discs (Klessen and Glover, 2016).

Aside from the radiation feedback considerations, it is also not entirely clear from where the progenitors of massive stars accrete all of their mass, if e.g. massive stars simply form from a very massive MC core progenitor in a *monolithic collapse* or if other mechanisms are at play. For instance, it has also been proposed that massive stars form by *competitive accretion*. In this scenario the collapse of larger, more massive regions of thousands of $M_\odot$ (e.g. MC clumps) is the driving force of star formation instead of the collapse of individual clumps. Via fragmentation these collapsing regions then give rise to smaller scale collapse, forming many lower mass protostars. The latter subsequently compete among each other to accrete material from the overall collapsing region, in which they are embedded (Bodenheimer, 2011).

Lastly, the mechanisms that limit the maximum mass for massive stars (and the mass limit itself) are also yet to be determined. Possible candidates are the radiative stellar feedback, an internal stability limit for stars with non-zero metallicity or fragmentation limiting the protostellar mass growth (Klessen and Glover, 2016).

1.1.5 Stellar Clusters

So far we have discussed star formation primarily on an individual object basis. In reality stars, in particular massive ones, form rarely in isolation but in *clusters* instead. A star cluster is a gravitationally bound collection of numerous stars. They are not to be confused with stellar associations, which are unbound, or multiple star systems, which are gravitationally bound but consist of only few stars ($\lesssim 10$). Contrary to galaxies, which are also

gravitationally bound systems, stellar clusters are not dominated by dark matter (Krause et al., 2020; Krumholz et al., 2019).

The formation of stars in clusters likely follows the following process. A larger dense molecular region of some 10^2 to 10^3 $M_\odot$ becomes Jeans unstable and starts to contract under its own gravity. Due to its highly turbulent interior, this region exhibits considerable substructure with high contrasts in density. During the overall contraction of the region, higher density substructures become themselves Jeans unstable, start to collapse and form accreting protostars on free-fall timescales much shorter than that of the global collapse. Protostars produced close to the global minimum of the gravitational potential may accrete large amounts of mass, being fed by the overall infall of the region (Klessen and Glover, 2016), whereas low-mass protostars in outskirt regions may only accrete material from their immediate surroundings (cf. competitive accretion).

In these clustered environments the protostars interact as they accrete and contract to become stars, be it through potential collisions (in very high protostellar density environments) or stellar feedback (as soon as the first massive stars are born). The ionising radiation of the first massive stars for instance quickly carves out bubbles of ionised hydrogen (i.e. H II regions) into the contracting region, disrupting or even completely dispersing the collapsing material, while the lower mass stars are still forming. In clustered star formation, the interaction between protostars, thus, is another important factor that may affect how much mass certain stars can ultimately attain (Klessen and Glover, 2016). Although stellar feedback of the young cluster ultimately disperses the molecular cloud, it is actually also a factor that possibly triggers star formation, as the gas entrained by the ionising radiation may collide with material in the surrounding GMC environment to form local overdensities that in turn collapse and form new stars (e.g. Walch et al., 2013). In that vein feedback from the first formed stars may induce the formation of multiple populations in a cluster environment.

Another interesting implication that follows from the cluster formation scenario outlined above is *mass segregation* in young clusters. As only the protostars close to the global gravitational potential minimum (i.e. cluster

centre) can maintain large accretion rates, massive stars should, thus, accumulate predominantly in the centre of the final cluster, whereas lower mass stars would be more likely to occupy the cluster's outskirts (Klessen and Glover, 2016). Clusters mass-segregated in this fashion have been observed (e.g. Zeidler et al., 2017), but it is not necessarily clear whether this mass segregation is truly primordial (i.e. a direct consequence of the cluster formation) or a result of the cluster's dynamical evolution. As a cluster dynamically relaxes, kinetic energies of the cluster constituents equalise, such that massive stars attain lower velocities, settling deeper in the gravitational potential well (Krause et al., 2020). It is also still debated whether the competitive accretion scenario truly forms the most massive stars in the centre of a cloud, as shielding mechanisms have been proposed that actually starve the most central regions of gas (fragmentation-induced starvation; Peters et al., 2010; Girichidis et al., 2012), thus preventing high accretion rates for protostars localised there (Klessen and Glover, 2016). Consequently, the topic of mass segregation is still actively researched and may provide crucial insights into the formation scenarios of stellar clusters.

1.1.6 Initial Mass Function

One of the most central relations in star formation research and the observation of young stellar clusters is the *initial mass function* (IMF). The IMF $\Phi(M)$ describes the distribution of stellar masses at birth, i.e. $dN = \Phi(M)dM$ indicates how many stars form with a mass in the range M and $M + dM$ (Beech, 2019). Theoretically explaining the observed IMF is not only a central and still actively studied problem in star formation research, but the IMF also plays an important role in many other disciplines from stellar to galactic astrophysics. For instance, constraining physical properties from the observed light of unresolved stellar populations in distant galaxies or modelling galaxy formation relies heavily on the assumed form of the IMF (Krumholz, 2020).

The observed IMF can be reasonably described by a broken power law, first introduced by Salpeter (1955) and later extended by Kroupa (2001), or a
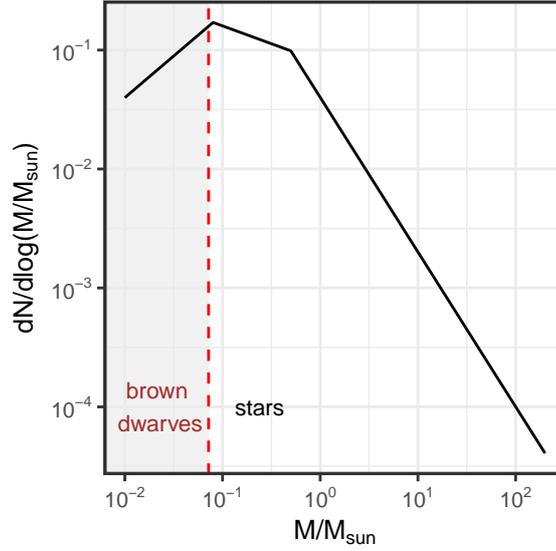
Fig. 1.2.: Diagram of the Kroupa (2001) IMF in terms of logarithm in mass. The presented curve ranges from 0.01 to $200\,M_\odot$ and the integral of the IMF is normalised to one over this range. The red dashed line and grey shaded region indicate the brown dwarf regime.

lognormal distribution (Chabrier, 2003). The Kroupa (2001) IMF has the following forms in terms of number in mass

$$
\frac{\mathrm{d}N}{\mathrm{d}M} = \Phi(M) \propto
\begin{cases}
M^{-0.3} & \text{for} \quad M < 0.08\,M_\odot \\
M^{-1.3} & \text{for} \quad 0.08\,M_\odot < M < 0.5\,M_\odot, \\
M^{-2.3} & \text{for} \quad M > 0.5\,M_\odot
\end{cases}
\tag{1.11}
$$

and in terms of logarithm in mass

$$
\frac{\mathrm{d}N}{\mathrm{d}\log(M)} = \xi(M) \propto
\begin{cases}
M^{0.7} & \text{for} \quad M < 0.08\,M_\odot \\
M^{-0.3} & \text{for} \quad 0.08\,M_\odot < M < 0.5\,M_\odot. \\
M^{-1.3} & \text{for} \quad M > 0.5\,M_\odot
\end{cases}
\tag{1.12}
$$

The form (see Figure 1.2) of the observed IMF has two important implications. Peaking in the range $0.1 - 0.5\,M_\odot$, we find that the formation of low-mass stars (i.e. M dwarfs) is preferred. On the other hand, although accounting for a large fraction of the overall stellar feedback, massive stars do not only form quickly and die young, but are also rarely born to begin with (Beech, 2019).
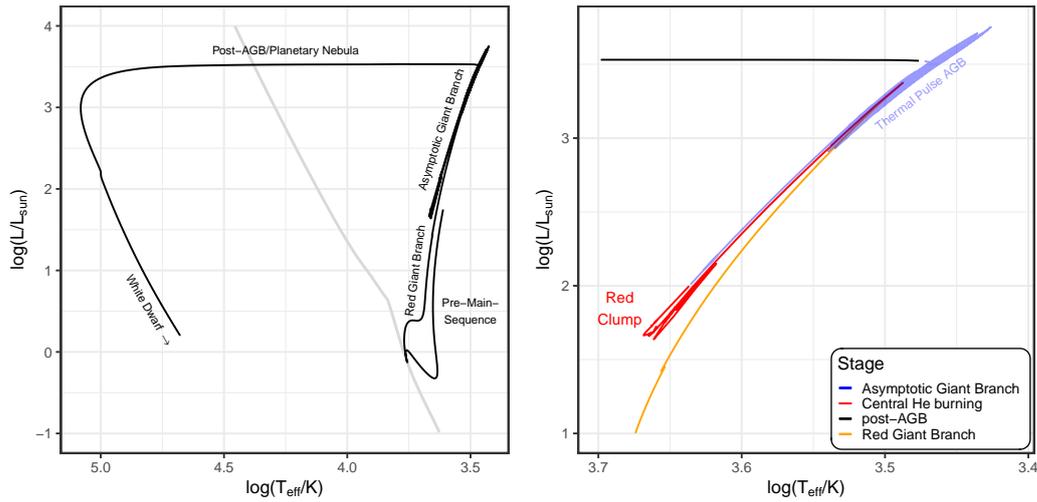
Fig. 1.3.: HRD showing the full (MIST; Dotter, 2016; Choi et al., 2016) evolutionary track for a $1\,M_\odot$ star from the beginning of the PMS to the end of its life as a white dwarf (left). Indicated in grey is the ZAMS. Right: Zoom-in on the post-main-sequence portion of the evolutionary track, highlighting the upper red giant branch, central helium burning phase and asymptotic giant branch.

From a theoretical standpoint the origin the IMF is still actively researched. The core mass function, i.e. the initial mass distribution of molecular cloud cores, appears strikingly similar in shape to the stellar IMF (although scaled to higher masses by a factor of ~ 3), so that it has been proposed that the IMF is simply a direct result of fragmentation in (giant) molecular clouds (Klessen and Glover, 2016). Another still-puzzling observation for star formation theory is the apparent universality of the IMF observed across a wide variety of star forming environments (Beech, 2019). Naively one would expect the IMF to directly depend on the individual natal environments of stars. For more detailed discussions of potential origins of the IMF and its universality we refer to e.g. Kroupa et al. (2013); Offner et al. (2014); Krumholz (2020).

1.2 Post-Main-Sequence Evolution

In this section we shall briefly outline the main evolutionary phases of stars after they end their lifetime on the main-sequence. For more details see e.g. Kippenhahn et al. (2012) or Srinivasan (2014).

Once a star has joined the main-sequence, it remains there as long as it is burning hydrogen in its core. Since the mass of a star and, thus, the available fuel for nuclear fusion are limited, it cannot burn hydrogen indefinitely, however. Instead, it exhausts its central hydrogen supply on the *nuclear timescale* t_{nuc},

$$t_{\mathrm{nuc}} = \frac{E_{\mathrm{tot}}}{L_*} = \frac{0.007 f_{\mathrm{H}} X M_* c^2}{L_*}, \tag{1.13}$$

where L_* and M_* denote the stars luminosity and mass, X is the star's mass fraction in hydrogen (typically $\sim 70\%$), and f_{H} indicates the fraction of hydrogen that is consumed over the main-sequence lifetime (Stahler and Palla, 2004). Using an estimated value of $f_{\mathrm{H}} \approx 0.1$, the main-sequence lifetime is

$$t_{\mathrm{nuc}} \approx 1 \times 10^{10} \left(\frac{M_*}{M_\odot} \right) \left(\frac{L_\odot}{L_*} \right) \ \mathrm{yr}. \tag{1.14}$$

Therefore, our Sun is expected to spend a total of 10 Gyr on the main-sequence, whereas a massive O-type star ($M_* \sim 50 \, M_\odot$, $L_* \sim 10^{5.55} L_\odot$) exhausts its fuel in a mere ~ 1.4 Myr (Stahler and Palla, 2004).

Figure 1.3 outlines the post-main-sequence evolution of a solar mass star in the HRD. Once the star has exhausted its central hydrogen supply, it is left with an inert helium core that is too cold to ignite any further fusion reactions. This core is surrounded by a shell that is still actively burning hydrogen and continuously feeds more helium to the inner core. Without the energy generated by the fusion reaction the helium core subsequently starts to contract under its own gravity as it is gaining mass. Through the energy release of this gravitational contraction, the envelope of the core begins to rapidly expand and the star moves to the right away from the main sequence in the HRD. As the energy released by the contracting core and the hydrogen-burning shell raises the star's luminosity by several orders of magnitude, the star then ascends the so-called *red giant branch* (RGB), while its envelope is steadily expanding (Stahler and Palla, 2004). Eventually the helium core reaches temperatures high enough to ignite the fusion of helium, producing carbon and oxygen. When this happens the star moves once again to the left in the HRD (see Figure 1.3, right panel). The central helium burning phase is a relatively stable state, and in the case of a $1 \, M_\odot$ star it lasts for $\sim 10^8$ yr.

Afterwards, the central helium supply, too, is exhausted and the star, again with an inert contracting core (consisting of C/O), now two burning-shells (one helium, one hydrogen) and expanding envelope, begins to ascend the *asymptotic giant branch* (AGB) towards the top right of the HRD (Stahler and Palla, 2004). As the star moves up along the AGB the two burning shells become thermally unstable and enter an alternating cycle, where either the helium- or hydrogen-burning shell dominates the star's luminosity, while the other ceases fusion. In these so-called *thermal pulses* (TP) the luminosity of the star can rise significantly for short periods of time, as the helium shell explosively reignites in events known as *helium shell flashes* (Srinivasan, 2014; Kippenhahn et al., 2012). The AGB and TP-AGB phases of the post-main-sequence evolution are accompanied by massive winds that eject substantial amounts of mass from the star. A star ends its AGB phase and enters the post-AGB evolution as soon as the winds have dispersed most of the stars envelope revealing the inert, contracting C/O core. The core now moves horizontally to the left in the HRD, potentially ionising the remnants of its envelope in the vicinity creating a "planetary nebula", until its remaining fuel for shell-burning runs out and it ends its life as a slowly-cooling *white dwarf*, supported against collapse by electron-degeneracy pressure (Kippenhahn et al., 2012; Stahler and Palla, 2004).

Contrary to the low and intermediate-mass stars, the very massive stars ($M \gtrsim 8\,M_\odot$) can maintain nuclear fusion reactions beyond the formation of a carbon-oxygen core. They repeat the cycle of central burning, fuel exhaustion, shell burning, core contraction and heating, and subsequent fusion ignition all the way up to an iron core, when no more energy can be released by fusion reactions. At this point, a massive star exhibits an onion-like structure with many shells engaged in different active shell-burning processes. Ultimately, the iron core collapses, forming a neutron star or black hole and ejecting its envelope in a violent explosion, a *supernova*, whose luminosity may (briefly) eclipse that of entire galaxies (Kippenhahn et al., 2012; Srinivasan, 2014; Stahler and Palla, 2004).

1.3 Observing Star Formation

In this section we will briefly discuss characteristic observational properties of still-forming stars, outline a common observation method and corresponding analysis tool, and introduce a few tracers of star formation.

Observational YSO classification and properties

We begin our discussion of observational characteristics of young stars by introducing a widely used classification scheme for young stellar objects (YSOs) that is based on their spectral energy distributions (SEDs). YSO is a common term used to describe the entire system of a still-forming star (i.e. central source + disc + envelope) regardless of the evolutionary state (Schulz, 2012). In this classification scheme YSOs are distinguished into four classes 0, I, II, and III.

Class 0 are YSOs that are likely in the early stages of the protostellar collapse, where most of the mass of the system is still located within the envelope rather than the central object. In this stage the YSO is optically thick and fully absorbs the accretion luminosity of the material that is falling onto the protostar. Consequently, the SED of class 0 systems consists of reprocessed emission of the overall cold (dusty) envelope, which falls primarily in the sub-millimetre regime, rendering the object effectively invisible at shorter wavelengths (Klessen and Glover, 2016; Schulz, 2012). As a protostars accumulates more and more material, forming an accretion disc, it ultimately thins out its infalling envelope through accretion and outflows (jets and winds). This process gradually reveals the inner regions of the accretion disc, shifting the protostellar SED towards shorter wavelengths. The YSO, thus, becomes observable in the infrared and is, at this point, identified as a *class I* source (Klessen and Glover, 2016; Schulz, 2012). Note that, although the term YSO is applicable to all evolutionary phases of forming stars, it is most commonly used to refer to class 0 and I sources and in particular the most massive protostellar systems (Schulz, 2012).

Once most of the envelope, but not the accretion disc, is dispersed and the central protostar becomes directly visible for the first time, i.e. the YSO is entering the PMS phase, it is observationally identified as a *class II* object. Low mass ($M \lesssim 2\,M_\odot$) objects of this type are also called *classical T Tauri stars* (CTTS). Prototyped by T Tau in the Taurus cloud, CTTS are characterised by effective surface temperatures in the range of 3000 to 7000 K and F to M spectral types. Their SEDs peak in the $1 - 10\,\mu\mathrm{m}$ range and show notable excess emission in the infrared regime, which is evidence for the presence of (flared) accretion discs. Aside from exhibiting strong $\mathrm{H}\alpha$ (656.3 nm) hydrogen line emission and X-ray emission, CTTS are also observed to be highly variable, with potential bursts of luminosity lasting several decades (FU Orionis objects), which has been accredited to episodes of increased accretion (Schulz, 2012; Stahler and Palla, 2004). The intermediate-mass counterparts ($2\,M_\odot \lesssim M \lesssim 8\,M_\odot$) of the CTTS are called *Herbig Ae/Be* stars. With effective surface temperatures in the range of $8,000$–$20,000$ K and spectral types between B0 and F, they, similar to the CTTS, are characterised by strong emission lines, in particular in $\mathrm{H}\alpha$, and excess IR emission (for more details see e.g. Schulz, 2012; Stahler and Palla, 2004).

Lastly, *class III* objects are PMS stars that have exhausted most of their circumstellar discs, leaving only a remnant debris disc, and appear to be no longer actively accreting new material, while they undergo their final contraction towards the main sequence. Observationally, low-mass class III objects are also referred to as *weak-lined T Tauri stars* (WTTS). Although they are still strong X-ray emitters, compared to CTTS the WTTS exhibit only weak emission in $\mathrm{H}\alpha$ and a vanishing IR excess (Schulz, 2012).

Photometry

One of the primary observational techniques to study star formation (and stars in general) is *photometry*. Contrary to spectroscopy, where the entire spectrum of a star is measured, in photometry a stellar spectrum is essentially observed piece-wise. Using (broad and narrow-band) filters that dictate a response function $R(\lambda)$ of the observing apparatus (i.e. the fraction of incident energy flux registered at wavelength λ), photometric observations consist of

measuring the integrated flux F_{fltr} of a star over a limited wavelength interval (i.e. where $R(\lambda) > 0$) as

$$F_{\text{fltr}} = \int_0^\infty R(\lambda) f_\lambda \mathrm{d}\lambda, \qquad (1.15)$$

where f_λ indicates the monochromatic energy flux from the star (Chromey, 2010). The measured flux, i.e. the star's brightness, over the filter's wavelength range is commonly given in *magnitudes*. Here we distinguish between the *apparent* m_{fltr} and *absolute* M_{fltr} magnitude of an observed object. Because the flux density $F = L/(4\pi r^2)$ we receive from a star (and thus the measured F_{fltr}) depends on our distance r to it, the *apparent* magnitude is defined as

$$m_{\text{fltr}} = -2.5 \log\left(\frac{F_{\text{fltr}}(r)}{F_{\text{fltr},0}(r)} \right) = -2.5 \log(F_{\text{fltr}}(r)) - m_{\text{fltr},0}, \qquad (1.16)$$

where $F_{\text{fltr},0}$ denotes a reference flux marking the zero point for the magnitude system. Typically the latter is calibrated to the observed fluxes of the star Vega, i.e. $m_{\text{fltr,Vega}} = m_{\text{fltr},0} = 0$. In contrast, the *absolute* magnitude serves as a measure of the *intrinsic* brightness of an object and is defined as the apparent magnitude we would measure if a given object were located at a distance of 10 pc:

$$M_{\text{fltr}} = -2.5 \log(F_{\text{fltr}}(10\,\text{pc})) - m_{\text{fltr},0}. \qquad (1.17)$$

The apparent and absolute magnitude are related via the *distance modulus* $m_{\text{fltr}} - M_{\text{fltr}}$ by

$$m_{\text{fltr}} - M_{\text{fltr}} = 5 \log\left(\frac{r}{10\,\text{pc}} \right). \qquad (1.18)$$

Thus, if we know the intrinsic brightness of an observed star, i.e. its absolute magnitude, we can directly recover the distance to the object from its apparent magnitude (Stahler and Palla, 2004). Typically, photometric filters are either referred to by their central (effective) wavelength or by an abbreviation in a photometric filter system. Commonly used for instance is the Johnson-Cousins UBVRI system, summarised in Table 1.2 (Bessell, 2005).

Tab. 1.2.: Central effective wavelengths and widths of the Johnson-Cousin broad-band filter system as listed in Bessell (2005).

	U	B	V	R	I
λ_{eff} (nm)	366.3	436.1	544.8	640.7	789.0
$\Delta\lambda$ (nm)	65	89	84	158	154
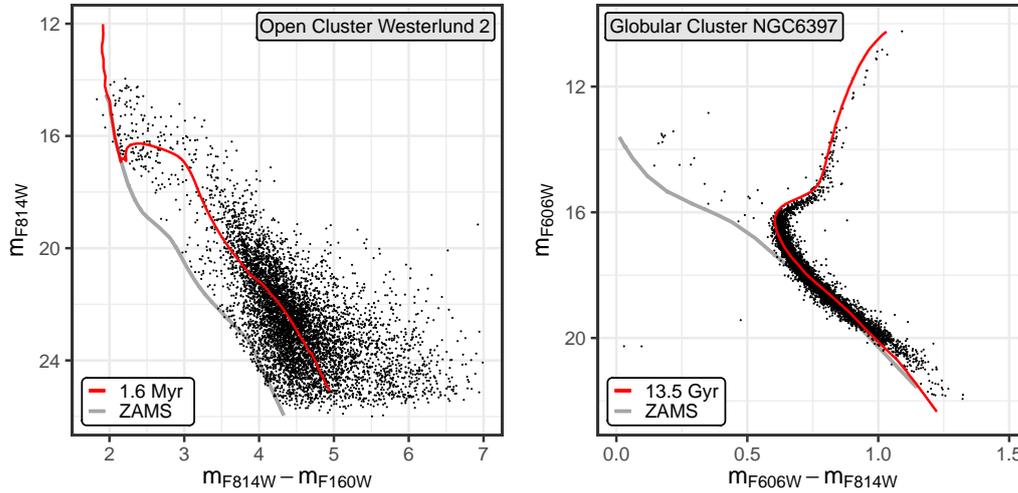
Fig. 1.4.: CMD of the open cluster Westerlund 2 (See Section 2.1 for details) in the Milky Way (left). Indicated in red is a 1.6 Myr PARSEC(Bressan et al., 2012; Chen et al., 2014, 2015; Tang et al., 2014) isochrone and the ZAMS in grey (solar metallicity, corrected for distance and average extinction of the cluster) for comparison. Right: CMD of the globular cluster NGC6397 (See Section 2.2). Indicated in red and grey are a 13.5 Gyr PARSEC isochrone and the ZAMS, respectively.

Colour Magnitude Diagrams and stellar clusters

Among the central outcomes of photometric surveys of e.g. star-forming regions and young stellar clusters is the colour-magnitude diagram (CMD). The CMD is the observational analogue to the theoretical HRD (i.e. the one relating luminosity to effective surface temperature). As a proxy for the luminosity of an object the CMD uses the measured apparent (or absolute if the distance to the object is known and can be corrected for) magnitude in one of the survey filters and relates it to the object's *colour index*. The colour

index is given as the difference in apparent magnitude between two filters with effective (central) wavelengths λ_1, λ_2 following

$$\mathrm{CI} = m_{\lambda_1} - m_{\lambda_2}, \tag{1.19}$$

where $\lambda_1 < \lambda_2$. Assuming that a star is mostly a blackbody emitter, the colour index is a direct measure of the form of the star's spectrum, i.e. indicates the overall slope of the spectrum between the two central filter wavelengths. Additionally, it measures the star's *effective temperature*. A hot (blue) star, with an emission maximum at shorter wavelength, has a small colour index as $m_{\lambda_1} < m_{\lambda_2}$, whereas a colder (red) star with peak emission at longer wavelength has a large colour index since $m_{\lambda_1} > m_{\lambda_2}$ (Chromey, 2010). Often the colour index is denoted via the filter names, e.g. $B - V$ indicates the colour index $m_{\mathrm{B}} - m_{\mathrm{V}}$ in the Johnson-Cousin filter system (cf. Table 1.2).

In the Milky Way, gravitationally bound stellar clusters are identified as either *open* or *globular* clusters. Open clusters are located in the Galactic disc, young ($\lesssim 1$ Gyr) and generally less massive ($M_{\mathrm{tot}} \lesssim 10^5\,M_\odot$), whereas globular clusters are found in the Galactic bulge and halo, generally old (> 1 Gyr), but massive ($M_{\mathrm{tot}} \gtrsim 10^4\,M_\odot$). The latter are believed to be survivors from the earliest phases of star formation in the Universe and some are estimated to be so old (> 13 Gyr) that they provide constraints for the age of the Universe itself (Krause et al., 2020). CMDs from photometric surveys provide an important tool to discern intrinsic properties of stellar clusters. CMDs show for instance that open clusters are mostly coeval single populations with a single main-sequence, whereas CMDs of globular clusters most of the time exhibit multiple main-sequences, indicating the presence of multiple stellar populations (of different metallicity; Krause et al., 2020).

CMDs can also be used to estimate a cluster's age when combined with stellar evolutionary models. This is done by constructing *isochrones*, i.e. loci of *constant age* across the full mass range (e.g. $0.08 - 300\,M_\odot$) of evolutionary tracks, and comparing these to the observed cluster CMDs. Figure 1.4 shows examples of a (rough) age estimate using PARSEC isochrones (Bressan et al., 2012; Chen et al., 2014, 2015; Tang et al., 2014) for a young open cluster, Westerlund 2, in the left panel and for an old globular cluster, NGC 6397,

on the right. For more details on these two clusters see Sections 2.1 and 2.2, respectively. Note that NGC 6397 does *not* show evidence of multiple stellar populations in the presented data, and that the Westerlund 2 CMD is missing some of the most massive stars due to saturation effects.

We can estimate a cluster's age using two distinct features in the CMD. At a given age of a young cluster only those constituents exceeding a certain mass threshold (cf. Kelvin-Helmholtz timescale Eq. (1.10)) have already evolved onto the main-sequence, whereas lower-mass stars are still in their PMS evolutionary phase. This introduces an age-characteristic bend in the CMD of a young cluster called the *main-sequence turn-on* and is determined by the lowest-mass cluster main-sequence star. By matching this bend of a synthetic isochrone to the observed one in the CMD, as done in the left panel of Figure 1.4, we can, thus, estimate a cluster's age (Gouliermis, 2012; Stahler and Palla, 2004). For an old cluster we can use the same approach using the second characteristic feature in the CMD, the *main-sequence turn-off*. The turn-off is defined by the minimum mass of the stars in a cluster that are just *departing* from the main sequence, i.e. all cluster stars more massive than this threshold are already in their post-main-sequence evolution given the cluster's age (Gouliermis, 2012; Stahler and Palla, 2004). This turn-off, too, induces a bend in the CMD of the cluster, as demonstrated by the old NGC 6397 in the right panel of Figure 1.4, that can be matched to an isochrone to determine the age of the cluster. Shown here is only a simple approximation to demonstrate the idea of the approach, real applications use, of course, more sophisticated isochrone fitting techniques to deal with e.g. photometric uncertainties and potential age gradients in a cluster's population.

Using the isochrones derived from stellar evolutionary models, we can also highlight one of the difficulties that occur when studying young stellar clusters and PMS stars. Figure 1.5 compares a series of young (PMS) isochrones between 1 and 10 Myr to very old isochrones in the range of 0.5–10 Gyr in the filter combination of the Westerlund 2 example from Figure 1.4. In this diagram we find that the post-main-sequence portions of the old isochrones (blue) *overlap* with the PMS parts of the young isochrones (black). Suppose we conduct an observation of a young stellar cluster. If now an old population
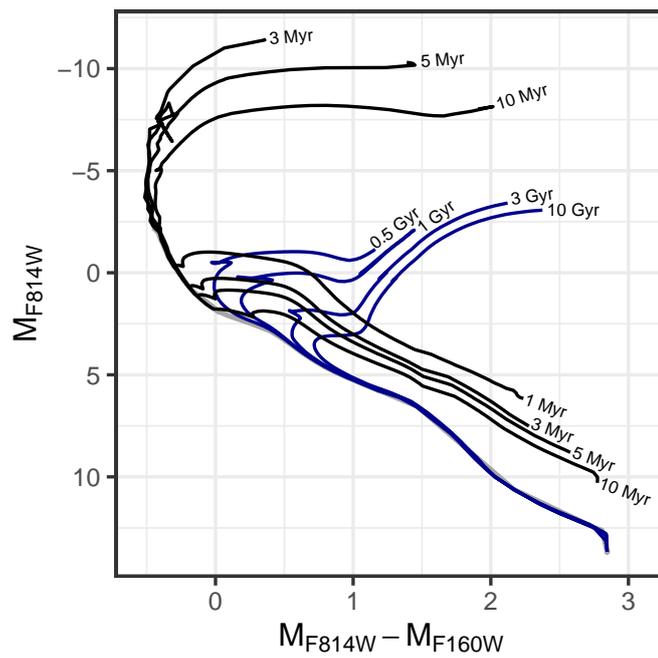
Fig. 1.5.: Synthetic CMD of PARSEC (Bressan et al., 2012; Chen et al., 2014, 2015; Tang et al., 2014) isochrones with ages between 1–10 Myr (black) and 0.5–10 Gyr (blue). The filter combination matches the left panel of Figure 1.4. For comparison the grey line marks the ZAMS.

(e.g. the field) falls within the same line-of-sight as the cluster (either in the foreground or background), and no secondary measures to determine cluster membership (e.g. individual distance measures) are available, then this overlap of isochrones indicates that we could confuse an old post-main-sequence star for a PMS source. This inherent degeneracy (in L, T_{eff}) is among the subjects of our study in Section 3.2.

Extinction and the red clump

When observing stars and stellar clusters one of the major difficulties we face is interstellar *extinction*, i.e. the fact that any dust (or gas) between us and the source will either absorb or scatter the light emitted by the object, dimming it in the process. In photometry we can formulate this effect by modifying the relation in Eq. (1.18) between apparent and absolute magnitude with an additional extinction term as follows

$$m_\lambda = M_\lambda + 5\log\left(\frac{r}{10\,\mathrm{pc}}\right) + A_\lambda, \tag{1.20}$$

where, for simplicity, we have assumed a perfect monochromatic filter of wavelength λ here and A_λ denotes the extinction at that wavelength (Stahler and Palla, 2004). The extinction value A_λ is strongly dependent on λ and for dust tends to decrease with longer wavelength, i.e. interstellar dust absorbs UV light much more strongly than infrared emission. From Eq. (1.20) it is also immediately evident that extinction alters the observed colour index of a source. This is quantified by the *colour excess*, defined as (for e.g. the B and V band) the difference

$$E(B-V) = (B-V) - (B-V)_0 \tag{1.21}$$

between the observed colour $(B-V)$ and the intrinsic colour $(B-V)_0$ of the source. Inserting Eq. (1.20) into the definition (1.21) we find

$$E(B-V) = A_{\mathrm{B}} - A_{\mathrm{V}}. \tag{1.22}$$
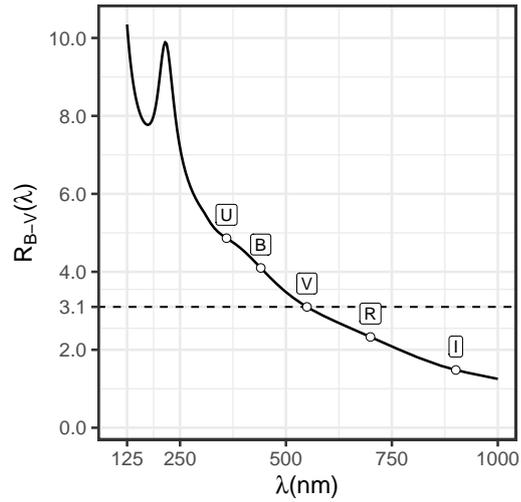
Fig. 1.6.: Extinction law $R_{B-V}(\lambda)$ for the diffuse ISM in the Milky Way between 125 nm and 1000 nm as given by the Cardelli et al. (1989) model for $R_V = A_V/(B - V) = 3.1$. The marked points give $R_{B-V}(\lambda)$ for the Johnson-Cousin photometric system introduced in Table 1.2.

Using the extinction and colour excess we can also define the normalised *total extinction*

$$R \equiv \frac{A_V}{E(B - V)}, \tag{1.23}$$

which also denotes the slope of the reddening vector in a CMD, i.e. the direction in which extinction dislocates a star's position away from its true CMD location. Measuring R over all wavelengths in dependence of the colour excess (e.g. $E(B - V)$) between two filter bands, i.e.

$$R_{B-V}(\lambda) = \frac{A_\lambda}{E(B - V)}, \tag{1.24}$$

recovers the *extinction* or *reddening law* of the sight-line towards the object (Stahler and Palla, 2004). As an example, Figure 1.6 shows the extinction law $R_{B-V}(\lambda)$ for the diffuse interstellar medium in the Milky Way as given by the Cardelli et al. (1989) model for $R_V = A_V/(B - V) = 3.1$, highlighting the decrease in extinction with increasing wavelength.

One way to constrain the reddening law within an observed field of view is to make use of a distinct CMD feature called *red clump* (RC). As we have discussed in Section 1.2, stars will enter a phase of central helium burning
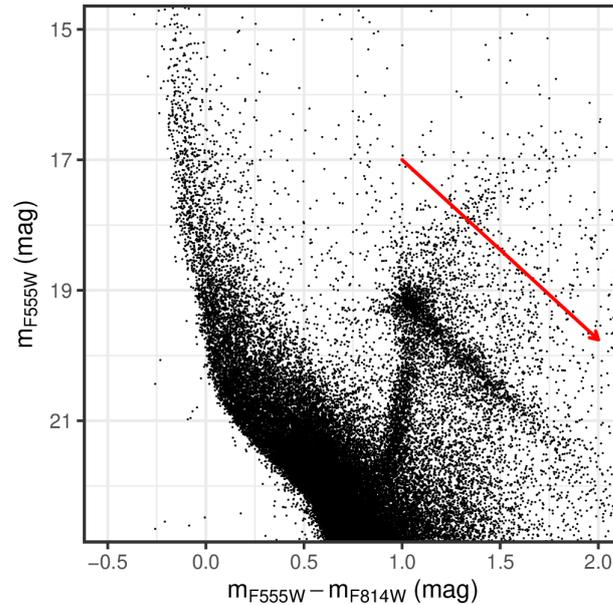
Fig. 1.7.: CMD of the MYSST survey (see Sections 2.5, 3.3, 3.4) of the star-forming region N44 in the Large Magellanic Cloud. Highlighted is the extinction-elongated red clump in comparison to the reddening vector (red arrow).

at some point during their post-main-sequence evolution. In particular for low-mass stars ($M_* \lesssim 2\,M_\odot$) this phase of central helium burning is both a fairly stable and long-lived state ($\sim 10^8$ yr), such that observing post-main-sequence stars in this phase is not unlikely (Stahler and Palla, 2004). Additionally, the mass of the inert helium core at the point of helium ignition is almost equal for most low-mass stars *independent* of their total mass. Consequently, over a mass range of (at least) $0.5\,M_\odot \lesssim M_* \lesssim 1.7 - 2.5\,M_\odot$ these low-mass stars actually exhibit a very similar luminosity and effective temperature, except for relatively small deviations caused by different envelope masses and chemical compositions (Girardi, 2016). Because of this, low-mass helium-burning stars appear as a distinct clump in the CMD when we conduct observations that capture the old field population of galaxies. This is the red clump. For a more extensive review of the properties of the RC see e.g. Girardi (2016).

Being an easily identifiable feature in the CMD and a fairly (intrinsically) homogeneous set of objects, the RC can be used to determine individual extinction measures under the assumption that, given this intrinsic homogeneity, any observed colour-magnitude spread can be attributed to (differential) ex-

tinction (Girardi, 2016). In particular, the RC feature is often found to spread along the reddening vector owing to extinction in observations of external galaxies (e.g. the Large Magellanic Cloud, see Section 2.3). Figure 1.7 shows an example of this extinction-elongation of the RC feature in the CMD in comparison to the reddening vector (red arrow) from the MYSST survey of star-forming region N44 in the Large Magellanic Cloud (see Sections 2.5, 3.3 and 3.4 for details on the survey and N44). Measuring the slope of the extinction-elongated RC, we can, thus, determine the reddening vector R and even constrain the reddening law $R(\lambda)$, if we have access to a panchromatic survey (i.e. a wide range of photometric filters). For example applications of this approach see e.g. De Marchi and Panagia (2014); De Marchi et al. (2016).

Tracers of Star Formation

Observing young stellar populations and PMS stars directly is the straightforward way to investigate star formation. There are also other observational tracers that indicate potential future, on-going or recent star formation. In this section we give a few examples (but not an exhaustive list) of such alternative tracers, which are in particular considered in our study in Section 3.4.

Being the very material stars form out of, molecular clouds are certainly the first objects to look for. However, observing molecular hydrogen is actually not a trivial matter. Being a homonuclear, symmetric (no permanent dipole moment), and low-mass molecule, H_2 basically never occupies excited states at the temperatures of GMCs (except for shock heated regions, which overall are rare), and is, thus, practically incapable of emitting. The only other way to directly observe H_2 in GMCs is through UV absorption lines. However, this requires a bright UV background source, which is generally not a common occurrence (Krumholz, 2020; Klessen and Glover, 2016).

GMCs are not only host to H_2, but to other molecular species as well, which can be used as observational proxies. The most common choice is carbon monoxide CO, the second most abundant molecule in the ISM. Being a more massive, heteronuclear molecule with a strong dipole moment, the rotational transitions of CO can be excited even at the low temperatures in

GMCs (e.g. the CO $J = 1$ state lies at only $5.5K$ above the ground state; Krumholz, 2020). CO is, therefore, one of the main observational targets to identify GMCs and measure their masses. Additionally, together with its isotopologues ^{13}CO and $C^{18}O$, it can also be used to probe various density regimes of molecular clouds (^{12}CO for $n \sim 100 - 300\,\mathrm{cm}^{-3}$; ^{13}CO and $C^{18}O$ for $n \sim 10^3\,\mathrm{cm}^{-3}$, e.g. molecular clumps) up to about $n \sim 10^5\,\mathrm{cm}^{-3}$, where it then becomes optically thick and may freeze out (Bodenheimer, 2011).

Massive main-sequence stars (of the O-type to early B-type variety) are themselves signposts of recent and on-going star formation. Because their main-sequence lifetimes are so short (cf. Eq. (1.10)), they have to be young and likely close to their natal environment (where other stars may still be forming) when observed in this phase. Very massive stars ($M_* > 10\,M_\odot$) also emit a significant amount of ionising UV radiation, which impacts their natal gas clouds. Consequently, sites of massive star formation are often accompanied by H II regions, i.e. hydrogen from the parental cloud (or the remnants thereof) ionised by the stellar UV radiation (Krumholz, 2020; Klessen and Glover, 2016). Observing and identifying H II regions can, thus, directly lead to active star-forming centres. H II regions are best observed via hydrogen recombination lines, in particular the optical $H\alpha$ and $H\beta$ lines at 656.3 nm and 486.1 nm, corresponding to the $3 \to 2$ and $4 \to 2$ electronic transitions of atomic hydrogen. H II regions emit these atomic hydrogen lines because hydrogen ions in H II regions can recombine to excited states instead of the ground state, prompting a radiative decay back to the ground state accompanied by line emission (Krumholz et al., 2019). H II regions also exhibit metal recombination lines from either singly or multiply ionised species such as N^+ or O^{++} (Klessen and Glover, 2016).

Lastly, dust emission observations also play a role in the study of star formation. As discussed in Section 1.1, especially in the early phases of protostellar evolution the central object is almost completely obscured by its surrounding natal envelope. The only way to observe these sources is through the reprocessed emission from the dusty envelopes in the infrared and sub-mm wavelength regimes (Klessen and Glover, 2016). Additionally, at later stages the presence of accretion discs around young objects is connected to excess IR emission from the dust in the disc, again reprocessing radiation

from the central source. Consequently, mid- to far-IR observations play an integral part in the search for YSOs (see e.g. Chen et al., 2009; Carlson et al., 2012, for example YSO studies). In general bright far-IR emission has been connected to regions of active star formation by several studies (e.g. Skibba et al., 2012; Javadi et al., 2017), as the dusty remnants of the natal molecular clouds, too, absorb and reemit the radiation of young stars to longer wavelengths (see also Casey et al., 2014, for a review).

1.4 Machine Learning

Machine learning (ML) is a discipline in computer science that falls under the greater subjects of applied statistics and artificial intelligence (AI). Moving beyond AI approaches that employ fixed knowledge bases and strict logic rules to solve tasks and perform inference, ML focuses on algorithms that acquire knowledge by themselves by extracting information and patterns from raw data. In this fashion ML has allowed computers to successfully solve and make predictions for a large variety of tasks that require knowledge of the real world, where the more rigid knowledge-based AI approaches have previously failed (Goodfellow et al., 2016).

In this section we first provide a short overview of general concepts related to the construction and application of ML approaches. Afterwards, we describe three specific algorithms more in detail, namely the support vector machine, the random forest, and the random sample consensus, which are the main models of our studies in Sections 3.1, 3.3 and 3.4. Lastly, we briefly introduce density estimation and the expectation-maximisation algorithm. Note that an exhaustive discussion of all the ML and statistics concepts applied throughout the studies in Chapter 3 is beyond the scope of this thesis, so that we shall refer to the more detailed literature, e.g. Bishop (2009), Hastie et al. (2009), James et al. (2017), Goodfellow et al. (2016), for all ML methodologies that are not captured in the following.

1.4.1 General Concepts

ML algorithms learn from data. To best describe what this entails, let us consider the concepts of *tasks* T, a *performance measure* P and *experience* E (Goodfellow et al., 2016). Referencing a definition by Mitchell (1997), *learning* certain tasks T means that an ML algorithm improves its performance (as measured by P) at the given tasks T with experience E (the data).

Tasks

The task refers to the problem the ML algorithm is supposed to solve. Although the list of *specific* tasks is in principle infinite, they can be categorised. The task categories relevant to this thesis, i.e. the subjects of our studies in Chapter 3, are *classification*, *regression* and *density estimation*. One of the main differences between these categories is the way a corresponding ML algorithm processes its inputs for the task. A single input, referred to as an *example* or *observation*, is typically represented by a vector $\mathbf{x} \in \mathbb{R}^n$, where each entry x_i denotes an observed property, a *feature*, of the object to be processed (Goodfellow et al., 2016). Learning a task commonly requires an ML algorithm to produce a function $f(\mathbf{x}) = \mathbf{y}$ that maps an input observation to a desired output quantity $\mathbf{y}$.

In a classification task the ML algorithm has to categorise an input into one of k different classes, i.e. it has to find a function $f : \mathbb{R}^n \to \{1, \ldots, k\}$. Instead of a class label $y \in \{1, \ldots, k\}$ the input is often also mapped onto a vector of probabilities for the different classes. An example classification task is *object recognition*, i.e. identifying an object shown within an input image (Bishop, 2009; Goodfellow et al., 2016).

In a regression problem the ML algorithm is asked to predict a continuous numeric value given the input, i.e. it has to determine a mapping $f : \mathbb{R}^n \to \mathbb{R}$ or $f : \mathbb{R}^n \to \mathbb{R}^m$ in the multivariate case. There are many examples for regression tasks, in particular in the natural sciences, where we often

want to predict some underlying physical parameters of a system given measurements in an experiment (Bishop, 2009; Goodfellow et al., 2016).

In density estimation the task is to find a function $p_{\mathrm{model}} : \mathbb{R}^n \to \mathbb{R}$ such that $p_{\mathrm{model}}(\mathbf{x})$ represents a probability density distribution over the space that the examples $\mathbf{x}$ are drawn from. Specifically, this means that an algorithm needs to learn the structure of and distinct patterns in the data it has seen (Bishop, 2009; Goodfellow et al., 2016).

Experience

In general experience refers to the set of data that an ML algorithm sees while learning a task (also called *training*). In this context a data set is simply a collection of N examples, each consisting of n features. Although most ML algorithms experience a complete data set (of fixed size) during the training procedure (except on-line and reinforcement learning techniques; see e.g. Bishop, 2009; Goodfellow et al., 2016, for details), based on the way they operate on the data, they can be distinguished into two categories, *supervised learning* and *unsupervised learning* algorithms (Goodfellow et al., 2016).

For every example in the training data a supervised learning algorithm is provided with an additional corresponding *label* or *target*. In other words, during the learning procedure for every example $\mathbf{x}$ the algorithm is told what the output $\mathbf{y}$ of the function that it is constructing should be. Classification and regression are typical examples for tasks that are solved via supervised learning. On the other hand, an unsupervised learning algorithm has to find useful properties of the structure of the data that it is presented with, without any additional information. Unsupervised learning tasks include e.g. density estimation and clustering (Goodfellow et al., 2016).

Performance measure

To learn a specific task, an ML algorithm requires a performance measure, which quantifies how well the algorithm does at the task. The type of perfor-

mance measure depends primarily on the nature of the task. In classification for instance the *accuracy*, i.e. the fraction of correctly classified objects in a data set, is a typical choice (see also Appendix A4 of Ksoll et al., 2018, in Section 3.1 for more classification-specific performance measures). In a regression task on the other hand e.g. the *mean squared error* between predicted and true target value is employed. The performance measure plays an integral part in constructing an ML algorithm as it determines the quantity that is optimised during the training procedure (Goodfellow et al., 2016).

The true goal and challenge of any ML approach is the prediction on entirely new data, which has not been seen during the training process of the ML model. To ascertain a model's ability to *generalise* to new data, a second dataset (with known targets) is prepared, the *test* set, which is held-out during training. The performance of a fully-trained model on this unseen test set represents an estimate of the *generalisation error*. This concept works under the assumption that both the training and test set are drawn from the same data generating distribution, so that minimising the *training error* on the training set implies an improvement for the prediction on unseen data (Goodfellow et al., 2016). Consequently, the *test error* is larger than or equal to the training error.

Over-/underfitting and regularisation

This relates to two common issues occurring when training an ML approach, namely *overfitting* and *underfitting*. Overfitting refers to the problem that the discrepancy between training and test error becomes too large, i.e. the model does extremely well on the training data (likely by simply memorising it completely) but generalises poorly. On the other hand, underfitting describes the situation where training and test error may be similar, but the model does not achieve an acceptable performance on the training data, i.e. it did not fully learn the task yet (Goodfellow et al., 2016).

One technique to combat overfitting when constructing an ML model is called *regularisation*. Broadly speaking, it is defined as any modification to the learning procedure that aims at reducing the generalisation but not the

training error (Goodfellow et al., 2016). Consider for instance a simple linear regression model of the form

$$y = \mathbf{w}^T \mathbf{x} + \mathbf{b} \tag{1.25}$$

with weights $\mathbf{w}$ and bias b, which is trained by minimisation of the mean squared error (MSE). A commonly chosen regularisation approach for this setup is to extend the MSE performance measure with the L2 norm of the weights, i.e.

$$P(\mathbf{w}) = \text{MSE} + \lambda \mathbf{w}^T \mathbf{w}. \tag{1.26}$$

In this scenario the regularisation term $\Omega(\mathbf{w}) = \mathbf{w}^T \mathbf{w}$ introduces a preference for smaller weights, where the scaling factor λ indicates the severity of this preference. This particular regulariser is also referred to as *weight decay* in the context of neural networks (Bishop, 2009; Goodfellow et al., 2016).

Hyperparameters and cross-validation

Most ML algorithms entail two sets of parameters. The first set are parameters that are determined during the training process by optimising the given performance measure. The second set on the other hand comprises a form of *settings* that influence the algorithm's performance and the solution it derives, but are fixed before and during training. The latter are referred to as *hyperparameters*. The scaling factor λ in Eq. (1.26) from the linear regression example above is for instance one such hyperparameter.

The proper choice of hyperparameters is often integral to the success of an ML model. To determine the hyperparameters a second data set, which the model does not get to see during training, is employed, the so-called *validation* set. The validation set is not to be confused with the actual test set and it is important that they are distinct from each other. The test set may only provide a reasonable estimate of the generalisation error if it is neither involved in the fit- nor hyperparameter determination of the model. In most cases the validation set is created by randomly selecting a subset of the total available training data. The hyperparameters are then determined by e.g. performing a random search in the hyperparameter space, i.e. a series

of random sets of hyperparameters is generated, for each set the model is trained on the remaining training data and afterwards evaluated on the validation set. The model and set of hyperparmaters that best perform on the validation set are then chosen as the final model (Goodfellow et al., 2016).

If only little total training data is available, so that a randomly selected validation set is very small, estimating hyperparameters via validation will suffer from large statistical uncertainty. In this case a more efficient procedure to determine hyperparameters is *k-fold cross-validation*. In this approach the available training data is first split into k non-overlapping random subsets. The validation error is then estimated as an average over k different realisations of the model. Specifically, in realisation i the ith subset of the data is used as the validation set and the model trained on the remaining $k - 1$ subsets (Goodfellow et al., 2016).

1.4.2 Support Vector Machine

The support vector machine (Cortes and Vapnik, 1995) is a widely-used and fairly successful classifier for many classification problems. Making use of the *kernel trick* to implicitly map a given classification task to a higher dimensional feature space and determining an optimal class-separating hyperplane there, the SVM quite efficiently computes non-linear decision boundaries in the original feature space. It is a generalisation of the *support vector* and *maximal margin* classifiers (James et al., 2017; Hastie et al., 2009).

Maximal Margin and Support Vector Classifier

The basic idea of the maximal margin classifier is that of the *optimal separating hyperplane*, i.e. finding a flat affine $n - 1$ dimensional subspace of a n-dimensional space (e.g. a 2D-plane in a 3D space) that describes the solution to a set of linear equations

$$\mathbf{x}^T \beta + \beta_0 = 0, \tag{1.27}$$

where β denotes a vector of coefficients and β_0 is a constant. In a classification task the optimal separating hyperplane is constructed such that it perfectly separates all N training examples $\mathbf{x}_i$ according to their class labels y_i ($y_i \in \{-1, 1\}$), i.e.

$$y_i(\mathbf{x}_i^T \beta + \beta_0) > 0 \quad \forall i = 1 \ldots N. \tag{1.28}$$

Classifying a new observation $\hat{\mathbf{x}}$ is then straightforward, as the class can simply be determined by the side of the hyperplane $\hat{\mathbf{x}}$ falls on (James et al., 2017), i.e. computing the sign

$$G(\hat{\mathbf{x}}) = \mathrm{sign}[\hat{\mathbf{x}}^T \beta + \beta_0]. \tag{1.29}$$

The *margin* denotes the minimum (perpendicular) distance of all training examples $\mathbf{x_i}$ to a given hyperplane. Therefore, the *maximal margin classifier* is defined as the classifier following Eq. (1.29) that uses the hyperplane with the largest margin, i.e. the *optimal separating*/*maximal margin* hyperplane.

Training examples $\mathbf{x}_i$ that lie exactly on the margin are referred to as *support vectors* as their position essentially defines ("supports") the hyperplane. To construct the maximal margin classifier the hyperplane satisfying the following optimisation problem has to be found:

$$\max_{\beta, \beta_0, ||\beta||=1} M, \tag{1.30}$$

subject to

$$y_i(\beta_0 + \mathbf{x}_i^T \beta) \geq M \quad \forall i = 1, ..., n, \tag{1.31}$$

where M, $M > 0$, represents the width of the margin.

The *support vector classifier* generalises the maximal margin approach to cases where the training examples $\mathbf{x}_i$ are not linearly separable, i.e. no solution to Eqs. (1.30), (1.31) with $M > 0$ exists. To achieve this, the concept of a *soft margin* is introduced. Rather than determining the hyperplane that *perfectly* separates the training data, the support vector classifier aims to find the hyperplane that *best* separates the data, while allowing for a few

training examples to fall on the incorrect side of the margin or hyperplane. This hyperplane is the solution to

$$\max_{\beta,\beta_0,\epsilon_1,...,\epsilon_n} M, \qquad ||\beta|| = 1 \tag{1.32}$$

subject to

$$y_i(\beta_0 + \mathbf{x}_i^T \beta) \geq M(1 - \epsilon_i) \qquad \epsilon_i \geq 0,$$
$$\sum_{i=1}^{N} \epsilon_i \leq C, \tag{1.33}$$

where C ($C > 0$) denotes a *cost* parameter, M is the width of the margin, and $\epsilon_1, ..., \epsilon_n$ are *slack variables*. The latter allow individual training examples to fall on the wrong side of the margin or hyperplane. ϵ_i denotes the position of training example $\mathbf{x}_i$ w.r.t. the margin and hyperplane, i.e. $\epsilon_i = 0$ implies that $\mathbf{x}_i$ is on the correct side of the margin, $\epsilon_i < 0$ that $\mathbf{x}_i$ lies on the wrong side of the margin (*violating the margin*) and $\epsilon_i > 1$ that $\mathbf{x}_i$ falls onto the wrong side of the hyperplane.

The cost C, bounding $\sum_{i=1}^{N} \epsilon_i$, determines the number and severity of tolerated margin violations, such that a large value of C results in a wide margin, whereas a small C leads to the opposite. Similar to the maximal margin classifier, the support vector classifier is defined completely by the training examples that either lie directly *on* the margin or *violate* it, i.e. the *support vectors* (James et al., 2017).

Support Vector Machine

The support vector classifier performs well in a two-class setting if the classes are reasonably divided by a linear boundary, but will fail significantly without modification if the class boundaries are highly non-linear. To solve such a classification task with a support vector classifier one has to embed the original feature space into a higher dimensional one and construct the classifier there. Although the classifier's decision boundary remains linear in this higher dimensional space, it will translate to a non-linear threshold in the

original feature space. This is the basic idea of the support vector machine (SVM; Cortes and Vapnik, 1995).

The optimisation problem of the support vector classifier, i.e. Eqs. (1.32) and (1.33), can be reformulated as

$$f(\mathbf{x}) = \beta_0 + \sum_{i=1}^{N} \alpha_i \langle \mathbf{x}, \mathbf{x}_i \rangle \tag{1.34}$$

where $\langle \mathbf{x}_i, \mathbf{x}_{i'} \rangle = \sum_{j=1}^{n} x_{ij} x_{i'j}$ denotes the inner product and α_i $(i = 1, ..., n)$ are n parameters (one per training example) that are only nonzero for support vectors.

An SVM generalises the inner product with a non-linear kernel function (e.g. a radial kernel), which *implicitly* transforms the classification task to a higher dimensional space. There the SVM constructs a support vector classifier, resulting in the desired non-linear class boundary in the original feature space. Aside from avoiding an explicit transformation to the higher dimensional space, this *kernel trick* also allows the latter to become infinite-dimensional (as e.g. with the radial kernel) and is computationally efficient (see e.g. Hastie et al., 2009, for more details).

To derive class probabilities with an SVM instead of the direct class labels, a sigmoid function can be fit to the SVM decision value f:

$$P(y = 1|f) = \frac{1}{1 + \exp(Af + B)}, \tag{1.35}$$

where A and B are estimated by minimizing the negative log-likelihood function (Platt, 1999).

1.4.3 Random Forest

The *random forest* (RF, Breiman, 2001) is a classifier that is based on the concept of *bagging*. Bagging is a general purpose method to reduce variance in machine learning approaches by constructing many individual models with high variance and low bias, and averaging over their predictions (James

et al., 2017). One such classifier, suffering from high variance, but low bias, is the *decision tree*, which is the underlying model of the random forest.

Decision Tree

The basic idea of the decision tree classifier is that, in an ideal classification scenario, the feature space can be hierarchically partitioned such that all training examples (and new observations) are correctly classified. Such a hiearchical partitioning can be represented by the end points (*leaves*) of a tree model, where each branching point (*node*) of the tree denotes a split of the feature space along one of its axes, e.g. a binary split separating axis x_j into $x_j < t$ and $x_j \geq t$ (where t denotes the split threshold). Constructing the decision tree, the feature space is recursively split in this fashion until a stopping criterion is reached, assigning subsets of the training data to each node and the final leaves along the way. With this assignment, each node i possesses a probability distribution p_{ik} over the classes k (Venables and Ripley, 2002).

To decide where to perform a split during construction of the tree, an *impurity* measure is introduced for each node. A node is considered pure if all of its assigned training examples stem from the same class. At any given node then the split is performed that reduces the average impurity of the tree the most. A typical choice for this impurity measure is the *Gini index*

$$G = \sum_{j \neq k} p_{ij} p_{ik} = 1 - \sum_k p_{ik}^2, \tag{1.36}$$

which quantifies the training error if a node's assigned training examples were classified as class k rather than the majority class in the node. The Gini index is zero for a pure node (Venables and Ripley, 2002).

Stopping criteria for the construction of a decision tree can be all nodes becoming pure, the tree reaching a prescribed maximum depth, or no available split decreasing the average impurity by more than a predefined threshold. If the tree construction is terminated before all leaves are pure, the leaves

assume the majority class of their assigned training examples (Venables and Ripley, 2002).

At prediction time, a new observation is simply propagated along the tree, according to its feature vector, from the *root* (i.e. the first split/branching point) to one of the leaves. The class of the final leaf is the prediction for the class of the new observation (Hastie et al., 2009).

Unlike many machine learning approaches, the decision tree classifier can deal, to some degree, with observations that have an incomplete feature vector. This is achieved by either propagating a new observation as far down the tree as its incomplete feature vector allows, assigning the class of the deepest non-terminal node reached, or by employing *surrogate splits*. In the latter approach each node keeps a list of secondary split criteria, which best approximate the primary split of the node, if the required feature is missing. This list is then considered (in order of best approximation) for prediction of a new observation that misses the primary split feature at a given node (Hastie et al., 2009).

Random Forest Classifier

Following the bagging idea, a random forest classifier constructs B decision trees during training, each of which is built on one of B different bootstrapped datasets that are randomly sampled from the total training data (James et al., 2017; Hastie et al., 2009). To further improve the reduction of variance, beyond just bagging, the construction of the decision trees is slightly modified, with the goal of reducing correlation between the B trees. Instead of selecting the feature for a split that would most reduce the tree's impurity, m random features out of the n-dimensional feature space are chosen at each split decision and then the best feature for the split is selected only among these (again via the impurity criterion). If a large number of the features are correlated a small value of m is preferred, but a generally decent choice for classification tasks is $m = \sqrt{n}$ (Hastie et al., 2009). This modification of the decision tree construction *decorrelates* the B trees by preventing that a

strong predictor feature dominates the split criteria in all trees built[1]. Each of the B decision trees is then typically grown until a minimum terminal node size is reached.

At prediction time the random forest classifier performs a majority vote among its decision trees, i.e. each tree classifies the new observation, their votes are counted and the most voted for class is assigned to the query example (Hastie et al., 2009).

1.4.4 RANSAC

The Random Sample Consensus (RANSAC; Fischler and Bolles, 1981) algorithm is a robust ML method to fit models to datasets that are contaminated by outliers (e.g. noise) to the underlying data-generating distribution. Specifically, the RANSAC approach assumes that the data consists primarily of inliers to a single distribution/model, whose parameters are to be determined by the fit to the data, and a few outliers. To determine the model parameters, RANSAC first randomly samples a series of subsets from the query dataset. Afterwards, it fits the target model to each of the subsets, gathering a collection of sets of fit parameters. Given the earlier assumption on inliers and outliers, most of these random subsets should consist only of inliers to the target model, and, therefore, should return (close to) the same fit parameters. The fits to the minority of subsets that contain one or multiple outliers, however, do not return a consistent set of fit parameters. Consequently, the parameters of the target, data-generating model can be recovered by determining the most voted for set of fit parameters among all of the random subsets (Forsyth and Ponce, 2003).

To achieve this, RANSAC proceeds as follows: First a random subset of N examples is drawn from the data, where N denotes the minimum number of points required to fit the target model (e.g. 2 for a line). Then the number

---

[1]Suppose there is one strong predictor feature along with several moderately strong ones. If the B decision trees are constructed according to the standard procedure, this strong predictor feature would dominate the split criteria. With the random sampling modification, however, on average $(n - m)/n$ nodes do not even consider the dominant feature for the split.

of remaining data points is determined that agree with this fit, i.e. the inliers to this set of fit parameters. This agreement is measured by computing a distance measure between the data and fit, and comparing this distance to a prescribed acceptance threshold. A set of fit parameters is determined to be a good fit if enough inliers are found. Once such a model fit is identified, the model is then refined by reestimating its parameters on the entire set of inliers (the *consensus* set; Forsyth and Ponce, 2003).

This process is repeated k times in order to find the best model as given by a final fitting error. Here, the number k of random subsets of size N is chosen such that the probability p of drawing subsets that contain only outliers is minimised. It is given by the following equation:

$$k = \frac{\log(p)}{\log(1 - w^n)},$$

(1.37)

where w denotes the fraction of inliers in the data. Although the inlier fraction w is often not known a priori, k can still be iteratively determined. Starting from an initial low estimate for w, a random subset is generated and the actual inlier fraction for this subset determined by the RANSAC fit. With this new guess for the true inlier fraction w, the parameter k is then reestimated and a new random subset drawn. This iterative process ultimately terminates once the number of actually drawn subsets exceeds the current estimate of k (Fischler and Bolles, 1981; Forsyth and Ponce, 2003).

1.4.5 Density estimation and the EM algorithm

In this section we shall briefly mention three additional ML approaches that are employed in our studies in Chapter 3, but play a less central role in the presented analysis.

In Ksoll et al. (2018) and Ksoll et al. (2021b) we employ a density estimation algorithm to estimate surface densities for stellar distributions. This is the *kernel density estimator*, which, given a sample of N examples $\{x_1, \ldots, x_N\}$

of a 1D distribution, estimates the density distribution $\hat{p}(x)$ at a given query point x as

$$\hat{p}(x) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right), \tag{1.38}$$

where h denotes a bandwidth factor and $K(x)$ a kernel function. One of the common kernel choices is a simple standard normal distribution $K(x) = 1/\sqrt{2\pi}\exp(-0.5x^2)$. For our purposes this can be extended to 2D using a multivariate standard normal kernel, following

$$p(\mathbf{x}) = \frac{1}{Nh} \sum_{i=1}^{N} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{||\mathbf{x} - \mathbf{x}_i||^2}{2h^2}\right), \tag{1.39}$$

where now $\mathbf{x}, \mathbf{x}_N \in \mathbb{R}^2$ (Bishop, 2009).

In the same two studies we also fit a *Gaussian mixture model* to a 1D density distribution using the *Expectation-Maximisation* algorithm. Similar to the kernel density estimator, Gaussian mixture models are a form of density estimation. They model the target distribution with a linear combination of Gaussian distributions of the form

$$\hat{p}(\mathbf{x}) = \sum_{m=1}^{M} \alpha_m \Phi(\mathbf{x}; \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m), \tag{1.40}$$

where α_m denotes the mixing proportions (with $\sum_{m=1}^{M} \alpha_m = 1$) for the M mixture components with mean $\boldsymbol{\mu}_m$ and covariance $\boldsymbol{\Sigma}_m$ (Bishop, 2009).

The Expectation-Maximisation (EM; Dempster et al., 1977) algorithm is an iterative approach to solve problems via maximum likelihood optimisation. It is an effective numerical alternative in cases where optimising the maximum likelihood becomes analytically unfeasible. Optimising via maximum likelihood with the EM algorithm often includes the introduction of *latent variables* as a means to simplify the problem. To fit a Gaussian mixture model for instance, it is conceptually helpful to introduce a label for every available example x_i that indicates to which mixture component it belongs to, but cannot actually be observed. Starting from an initial guess of the parameters of the model, which is to be fitted via maximum likelihood to the data, the EM algorithm alternates between an expectation (E) and maximisation (M) step.

In the E-step it computes the expectation of the log-likelihood over the data given the current parameter guess. Afterwards, in the M-step an update of the model parameters is derived by maximising this expected log-likelihood with respect to the model parameters. A more thorough description of this approach is beyond the scope of this section and we shall refer to the literature, e.g. McLachlan and Peel (2000); Bishop (2009); Hastie et al. (2009), or Appendix B in Ksoll et al. (2018) in Section 3.1 for more details.

1.5 Deep Learning

Deep Learning describes a sub-discipline of machine learning that, over the last two decades, has gathered a lot of popularity and success on a variety of complicated tasks, ranging from classification and regression over image, speech and text recognition to autonomous driving (to name only a few). It focuses on the construction, application and analysis of a particular set of models called *neural networks*. Although there are many different neural network architectures, in their essence neural networks consist of long (*deep*) chains of *layers*, each representing a (non-linear) transformation and comprised of a collection of *artificial neurons*, hence the name *deep learning* (Goodfellow et al., 2016).

In this section we provide a short, but not exhaustive, overview of both basic deep learning concepts, as well as the specific neural network approach employed in our study in Section 3.2. To that end Section 1.5.1 outlines the basic structure of an artificial neuron, the simplest neural network architecture, and general steps of the training procedure. In Section 1.5.2 we then briefly discuss the neural network approach in the context of solving inverse problems and afterwards present the invertible neural network architecture in more detail as one of the central methodologies applied in this thesis.
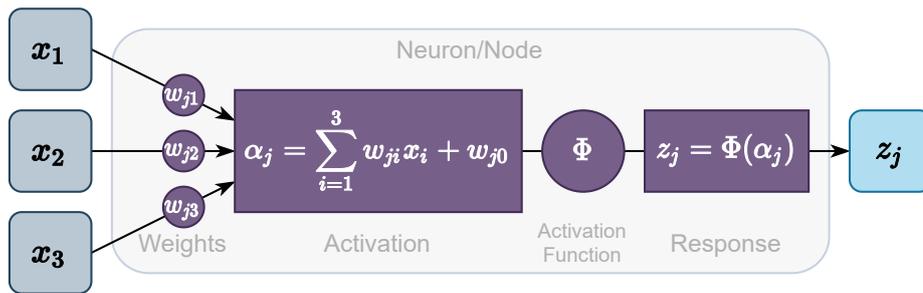
Fig. 1.8.: Schematic of a single artificial neuron used in most neural network architectures.

1.5.1 Neural Networks

Although neural networks have attained their current popularity and success only fairly recently, neural network research and the formulation of some of its basics date back to the 1950s and 1960s, where they were initially inspired by neuroscience (Bishop, 2009; Goodfellow et al., 2016). However, early misconceptions of neural network capabilities led to a temporary disinterest in the discipline. Combined with the need for a few key algorithmic and architectural innovations, and the fact that neural networks benefit substantially from both large data sets and sufficient (parallel, i.e. GPU) computing power, it was not until about 2006 that the machine learning community started to realise the actual potential of the neural network approach (Goodfellow et al., 2016).

Basic Architecture

The basic building blocks of any neural network are the *artificial neurons* (also known as *nodes*), which, as the name implies, are inspired by the neurons in the human brain (Bishop, 2009). Figure 1.8 shows a schematic of a single (artificial) neuron. Given an N-dimensional input (signal) $\mathbf{x}$ (here $N = 3$) the neuron j first computes a linear combination of the input often denoted as *activation*

$$a_j = \sum_{i=1}^{N} w_{ji} x_i + w_{j0}, \qquad (1.41)$$
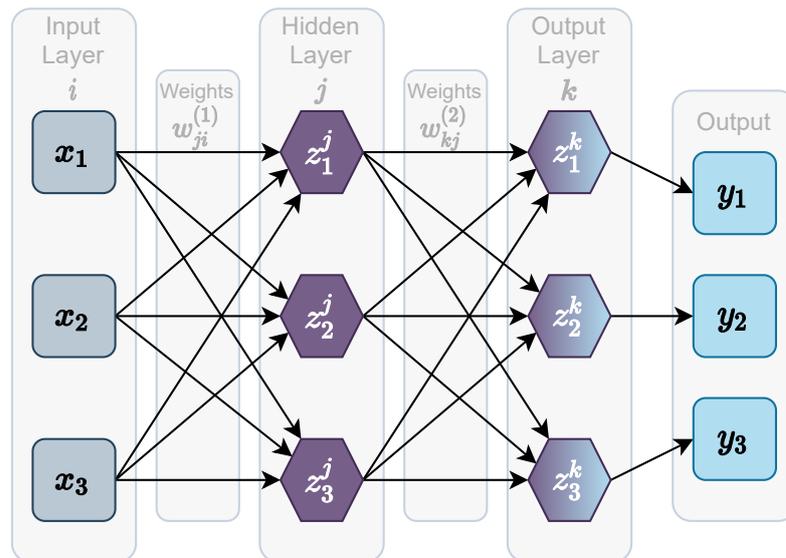
Fig. 1.9.: Architecture of a simple fully connected feed-forward neural network with one hidden layer of size three.

where w_{ji} denotes a set of (trainable) *weights* and w_{j0} is the so-called *bias* of node j. Afterwards, the activation a_j is transformed by a differentiable *activation function* $\Phi(\cdot)$ to determine the *response* of the neuron

$$z_j = \Phi(a_j). \qquad (1.42)$$

Combining a collection of M (often called *width*) such neurons into a *layer* and then chaining these layers, constructs the simplest neural network, the *fully-connected feed-forward* network or *multilayer perceptron* (Bishop, 2009). Fully-connected indicates that in every layer each neuron propagates its response to all neurons of the subsequent layer. Feed-forward denotes that in the combined transformation f, defined by this chain, information only flows from an input $\mathbf{x}$ to a final output $\mathbf{y}$ of the network and no output of the model is fed back into it at any point (Goodfellow et al., 2016).

Figure 1.9 presents an example architecture for a simple fully-connected neural network. Note that the responses of the last layer in the chain, the *output layer*, provide the final network output $\mathbf{y}$ and, therefore, its width D and activation function $\sigma(\cdot)$ are determined by the task that the network aims to solve (Bishop, 2009). In the given example the output layer could return e.g. three real valued parameters of a regression problem or three class

probabilities for a classification task. The layers in the network between the input and the output are referred to as *hidden layers* since the supervised training data does not provide the desired output for these layers (Goodfellow et al., 2016). The simple example in Figure 1.9, therefore, constitutes a neural network with one hidden layer of width $M = 3$, input size $N = 3$ and output dimension $D = 3$. The transformation f performed by this network can be formulated as

$$y_k = f_k(\mathbf{x}; \theta) = \sigma \left[\sum_{j=1}^{M} w_{kj}^{(2)} \Phi \left(\sum_{i=1}^{N} w_{ji}^{(1)} x_i + w_{j0}^{(1)} \right) + w_{k0}^{(2)} \right], \qquad (1.43)$$

where $\mathbf{x}$ denotes the network input vector and θ the entire set of network weights and biases (Bishop, 2009).

For ease of notation, we can also introduce an additional constant input $x_0 = 1$ to each layer in a network so that the biases can be absorbed into the set of weights (Bishop, 2009). The total transformation for the example network above then simplifies to

$$y_k = f_k(\mathbf{x}; \theta) = \sigma \left[\sum_{j=0}^{M} w_{kj}^{(2)} \Phi \left(\sum_{i=0}^{N} w_{ji}^{(1)} x_i \right) \right]. \qquad (1.44)$$

The Activation Function

Equation (1.43) indicates that the type of transformation that can be represented by a neural network depends on the chosen activation functions Φ of the hidden layers. If the activation functions are linear transformations, then the network is just a linear composition of successive linear transformations, rendering f itself a linear transformation (Bishop, 2009). Using a non-linear activation function, however, such as a sigmoid function or the now-widespread rectified linear unit, it has been shown that any neural network with at least one hidden layer (and a linear output layer) can approximate any (Borel-measurable) function from one finite-dimensional space to another with an arbitrarily small, but non-zero, error, provided the hidden layers have enough nodes. This is known as the *universal approximation theorem* for neural networks, which also applies to functions between two

discrete finite-dimensional spaces (Goodfellow et al., 2016). Although a large variety of activation functions exists, including e.g. the logistic sigmoid and the hyperbolic tangent, among the most widely-used and successful today is the *rectified linear unit* (ReLU) $\Phi(z) = \max(0, z)$ and variations thereof.

The Loss Function

With the basic architecture of a simple neural network established, we can now describe the training procedure. In the following we will only consider the supervised training case, i.e. a training data set of N observations $\mathbf{X} = \{\mathbf{x_1}, \ldots, \mathbf{x_N}\}$ and their corresponding targets $\mathbf{Y} = \{\mathbf{y_1}, \ldots, \mathbf{y_N}\}$ are given. Training a neural network consists of finding the optimal set of network parameters θ, i.e. weights and biases, such that the network transformation $f(\mathbf{x}; \theta) = \hat{\mathbf{y}}$ best approximates the true underlying data-generating function $f^*(\mathbf{x}) = \mathbf{y}$ (Goodfellow et al., 2016). To achieve this we have to define an objective function that is optimised during training.

In deep learning this objective function $L(\theta)$ is often referred to as the *loss* and can be written as an average over the training set

$$L(\theta) = \mathbb{E}_{(\mathbf{X},\mathbf{Y})} \mathcal{L}(\mathbf{y}, f(\mathbf{x}; \theta)), \tag{1.45}$$

where $\mathcal{L}(\mathbf{y}, f(\mathbf{x}; \theta))$ denotes the per-observation *loss function*, penalising a discrepancy between the targets $\mathbf{y}$ and the network outputs $f(\mathbf{x}; \theta)$ (Bishop, 2009; Goodfellow et al., 2016). In most applications this objective function decomposes to a sum over the individual training observations

$$L(\theta) = \mathbb{E}_{(\mathbf{X},\mathbf{Y})} \mathcal{L}(\mathbf{y}, f(\mathbf{x}; \theta)) = \sum_{i=1}^{N} \mathcal{L}(\mathbf{y}_i, f(\mathbf{x}_i; \theta)). \tag{1.46}$$

Like the activation function in the output layer of a neural network, this loss function needs to be tailored to the task at hand. For a regression task for instance a simple squared error can be employed

$$L(\theta) = \frac{1}{2} \sum_{i=1}^{N} ||\mathbf{y}_i - f(\mathbf{x}_i; \theta)||^2, \tag{1.47}$$

whereas e.g. the *cross entropy loss*

$$L(\theta) = \sum_{i=1}^{N} y_i \ln(f(\mathbf{x_i}; \theta)) + (1 - y_i) \ln(1 - f(\mathbf{x}, \theta)) \qquad (1.48)$$

is used for binary classification tasks, i.e. $y_i \in \{0, 1\}$ (Bishop, 2009).

Gradient descent

Finding the set of network parameters θ that minimises the objective function $L(\theta)$ pertains to solving

$$\nabla_\theta L(\theta) = 0. \qquad (1.49)$$

This equation defines stationary points of the objective function, which need to be further distinguished into minima, maxima and saddle points. The optimal set of parameters $\hat{\theta}$ is the minimum at which $L(\theta)$ takes the smallest value across the entire θ-space, the *global minimum* (Bishop, 2009). Due to the (usually) highly non-linear dependence of the objective function on the network parameters θ, however, there are often many different sets of θ that fulfil the criterion in Eq. (1.49), but are merely *local minima*. Beside these different minima, there is also the issue of *weight space symmetry*, meaning that for every minimum $\theta_{\min}$ there is a large family of equivalent minima that can be reached by simply reordering nodes and connections within a hidden layer (Bishop, 2009; Goodfellow et al., 2016). Because of this, it is evident that an analytical solution to Eq. (1.49) for the global minimum is practically impossible to determine. Fortunately, finding the global minimum is often not necessary as a "good" local minimum may already allow the neural network to solve a given task reasonably well.

To find such a local minimum numerically, one of the most efficient methods is an iterative procedure, in which, starting from an initial guess $\theta^{(0)}$, the network weights are successively updated such that the objective function decreases with each update (Bishop, 2009). The simplest and most-used update rule is the *gradient descent* approach

$$\theta^{(t+1)} = \theta^{(t)} - \lambda \nabla_\theta L(\theta^{(t)}), \qquad (1.50)$$

with t indicating the iteration step, where the weights are updated by moving in the direction of the loss gradient $\nabla_\theta L(\theta^{(t)})$ (evaluated at the current guess $\theta^{(t)}$) in θ-space with steps of size λ (*learning rate*).

If the training set $(\mathbf{X}, \mathbf{Y})$ is particularly large then this iterative optimisation scheme can become very computationally expensive, because the loss gradient $L(\theta^{(t)})$ has to be evaluated at every weight update for the entire training set. In practice, however, evaluating the gradient on the entire training set may not even be necessary, because a reasonable estimate of the gradient can already be made from a much smaller subset of the data (Goodfellow et al., 2016). This is the idea behind *stochastic gradient descent*, where the update rule is modified to

$$\theta^{(t+1)} = \theta^{(t)} - \lambda \nabla_\theta \sum_{i=1}^{m} \mathcal{L}(\mathbf{y}_i, f(\mathbf{x}_i; \theta^{(t)})), \qquad (1.51)$$

and $m < N$ denotes the size of the chosen subset. If $m = 1$ this method is refers to the "true" *stochastic* gradient descent, whereas for $m > 1$ it is called *mini-batch* stochastic gradient descent.

Backpropagation

Even with a stochastic gradient descent approach, however, computing the loss gradient can become computationally expensive, in particular if a given network entails a large number of weights. When implemented in a naive, brute-force fashion the calculation of the loss gradient also contains a lot of repetition. To avoid these issues and efficiently optimise a neural network, a particular scheme for computing the loss gradient called *backpropagation* is employed.

The first step of the backpropagation scheme actually consists of a forward pass through the network, i.e. given a training observation $\mathbf{x}_n$, the responses $z_i^{(k)}$ and activations $a_i^{(k)}$ of all network layers (including the output layer) are computed. Consider now the derivative of the loss $\mathcal{L}_n(\theta) \equiv \mathcal{L}(\mathbf{y}_n, f(\mathbf{x}_n; \theta))$ for observation $\mathbf{x}_n$ with respect to one of the weights $w_{ji}^{(k+1)}$ between node i in hidden layer k and node j in layer $k + 1$. We can immediately see that

the loss gradient $\partial\mathcal{L}_n(\theta)/\partial w_{ji}^{(k+1)}$ depends on the weight $w_{ji}^{(k+1)}$ only via the activation $a_j^{(k+1)}$ of node j, so that we can write

$$\frac{\partial\mathcal{L}_n(\theta)}{\partial w_{ji}^{(k+1)}} = \frac{\partial\mathcal{L}_n(\theta)}{\partial a_j^{(k+1)}} \frac{\partial a_j^{(k+1)}}{\partial w_{ji}^{(k+1)}} \tag{1.52}$$

using the chain rule for partial derivatives (Bishop, 2009). Defining

$$\delta_j^{(k+1)} \equiv \frac{\partial\mathcal{L}_n(\theta)}{\partial a_j^{(k+1)}} \tag{1.53}$$

and using Eq. (1.41) to find

$$\frac{\partial a_j^{(k+1)}}{\partial w_{ji}^{(k+1)}} = z_i^{(k)}, \tag{1.54}$$

we can then write

$$\frac{\partial\mathcal{L}_n(\theta)}{\partial w_{ji}^{(k+1)}} = \delta_j^{(k+1)} z_i^{(k)}. \tag{1.55}$$

Eq. (1.55) indicates that the we can compute the entire loss gradient as soon as we know δ for every hidden and output layer of the network (Bishop, 2009). Using again the chain rule of partial derivatives we can write for δ_j^{k+1} in a hidden layer

$$\delta_j^{(k+1)} \equiv \frac{\partial\mathcal{L}_n(\theta)}{\partial a_j^{(k+1)}} = \sum_b \frac{\partial\mathcal{L}_n(\theta)}{\partial a_b^{(k+2)}} \frac{\partial a_b^{(k+2)}}{\partial a_j^{(k+1)}}. \tag{1.56}$$

This follows from the fact that a variation of $a_j^{(k+1)}$ influences $\mathcal{L}_n(\theta)$ only via the b nodes it is connected to in the consecutive layer $k+2$ (Bishop, 2009). Using Eqs. (1.53), (1.41) and (1.42) in Eq. (1.56) we find

$$\delta_j^{(k+1)} = \Phi'\left(a_j^{(k+1)}\right)\sum_b w_{bj}^{(k+2)}\delta_b^{(k+2)}. \tag{1.57}$$

Eq. (1.57) shows that we can immediately determine the $\delta^{(k+1)}$ for any hidden layer by propagating the $\delta^{(k+2)}$ of its consecutive layer *backwards* in the network. This property gives rise to the efficient *backpropagation* scheme to calculate the total loss gradient: First, pass a training observation x_n forward through the network to determine the activations and responses of all layers. Second, determine the $\delta_j^{(K)}$ for the output layer K, which are readily

computable given the loss function and the targets y_n. Lastly, determine the $\delta_j^{(k)}$ for all hidden layers by recursively propagating $\delta_j^{(k+1)}$ backwards in the network starting from the now known $\delta_j^{(K)}$ (Bishop, 2009).

The Adam optimiser

After determining the gradient of the loss using the backpropagation scheme, the updates for the weights θ can be determined by some flavour of gradient descent algorithm. Although these algorithms all follow the same basic concept, they vary in how fast they reach convergence and how prone they are to getting stuck in a suboptimal local extremum. In the following we briefly describe one of the most widely used stochastic gradient descent optimisers, the *Adam* (adaptive moment; Kingma and Ba, 2014) optimiser, which we employ in our study in Section 3.2.

The Adam optimiser employs both an adaptive learning rate approach and incorporates the concept of *momentum* (Goodfellow et al., 2016). Momentum is inspired by the physical analogue and entails record-keeping and accumulation of the loss gradients during the optimisation procedure. Specifically, it means that the weight updates do not just follow the loss gradient of the current mini-batch, but instead also take into account the gradients of previous iterations (through an exponentially decaying moving average). Consequently, the update procedure does no longer abruptly change direction in weight space but does so more gradually instead. This helps convergence in the presence of either noisy or small but consistent loss gradients (Goodfellow et al., 2016).

An adaptive learning rate serves both to avoid overshooting potentially good solutions and mitigating slow convergence. Given a fixed learning rate and a large loss gradient, the weight update may miss the minimum the gradient is pointing to simply by taking too big a step in weight space. On the other hand, if the loss gradient is very small, the optimiser's progress towards the optimum may be slow and time-consuming. To solve these issues, adaptive learning rate approaches, therefore, reduce the step size of the weight update in the presence of a large gradient and do the opposite for small gradients (Goodfellow et al., 2016).

The Adam optimiser incorporates these two principles by using first $\mathbf{s}$ and second $\mathbf{r}$ moments of the loss gradient. Starting from an initial guess $\theta^{(0)}$ for the weights and initialising the moments $\mathbf{r}$, $\mathbf{s}$ and the time step t to zero, first the gradient $\mathbf{g}$ of the loss with respect to the current mini-batch of size m is computed via

$$\mathbf{g} = \frac{1}{m} \nabla_\theta \sum_{i=1}^{m} \mathcal{L}(\mathbf{y}_i, f(\mathbf{x}_i; \theta)). \tag{1.58}$$

After setting $t \leftarrow t + 1$, the first and second moments are computed

$$\begin{aligned}
\mathbf{s} &\leftarrow \beta_1 \mathbf{s} + (1 - \beta_1)\mathbf{g} \\
\mathbf{r} &\leftarrow \beta_2 \mathbf{r} + (1 - \beta_2)\mathbf{g} \odot \mathbf{g},
\end{aligned} \tag{1.59}$$

where $\odot$ denotes element-wise multiplication, and then corrected for the bias of the zero initialisation:

$$\begin{aligned}
\hat{\mathbf{s}} &= \frac{\mathbf{s}}{1 - \beta_1^t} \\
\hat{\mathbf{r}} &= \frac{\mathbf{r}}{1 - \beta_2^t}.
\end{aligned} \tag{1.60}$$

β_1 and β_2 denote exponential decay rates for the moments here and constitute the set of hyperparameters for the Adam optimiser, together with the initial learning rate λ_0. Lastly, the weight update is computed according to

$$\theta \leftarrow \theta - \lambda_0 \frac{\hat{\mathbf{s}}}{\sqrt{\hat{\mathbf{r}}} + \delta}, \tag{1.61}$$

where δ denotes a small non-zero constant introduced for numerical stability and all operations are again performed element-wise. The first moment $\hat{\mathbf{s}}$ incorporates the momentum principle, whereas the second moment $\hat{\mathbf{r}}$ is used to adapt the learning rate (Goodfellow et al., 2016).

1.5.2 Invertible Neural Networks

Inverse Problems

In science simulations and modelling are fundamental approaches to under-
stand and predict real world phenomena. After constructing theories that
describe a certain phenomenon, these tools allow to predict the outcome
of some measurements given a complete description of the system. For
instance, knowing the underlying physical properties of a star (e.g. mass,
age, chemical composition etc.) one can simulate how a certain telescope
would observe this source (e.g. synthetic photometry). The task of predicting
the result of measurements is often referred to as a *forward problem*, as
it aims to infer the effects of something given its causes (Tarantola, 2005;
Richter, 2015).

In many scientific applications the more interesting part, however, is to
actually recover the underlying parameters that characterise the system
from the measurements. Inverting the associated forward problem, these
tasks are called *inverse problems*, that is, aim to determine the causes of
something given its effects. Whereas the forward problem is usually well
understood in the sense that it (the simulation) is often deterministic, easily
computed and has a unique solution, the inverse problem most of the time
is not. Be it due to critical information loss in the mapping to a given set of
observables or an intrinsic property of the phenomenon itself, in the inverse
problem there are often multiple solutions to a single query, i.e. the inverse
mapping is *degenerate* (Tarantola, 2005; Richter, 2015). For instance, given
the right combinations of stellar mass, age and chemical composition, two
entirely different stars may look almost identical in the HRD or observed CMD
(cf. Figure 1.5). This property makes inverse problems difficult to solve and,
in some cases, certain system parameters may be unrecoverable entirely
from the available observables (Ardizzone et al., 2018).

To adequately analyse inverse problems, a solver is necessary that can
provide the *complete posterior distribution* $p(\mathbf{x}|\mathbf{y})$ of the system's parameters
$\mathbf{x}$ conditioned on the given measurements $\mathbf{y}$. Only such a solver allows a
proper quantification of uncertainty, can reveal multi-modal distributions and

find degenerate or unrecoverable parameters in a given inverse problem (Ardizzone et al., 2018). Modelling conditional posterior distributions for inverse processes is in principle a task that can be tackled with Bayesian statistical approaches. However, many practical applications in science are often too complex for an exact Bayesian treatment. Because of this, the most widespread approach to tackle posterior prediction for inverse problems is the Markov Chain Monte Carlo (MCMC, Robert and Casella, 2004; Gamerman and Lopes, 2006) sampling method. Although undoubtedly successful in predicting posterior distributions, MCMC approaches are unfortunately very expensive in terms of computing time (Ardizzone et al., 2018). A more efficient alternative to MCMC is the so called approximate Bayesian computation (ABC, for a review see e.g. Sunnåker et al., 2013) approach, provided an implementation of the forward model (i.e. simulation solving the forward problem) exists. ABC combines this forward model with rejection sampling in order to compute an approximation of the posterior distribution.

Solving the inverse problem and predicting conditional posteriors has also seen attention in deep learning. For instance, neural networks predicting parametric representations of the true posteriors (Papamakarios and Murray, 2016; Siddharth et al., 2017) or dropout variational inference approaches (Gal and Ghahramani, 2015; Kingma et al., 2015) have been proposed. In principle, conditional generative adversarial networks (cGAN; Mirza and Osindero, 2014; Isola et al., 2017) or conditional variational auto encoders (cVAE; Sohn et al., 2015) are also suitable architectures to solve inverse problems (Ardizzone et al., 2018). Generative modelling, i.e. learning a non-linear transformation between a simple prior distribution and the actual data distribution (Deco and Brauer, 1995; Hyvärinen and Pajunen, 1999), has also great potential for this task. More specifically, this entails neural networks employing the concepts of *normalising flows* (Tabak and Vanden-Eijnden, 2010; Tabak and Turner, 2013), where normal densities are gradually transformed into the target data density (see e.g Rippel and Prescott Adams, 2013; Rezende and Mohamed, 2015; Tomczak and Welling, 2016; Trippe and Turner, 2018, for applications), and *auto-regressive flows* (Kingma et al., 2016), which decompose the density following the Bayesian chain rule (see also e.g Huang et al., 2018; Papamakarios et al., 2017; Uria et al., 2016).

The Invertible Neural Network

In this thesis we employ a recently proposed novel approach to solving inverse problems called *invertible neural network* (INN, Ardizzone et al., 2018). As defined in Ardizzone et al. (2018), INNs are a particular neural network architecture that encompasses three key properties:

i) The input-output transformation is bijective, i.e. it has an inverse,

ii) both the input-output and the inverse output-input mappings are efficiently computable, and

iii) the Jacobian of both the forward and inverse mapping is tractable, so that posterior probabilities can be explicitly calculated.

A network that fulfils these criteria has the unique advantage that it can be trained to model the usually well understood forward process (i.e. the forward problem) and automatically provides the desired solution to the inverse problem for free by simply running it in reverse at prediction time. This characteristic ability of INNs also avoids one of the major difficulties in solving inverse problems, arising when the posterior distributions are to be modelled directly, namely the definition of an adequate supervised loss. If the chosen loss function does not match the shapes of the target posteriors, a method is likely to converge to an incorrect solution (Ardizzone et al., 2018).

To solve a given inverse problem the INN approach, as described in Ardizzone et al. (2018), aims to approximate the posterior distributions $p(\mathbf{x}|\mathbf{y})$ with a model $q(\mathbf{x}|\mathbf{y})$, or more specifically a deterministic function g represented by an INN. Given a well understood simulation $s(\mathbf{x}) = \mathbf{y}$ of the forward process, that maps the underlying (e.g.) physical parameters $\mathbf{x}$ (that cannot be directly measured) to a set of observable quantities $\mathbf{y}$, it is then assumed that this forward mapping is subject to an inherent loss of information. Due to the latter the observables $\mathbf{y}$ can no longer explain all the variance of the physical parameters $\mathbf{x}$, such that the mapping $\mathbf{y} \rightarrow \mathbf{x}$ becomes degenerate. To capture the information that is otherwise lost, the INN approach introduces

a set of additional, *latent variables* $\mathbf{z}$ to encode all variance of $\mathbf{x}$ that is not entailed in $\mathbf{y}$.

The INN is then trained to model the known forward process, i.e. it learns to associate the physical parameters $\mathbf{x}$ to unique pairs $[\mathbf{y}, \mathbf{z}]$ of observables and latent variables:

$$[\mathbf{y}, \mathbf{z}] = f(\mathbf{x}) = [f_\mathbf{y}(\mathbf{x}), f_\mathbf{z}(\mathbf{x})] = g^{-1}(\mathbf{x}), \tag{1.62}$$

where $f_\mathbf{y}(\mathbf{x}) \approx s(\mathbf{x})$. By virtue of its invertible architecture this procedure implicitly determines the inverse $\mathbf{x} = f^{-1}(\mathbf{y}, \mathbf{z}) = g(\mathbf{y}, \mathbf{z})$. Two necessary ingredients for training an INN in this setup are the dimension of the latent variables $K = \dim(\mathbf{z})$ and their prior distribution $p(\mathbf{z})$.

Let $N = \dim(\mathbf{x})$ and $M = \dim(\mathbf{y})$ be the nominal dimensions of the physical parameters and the observables, respectively, and let m be the intrinsic dimension of $\mathbf{y}$, with $m \leq M$. The information loss assumption automatically implies $N > m$, even if the nominal dimensions may satisfy $M > N$. Consequently, the dimension of $\mathbf{z}$ follows as $K = N - m$, as the relation $f = g^{-1}$ only holds if the nominal and intrinsic dimensions of both sides match. In the case that the nominal dimensions $M + K$ exceed N, the input vector $\mathbf{x}$ can be augmented by a vector of zeros $\mathbf{x}_0$ with dimension $M + K - N$, so that $\mathbf{x}$ is replaced by the concatenation $[\mathbf{x}, \mathbf{x}_0]$ (Ardizzone et al., 2018).

For simplicity $p(\mathbf{z})$ is assumed to be a K-dimensional multivariate standard Normal distribution and this shape is enforced as part of the training procedure. The learned function $g(\mathbf{y}, \mathbf{z}) = \mathbf{x}$, thus, retrieves the target posterior distribution $p(\mathbf{x}|\mathbf{y})$ by transforming this known prior distribution $p(\mathbf{z})$ to $\mathbf{x}$-space given the condition $\mathbf{y}$ (Ardizzone et al., 2018). In practice, the posterior distribution is constructed by sampling the latent variables according to their prior $p(\mathbf{z})$ and computing $g(\mathbf{y}, \mathbf{z}_i)$ for each sample $\mathbf{z}_i$ given the observation $\mathbf{y}$.

In summary the INN approach, thus, represents the posterior model $q(\mathbf{x}|\mathbf{y})$ as

$$q(\mathbf{x} = g(\mathbf{y}, \mathbf{z})|\mathbf{y}) = p(\mathbf{z})|J_{\mathbf{x}}|^{-1},$$

$$J_{\mathbf{x}} = \det\left(\left.\frac{\partial g(\mathbf{y}, \mathbf{z})}{\partial[\mathbf{y}, \mathbf{z}]}\right|_{\mathbf{y}, f_{\mathbf{z}}(\mathbf{x})}\right), \quad (1.63)$$

$$\mathbf{z} \sim p(\mathbf{z}) = \mathcal{N}(\mathbf{z}, 0, \mathbf{I}_K),$$

where $J_{\mathbf{x}}$ denotes the Jacobian determinant and $\mathbf{I}_K$ the K-dimensional unit matrix.

To create an INN, Ardizzone et al. (2018) employ a chain of reversible, so-called *affine coupling layers* based on the RealNVP architecture designed by Dinh et al. (2016). In these layers the input vector $\mathbf{u}$ is split into two halves $\mathbf{u_1}$ and $\mathbf{u_2}$, which are then transformed by two complementary affine transformations with element-wise multiplication $\odot$ and addition $+$,

$$\mathbf{v_1} = \mathbf{u_1} \odot \exp(s_2(\mathbf{u_2})) + t_2(\mathbf{u_2}),$$
$$\mathbf{v_2} = \mathbf{u_2} \odot \exp(s_1(\mathbf{v_1})) + t_1(\mathbf{v_1}), \quad (1.64)$$

where s_i and t_i ($i \in \{1, 2\}$) are arbitrarily complex mappings of $\mathbf{u_2}$ and $\mathbf{v_1}$, which can even be realised as small neural networks and do *not* need to be invertible themselves. Given the output vector $\mathbf{v} = [\mathbf{v_1}, \mathbf{v_2}]$ these affine transformations are trivially inverted,

$$\mathbf{u_2} = (\mathbf{v_2} - t_1(\mathbf{v_1})) \odot \exp(-s_1(\mathbf{v_1})),$$
$$\mathbf{u_1} = (\mathbf{v_1} - t_2(\mathbf{u_2})) \odot \exp(-s_2(\mathbf{u_2})). \quad (1.65)$$

To facilitate interaction between variables in the INN architecture, these affine coupling blocks can be alternated with random permutation layers, shuffling the input of the subsequent layer, such that the splits $\mathbf{u} = [\mathbf{u}_1, \mathbf{u}_2]$ vary for each reversible layer.

Training an INN for an inverse problem requires a bi-directional training regime. Here forward and backwards passes are alternated, accumulating loss gradients for both directions, before the network weights are updated by

back-propagation. This process, thus, entails the definition of three losses, $\mathcal{L}_\mathbf{x}$, $\mathcal{L}_\mathbf{y}$ and $\mathcal{L}_\mathbf{z}$.

$\mathcal{L}_\mathbf{y}(\mathbf{y}_i, f_\mathbf{y}(\mathbf{x_i}))$ is a standard supervised loss that penalises deviations of the network prediction $f_\mathbf{y}(\mathbf{x}_i)$ from the actual simulation outcome $\mathbf{y}_i = s(\mathbf{x}_i)$, e.g. a standard squared loss for an inverse regression problem.

The loss on the latent variables $\mathcal{L}_\mathbf{z}$ is invoked to ensure a) that the latent variables follow the prescribed prior normal distribution $p(\mathbf{z})$ and b) that the network converges such that $\mathbf{y}$ and $\mathbf{z}$ are independent (i.e. $p(\mathbf{z}|\mathbf{y}) = p(\mathbf{z})$) and do not twice encode the same information (Ardizzone et al., 2018). This is achieved by penalisation $\mathcal{L}_\mathbf{z}(q(\mathbf{y}, \mathbf{z}), p(\mathbf{y})p(\mathbf{x}))$ of a mismatch between the joint distribution of network outputs $q(\mathbf{y} = f_\mathbf{y}(\mathbf{x}); \mathbf{z} = f_\mathbf{z}(\mathbf{x})) = p(\mathbf{x})/|J_\mathbf{yz}|$ and the product of the marginal distributions of latent variables $p(\mathbf{z})$ and simulation outcomes $p(\mathbf{y}) = s(\mathbf{x})) = p(\mathbf{x})/|J_s|$. To avoid that $\mathcal{L}_\mathbf{z}$ counteracts $\mathcal{L}_\mathbf{y}$, i.e. that the weight updates worsen the $\mathbf{y}$-predictions, the $\mathcal{L}_\mathbf{z}$ gradients are blocked with respect to $\mathbf{y}$ (Ardizzone et al., 2018). A cheap, stable and easy-to-use choice for $\mathcal{L}_\mathbf{z}$ is the *Maximum Mean Discrepancy* loss (MMD, see Gretton et al., 2012, for details), which is a kernel-based approach to compare two probability distributions that are given only by samples. Note that the Jacobian determinants J_s and $J_\mathbf{yz}$ do not need to be known explicitly if $\mathcal{L}_\mathbf{z}$ is realised via the MMD loss (Ardizzone et al., 2018).

Lastly, $\mathcal{L}_\mathbf{x}(q(\mathbf{x}), p(\mathbf{x}))$ penalises a deviation of the distribution of backward predictions $q(\mathbf{x}) = p(\mathbf{y} = f_\mathbf{y}(\mathbf{x})p(\mathbf{z} = f_\mathbf{z}(\mathbf{x}))/|\mathbf{J_x}|$ from the prior data distribution $p(\mathbf{x})$, also implemented as an MMD loss. Ardizzone et al. (2018) prove that the INN prediction will converge to the true posterior distribution in the asymptotic limit that $\mathcal{L}_\mathbf{y}$ and $\mathcal{L}_\mathbf{z}$ reach zero, but in practice they find a residual dependency between $\mathbf{y}$ and $\mathbf{z}$ after a finite training time. Because of this, $\mathcal{L}_\mathbf{x}$ is introduced to improve convergence of the method.

The Conditional Invertible Neural Network

Although our study in Section 3.2 has seen ample experimentation with the initial Ardizzone et al. (2018) INN approach during the research phase, we ultimately settled for an architecture that improves upon the INN foundation.
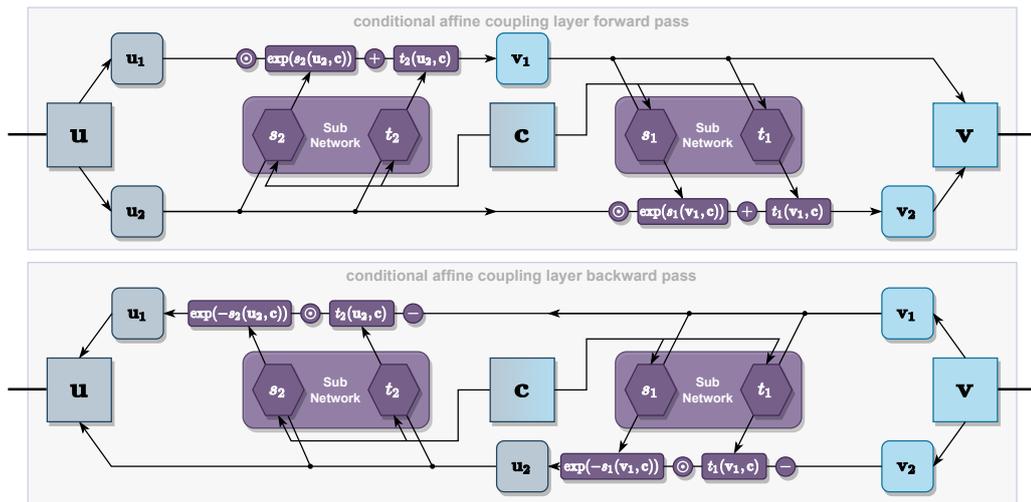
Fig. 1.10.: Schematic overview of a conditional affine coupling layer, outlining the forward and backward passes in the top and bottom panel, respectively. Following the Kingma and Dhariwal (2018) GLOW architecture, the transformations t_1, s_1 and s_2, t_2 are realised by a single sub-network each here, instead of having an individual sub-network for each transformation.

This extension is the *conditional invertible neural network* (cINN) proposed in Ardizzone et al. (2019).

The main modification of the INN architecture in the cINN is the introduction of additional conditioning inputs $\mathbf{c}$ to the affine coupling blocks. Since the sub-networks (or functions) s_i and t_i are only evaluated in the forward direction, even while inverting the network, one can concatenate an additional condition to the regular sub-network inputs, i.e. replace $s_2(\mathbf{u_2})$ with $s_2(\mathbf{u_2}, \mathbf{c})$ etc. in Eqs. (1.64) and (1.65), without compromising the overall invertibility of the network architecture. Figure 1.10 illustrates how this modification ties into the operation (forwards and backwards) of the affine coupling blocks.

In Ardizzone et al. (2019) this modification is proposed as a means to introduce a feature extraction network into the INN to facilitate an image colourisation/generation task by pre-processing input images into useful features (then used as $\mathbf{c}$). That said, the cINN architecture is easily applied to inverse regression problems by taking the observables $\mathbf{y}$ as the conditioning input $\mathbf{c}$. Consequently, the cINN does *not* predict the observables $\mathbf{y}$ in the forward process anymore, i.e. the forward and backward mappings are modified to $f(\mathbf{x}; \mathbf{c}, \theta) = \mathbf{z}$ and $g(\mathbf{z}; \mathbf{c}, \theta) = \mathbf{x}$ (where θ denotes the cINN model

parameters), respectively. The invertibility is, thus, given for fixed condition $\mathbf{c}$ as

$$f(\,\cdot\,;\mathbf{c})^{-1} = g(\,\cdot\,;\mathbf{c}). \tag{1.66}$$

Although these changes in the forward and backward mappings may appear significant in comparison to the original INN approach to inverse problems, the cINN still learns to encode all information about the parameters $\mathbf{x}$ in the latent variables $\mathbf{z}$ that is not captured by the observables $\mathbf{y}$, just like the standard INN. Likewise, posteriors for a given observation (condition) $\mathbf{y}$ are predicted by sampling from the known normal prior $p(\mathbf{z})$ of the latent variables using the inverted network g :

$$\mathbf{x_{posterior}} = g(\mathbf{z};\mathbf{c}=\mathbf{y}), \ \ \text{with } z \sim p(\mathbf{z}) = \mathcal{N}(\mathbf{z},0,\mathbf{I}_K), \tag{1.67}$$

where $\mathbf{I}_K$ is the $K \times K$ unity matrix with $K = \dim(\mathbf{z})$.

The cINN has a few advantages over the INN that improve stability and training efficiency. First, the cINN does not require zero padding of the input vector $\mathbf{x}$ in the case that the nominal dimension of the observables $\mathbf{y}$ exceeds that of $\mathbf{x}$. Since the forward process now only maps $\mathbf{x}$ to $\mathbf{z}$ (conditioned on $\mathbf{c}$), the invertibility is automatically given when the dimensions K and M of the latent variables and input parameters match. With the observables $\mathbf{y}$ provided as the conditioning input $\mathbf{c}$, their dimension can be arbitrarily large (like e.g. an image). As previously mentioned, this also provides the opportunity to integrate a feature-extraction network h into the cINN, which transforms a given condition $\mathbf{c}$ into an intermediate representation $\tilde{\mathbf{c}} = h(\mathbf{c})$. Such a network h can be trained jointly with the cINN, such that a transformation is learned that is specifically tailored to the cINN's needs.

Secondly, (although this is not necessarily exclusive to the cINN), Ardizzone et al. (2019) adapt a more computationally efficient version of the affine coupling blocks, replacing the Dinh et al. (2016) RealNVP setup by the GLOW (Generative Flow; Kingma and Dhariwal, 2018) architecture. The modification in the GLOW coupling layer is rather minor, instead of employing an invidiual sub-network for all four of the transformations s_i and t_i ($i \in \{1,2\}$), now the outputs of s_1, t_1 and s_2, t_2 are predicted by a single network each

(i.e. only two instead of four sub-networks). This effectively halves the total number of weights that need to be trained.

Lastly, Ardizzone et al. (2019) show that the cINN can be effectively trained by minimisation of the *maximum likelihood loss* alone. With only one loss function $\mathcal{L}$ to track, compared to the three individual losses $\mathcal{L}_\mathbf{x}$, $\mathcal{L}_\mathbf{y}$ and $\mathcal{L}_\mathbf{z}$ in the standard INN approach, this further improves stability and efficiency of the training process. With the prescribed prior distribution $p_Z(\mathbf{z})$ on the latent space Z, the forward model f essentially assigns a probability to any given input $\mathbf{x}$ in dependence on the condition $\mathbf{c}$ and the cINN model parameters θ. Making use of the change-of-variables formula, this can be written as

$$ p_X(\mathbf{x}; \mathbf{c}, \theta) = p_Z(f(\mathbf{x}; \mathbf{c}, \theta)) \left| \det \left(\frac{\partial f}{\partial \mathbf{x}} \right) \right|, \tag{1.68} $$

where $\partial f / \partial \mathbf{x}$ is the Jacobian matrix. Using Bayes' theorem the posterior over the model parameters θ can be determined as

$$ p(\theta; \mathbf{x}, \mathbf{c}) \propto p_X(\mathbf{x}; \mathbf{c}, \theta) p_\theta(\theta). \tag{1.69} $$

To train a cINN we have to find the optimal model parameters $\hat{\theta}$, i.e. solve

$$ \hat{\theta} = \arg\max_\theta p(\theta; \mathbf{x}, \mathbf{c}), \tag{1.70} $$

which can be achieved by minimisation of the negative log likelihood loss

$$ \mathcal{L} = \mathbb{E}_i \left[\log \left(p_X(\mathbf{x}_i; \mathbf{c}_i, \theta) \right) \right] - \log \left(p_\theta(\theta) \right). \tag{1.71} $$

Assuming the standard normal distribution for $p_Z(\mathbf{z})$ and a Gaussian prior on the network weights θ, this further simplifies to

$$ \mathcal{L} = \mathbb{E}_i \left[\frac{||f(\mathbf{x_i}; \mathbf{c_i}, \theta)||_2^2}{2} - \log |J_i| \right] + \frac{1}{2\sigma_\theta^2} ||\theta||_2^2, \tag{1.72} $$

where $\mathbf{x_i}$ is a training sample with corresponding condition $\mathbf{c_i}$ and J_i denotes the determinant of the Jacobian matrix $J_i = \det \left(\partial f / \partial \mathbf{x} |_{\mathbf{x_i}} \right)$ evaluated at $\mathbf{x_i}$. Here, the second term acts as the L2 weight regularisation, whereas the first is the maximum likelihood loss (Ardizzone et al., 2019).

1.6 Related Works

As stated in Section 1.0.1, among the central goals of this thesis is the exploration of machine learning methods for the identification and characterisation of PMS stars. In this section we briefly summarise a few example approaches in the literature related to these tasks.

Employing CMDs from surveys of star forming regions, PMS stars can be distinguished from more evolved contaminants, such as (low-mass) lower main-sequence (LMS) stars, by making use of reference fields. The latter are observations of the galactic field population (i.e. the primary source of contamination) in the vicinity of the target. Constructing CMDs for both the observed field and survey target, the true cluster population can be recovered (in the CMD) by statistically subtracting the field from the target CMD. See for example Gouliermis et al. (2007) for an application of this approach to the star-forming region LH 95 in the Large Magellanic Cloud.

In a more recent study, Zivkov et al. (2018) identify regions hosting PMS stars by using differential Hess diagrams. A Hess diagram is a 2D-histogram variant of the CMD, i.e. it indicates either number or density of stars in 2D bins in the CMD. Zivkov et al. (2018) distinguish young stellar populations from the field by identifying them as (spatial) regions that show a density excess in predefined areas (i.e. the PMS regions in the CMD) in a differential Hess diagram, i.e. the difference between the Hess diagrams of the observed star-forming region and the underlying field population. Zivkov et al. (2018) successfully apply this approach to several star-forming regions in the Large Magellanic Cloud using data from the photometric VISTA Survey of the Magellanic Clouds (VMC; Cioni et al., 2011) and derive PMS density contours for the observed regions. This approach can, however, not recover PMS sources individually, providing only estimates for the total number within the identified regions.

In Section 3.3 we introduce an ML approach to constrain the slope of the extinction-elongated RC (cf. Section 1.3). Previously, e.g. De Marchi et al. (2016) quantify the slope of the reddening vector via the RC feature in the CMD by employing unsharp masking. Starting from an image of the CMD,

centred on the RC, they subtract a second, Gaussian-blurred version of the CMD image from the original to sharpen the observed extinction-elongated RC. This filters out the noise around the RC feature (i.e. other RGB sources), narrowing it down to a sharp line, which can then be fitted to derive the slope of the reddening vector (De Marchi et al., 2016).

Regarding the prediction of stellar physical parameters from photometry, in particular probabilisitc isochrone fitting techniques based on Bayesian statistics have been investigated to derive posterior probability distribution functions of the properties of individual stars (see e.g. Valls-Gabaud, 2014, for a review). For instance Gordon et al. (2016) have developed such a Bayesian analysis tool for multi-band photometric surveys. Probabilistic Bayesian approaches also exist for the prediction of stellar parameters from full observed spectra (e.g. BONNSAI; Schneider et al., 2014, 2017). Bayesian inference of stellar physical parameters based on photometry has also been successfully applied to time-domain observations, e.g. light curves using photometric brightness variations (e.g. Miller et al., 2015), and time-series asteroseismic surveys (e.g. Bellinger et al., 2016).

In Section 3.2 we propose a novel neural network approach for the prediction of stellar physical parameters from photometry. Several recent studies in astronomy have employed neural networks for similar prediction tasks. For instance Sharma et al. (2020) construct a convolutional neural network to classify stellar spectra following the Harvard classification scheme and successfully test their method on data from the Sloan Digital Sky Survey. Using photometry and parallaxes from the second data release DR2 of the Gaia (Gaia Collaboration et al., 2016, 2018) survey, Kounkel et al. (2020) employ a neural network to predict age, extinction and distance for stellar clusters in the Milky Way and study the star-formation activity in the spiral arms. In a follow-up study McBride et al. (2020) extend the Kounkel et al. (2020) neural network to predict ages for PMS stars based on combined photometry from Gaia DR2 and the 2MASS (Two Micron All Sky Survey; Skrutskie et al., 2006) survey. A similar neural network approach to Kounkel et al. (2020) has been proposed by Cantat-Gaudin et al. (2020), who also employ Gaia data to predict physical parameters of stellar clusters, but take 2D histograms of the observed CMDs as inputs for their method instead.

Lastly, Olney et al. (2020) develop a deep convolutional neural network for the analysis of APOGEE spectra of YSOs and predict surface temperature, surface gravity and metallicity for these objects. As part of their training data construction they also utilise a secondary convolutional neural network to infer YSO parameters, i.e. ages, masses, extinction and surface temperature/gravity, from Gaia photometry in nine filter bands, a distance measure, as well as stellar radius and luminosity. Similarly to our approach in Section 3.2, Olney et al. (2020) train this auxiliary network using synthetic isochrone tables and manage to successfully predict surface temperatures for YSOs on real Gaia observations.

For a more detailed discussion of recent applications of various machine and deep learning approaches in astronomy, we refer to the reviews by Fluke and Jacobs (2020) and Baron (2019).

Targets and Data

This chapter briefly summarises the key properties of the target regions and corresponding surveys that are analysed in the four studies of this thesis. These are the young stellar cluster Westerlund 2 (Sections 2.1, 3.2), the old globular cluster NGC 6397 (Sections 2.2, 3.2) and two giant star-forming complexes located in the Large Magellanic Cloud (Section 2.3), the Tarantula Nebula (Sections 2.4 and 3.1) and LHα 120–N44 (Sections 2.5, 3.3, 3.4).

2.1 Westerlund 2

Westerlund 2 (Wd2, shown in Figure 2.1) is one of the two clusters that we analyse in Section 3.2 to test our cINN approach on real observational data. Harbouring a total of $3.7 \times 10^4 \, \mathrm{M_\odot}$ in stellar mass (Zeidler et al., 2017; Ascenso et al., 2007), Wd2 is the second most massive star-forming cluster in the Milky Way, just after Westerlund 1. Located in the Carina-Sagittarius arm, it lies at a distance of $4.16 \pm 0.33 \, \mathrm{kpc}$ (Zeidler et al., 2015) away from our Sun and is the ionising central stellar cluster of the H II region RCW49 (Rodgers et al., 1960). Wd2 consists of two almost coeval clumps (Zeidler et al., 2015) and exhibits evidence of mass segregation (Zeidler et al., 2017), i.e. its massive constituents appear more centrally clustered than their less massive counterparts. Wd2 hosts a rich ensemble of OB-type stars (Vargas Álvarez et al., 2013) and a large population of PMS stars, with the latter likely still accreting material from their circumstellar envelopes (Zeidler et al., 2016). Given its large mass and young age of only about 1–2 Myr (Zeidler et al., 2015; Vargas Álvarez et al., 2013), Westerlund 2 is an excellent local example of systems undergoing episodes of intense star formation, which are so bright that they can still be observed at distances of several Mpc in e.g. star

burst galaxies (Sabbi et al., 2020). Additionally, so far no supernova has been detected in Wd2, making it a prime target to study the pre-supernova evolution of a young stellar cluster (Zeidler et al., 2021).
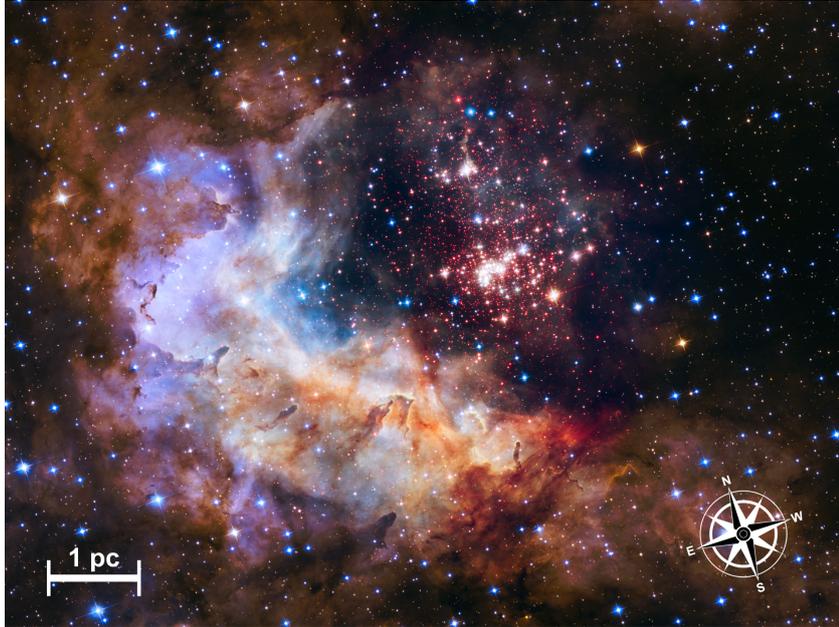
Fig. 2.1.: HST three-colour composite image of Westerlund 2. Original image by NASA, ESA, the Hubble Heritage Team (STScI/AURA), A. Nota (ESA/STScI), and the Westerlund 2 Science Team. Size scale and orientation indicators added for this thesis.

In Section 3.2 we analyse the photometric catalogue compiled by Sabbi et al. (2020), which combines multi-epoch HST imaging with the Wide Field Camera 3 (WFC3) in F814W with previous UVIS-IR channel WFC3 observations of Wd2 in the F160W filter (PI Nota, GO-13038). This photometric catalogue entails 9,267 sources in total, 6,268 of which are attributed to the Wd2 cluster, whereas the remainder are tentatively characterised as LMS foreground or background stars contaminating the line of sight. The left panel of Figure 1.4 in Section 1.3 shows the CMD of the cluster constituents. The catalogue is likely missing some of the most massive stars of Wd2, e.g. high-mass UMS sources, as they were saturated in the long 350s exposures in the F814W filters. In addition to the photometric catalogue of Sabbi et al. (2020) we make use of a gas extinction map of Wd2 created by Zeidler et al. (2015).

2.2 NGC 6397
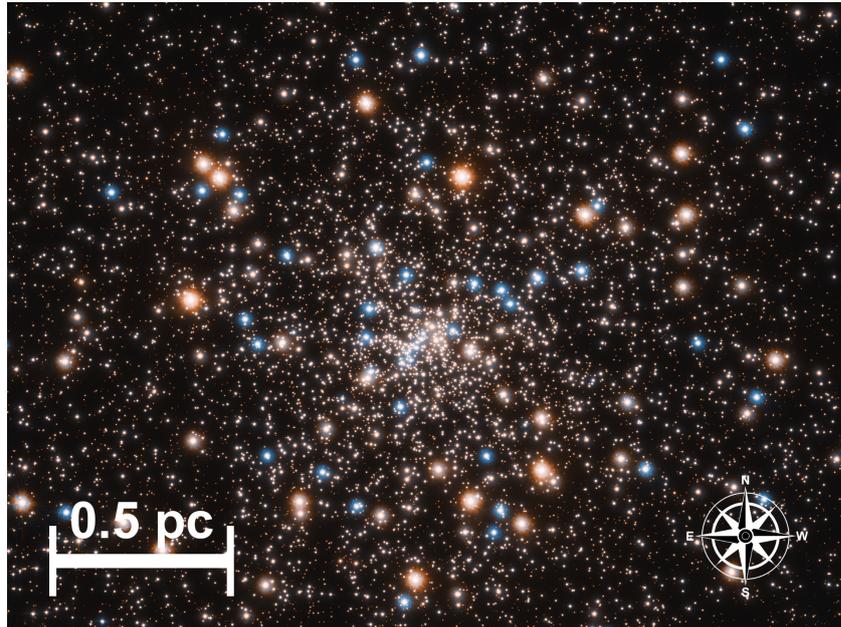
Fig. 2.2.: HST two colour composite image of the globular cluster NGC 6397. Original image by NASA, ESA, and T. Brown and S. Casertano (STScI). Size scale and orientation indicators added for this thesis.

The second target analysed in Section 3.2 is the old globular cluster NGC 6397 (Figure 2.2). At a distance of $2.39 \pm 0.17\,\mathrm{kpc}$ (Brown et al., 2018) NGC 6397 is located relatively close to our Sun and characterised by a very low metallicity of $[\mathrm{Fe/H}] = -2.02$ (Kraft and Ivans, 2003; Vulic et al., 2018). Recent studies indicate a cluster age between 12.6 ± 1.0 Gyr (Correnti et al., 2018) and 13.4 ± 1.9 Gyr (Brown et al., 2018). NGC 6397 is subject to only moderate extinction, as several studies place its colour excess $E(B-V)$ at a value between 0.183 mag (Gratton et al., 2003), 0.186 mag (Schlegel et al., 1998) and 0.187 mag (Anthony-Twarog et al., 1992). Given these properties, NGC 6397 is an excellent example for ancient metal-poor stellar populations. Study of metal-poor clusters like NGC 6397, in particular dating them accurately, may provide crucial insights into their formation environments (Correnti et al., 2018).

In Section 3.2 we employ the photometric catalogue of NGC 6397 provided by the *HST UV Globular Cluster Survey* (HUGS, Nardiello et al., 2018; Piotto

et al., 2015). The HUGS catalogues entail broadband photometry in five HST filters, i.e. F275W, F336W and F438W obtained with the WFC3 (UVIS channel), and F606W, F814W imaged with the wide field channel (WFC) of the Advanced Camera for Surveys (ACS). Following the pre-processing prescription detailed in Section 3 of Nardiello et al. (2018), the HUGS photometric catalogue of NGC 6397 entails a total of 4,831 stars (see also the right panel of Figure 1.4 for an example CMD of the NGC 6397 data).

2.3 The Large Magellanic Cloud
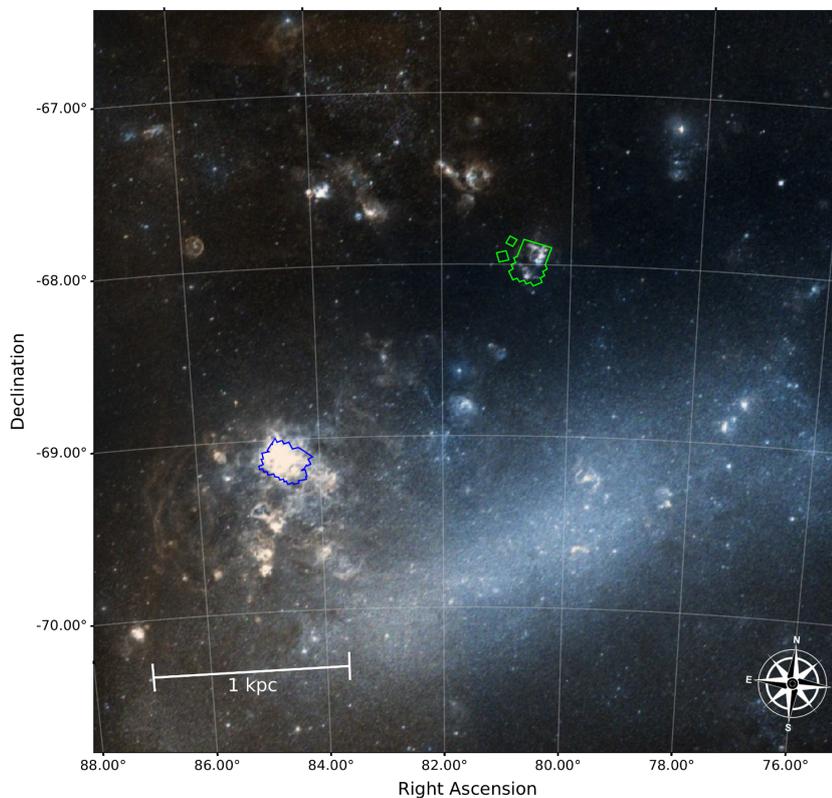
Fig. 2.3.: Colour composite image of the LMC in the neighbourhood of N44 and the Tarantula Nebula from the Digitized Sky Survey (Lasker et al., 1996). The blue and green outlines highlight the FoVs of the HTTP and MYSST surveys, presented in Sections 2.4 and 2.5, respectively. Image adapted from Ksoll et al. (2021b) with permission.

Two of our target regions in Chapter 3 are not located in our Milky Way, but in one of its dwarf satellite galaxies, the Large Magellanic Cloud (LMC). The

LMC is a gas-rich, one-armed spiral galaxy with an off-centre bar, prototyping the 'Magellanic Irregulars' class of galaxies (Besla et al., 2012). Similar to (but not quite to the extend of) its own smaller companion – the Small Magellanic Cloud (SMC) – the LMC is characterised by a very low metallicity environment. With $Z \simeq 1/3\ Z_\odot$ (Hill et al., 1995) the LMC exhibits conditions comparable to those in the early universe at the peak of the cosmic star formation activity, $z \sim 1.5$ (Madau et al., 1996; Pei et al., 1999), providing the best templates in the local universe for primitive star formation (Gouliermis, 2012). In addition, the LMC is subject to an exceptional star formation activity, hosting many large star-forming complexes, as well as the most impressive starburst in the Local Group, 30 Doradus. Located at a moderate distance of about 50 kpc (de Grijs et al., 2014; Pietrzyński et al., 2013; Panagia et al., 1991) and above the Galactic disc, the LMC can be observed at low interstellar extinction (Gordon et al., 2003) as it is largely unobscured by the Milky Way's gas and dust. This makes the LMC very attractive for photometric and spectroscopic surveys, providing one of the few opportunities (along with the SMC) to observe and resolve individual stars down to the very low-mass limit outside of our own galaxy. Given these properties, the LMC is a perfect target to study stellar birth by finding the PMS constituents of the star-forming complexes, recovering their star formation histories and investigating their stellar IMFs.

In the following, two such studies are presented, targeting the Tarantula Nebula and LHα 120–N44. Figure 2.3 shows a colour composite image of the LMC neighbourhood close to these two regions from the Digitized Sky Survey (Lasker et al., 1996). This figure also highlights the exact position of both regions as per the observed FoVs of the respective studies presented in Sections 2.4 and 2.5.

2.4 The Tarantula Nebula

The Tarantula Nebula, also known as 30 Doradus (shown in Figure 2.4), is the subject of Section 3.1. Located in the LMC, it is both the most massive and most luminous ($\log(L/\mathrm{erg\ s^{-1}}) = 39.66$; Pellegrini et al., 2012) star-forming

Fig. 2.4.: HST colour composite image of the Tarantula Nebula located in the LMC from observations taken as part of the Hubble Tarantula Treasury Project. Original image by NASA, ESA, E. Sabbi (STScI). Size scale and orientation indicators added for this thesis.

complex known in the Local Group (Kennicutt and Hodge, 1986). With an estimated area of $\sim 40.000 \ \mathrm{pc}^2$ the Tarantula Nebula is an enormous H II region, comparable to the luminous but unresolved H II structures that are observed in distant galaxies (Sabbi et al., 2013; Hunt and Hirashita, 2009; Oey et al., 2003). 30 Doradus is characterised by an exceptional star-forming activity, in particular in the core Radcliffe 136 (R136) of its ionising cluster NGC 2070 (Sabbi et al., 2013). R136 is often equated to a starbust-like environment (Sabbi et al., 2013) and considered a local template of unresolved starburst knots found in interacting galaxies at high-redshift due to its size and high rate of star formation (Schneider et al., 2018; Crowther et al., 2017; Heckman et al., 2004; Shapley et al., 2003; Meurer et al., 1997). The unique environment of the Tarantula Nebula is further highlighted by a collection of extreme objects harboured within the complex, such as the most massive known stars to date (200–300 $M_\odot$, Hainich et al., 2014; Bestenlehner et al., 2011; Crowther et al., 2010), the fastest rotating star (VFTS 102, Dufton et al., 2011) and the most energetic young pulsar (PSR J0537–910) inside the expanding supernova remnant N157B (Chen et al., 2006). Besides

the striking cluster R136 at the centre of NGC 2070, 30 Doradus is host to several other known stellar clusters, i.e. Hodge 301, SL 639 and NGC 2060, the latter itself harbouring cluster TLD 1 (Schneider et al., 2018). With all these centres where stellar birth has taken and still is taking place, the Tarantula Nebula is subject to a complex star formation history, which is still being extensively studied. NGC 2070 for instance exhibits a prolonged star-forming activity over the past 20 Myr (Cignoni et al., 2015; Walborn and Blades, 1997), characterised by several episodes (De Marchi et al., 2011) and a peak activity about 1–3 Myr ago (Cignoni et al., 2015). Hodge 301 on the other hand is much older, with an estimated age between 26.5 and 31.5 Myr (Cignoni et al., 2016). It has also been suggested that there may be sites of star formation between R136 and Hodge 301 that are triggered by the stellar feedback of these clusters, as observations find several massive O stars in dense knots of dust in between them (Brandner et al., 2001). With its complexity, being a multi-stage star-forming environment with dense stellar clusters of different ages and loose associations closely coexisting, the Tarantula Nebula provides an exceptional test bed for various scenarios of star formation in clusters (Sabbi et al., 2013).

The *Hubble Tarantula Treasury Project* (HTTP, Sabbi et al., 2013, 2016) is a large, deep, high spatial resolution, HST photometric survey that targets the Tarantula Nebula to study the star formation history of the region. The survey's field of view (FoV) captures the entire nebula, in particular the clusters NGC 2070 (with R 136 in its centre), NGC 2060 and Hodge 301, and covers an area of $16 \times 13 \ \mathrm{arcmin}^2$ on the sky, which corresponds to about $240 \times 190 \ \mathrm{pc}^2$ at the distance of the LMC. Employing both the ACS (WFC) and WFC3 (UVIS and IR channels) on board HST, the HTTP observations cover a wide spectral range from 0.27 to $1.6 \ \mu\mathrm{m}$ in one narrow- and six broad-band filters, i.e. F275W, F336W, F555W, F658N, F775W, F110W and F160W. The final HTTP stellar photometric catalogue entails a total of 822,204 stars with a detection in at least one of the seven filters. Both the spatial coverage and the fraction of detected stars varies strongly between the individual filters of the survey. The most stars are detected (taking photometric flag ≤ 2 as the detection criterion, see Sabbi et al., 2016) in the F775W, F160W and F110W filters with 79.9%, 75.2% and 75.1% of the total 822,204 sources covered. The least amount of sources are found in the UV (F275W), $H\alpha$

(F658N) and U (F336W) bands with detection rates of 5.6%, 16.1% and 18.5%, respectively (see also Table 1 of Ksoll et al. (2018) in Section 3.1). The largest spatial coverage is achieved in the F555W and F775W filters, whereas the least amount of the observed FoV is covered in F275W. Even though the detection rate is very high in the F110W and F160W filters, they also miss a large part of the FoV at the western edge, so that the best filter combination in terms of spatial coverage and detection rate is F555W and F775W with a total of 403,018 detected sources.

2.5 LHα 120–N44

Fig. 2.5.: HST two-colour composite image of the characteristic superbubble of the giant star-forming complex N44 in the LMC. This image was compiled from the MYSST observational data (Ksoll et al., 2021a) for this thesis.

LHα 120–N44 (or N44 for short; Henize, 1956) is the subject of Sections 3.3 and 3.4, and like the Tarantula Nebula an active giant star-forming region situated in the LMC (green outline in Figure 2.3). Although is is not quite as extreme an environment as the Tarantula Nebula, N44 is one of the most luminous giant complexes of H II regions in the entire LMC (just behind 30 Doradus and N11; Pellegrini et al., 2012; Kennicutt and Hodge, 1986). The

most striking characteristic feature of N44 is an enormous central super-bubble, shown in Figure 2.5, with several compact H II regions along the bubble's rim (McLeod et al., 2019; Pellegrini et al., 2012). This bubble is driven by the feedback of massive stars in its interior, namely the OB association LH 47 (Lucke and Hodge, 1970), clearing out the gas and dust of their natal environment. The stellar feedback at play becomes evident in X-ray observations, revealing very hot gas ($T \sim 10^6$) heated by stellar winds and supernovae (Jaskot et al., 2011). CO observations indicate that N44's star-forming activity stems from one giant molecular cloud complex (Wong et al., 2011; Fukui et al., 2001), which entails an intricate hierarchical ISM structure visible in dust maps from the Herschel Space Telescope (Hony et al., 2010).

Several studies highlight the youthfulness of N44's cluster constituents, revealing another two OB associations (LH 48 and 49; Lucke and Hodge, 1970), along with more than 30 spectroscopically confirmed massive O-type stars (McLeod et al., 2019; Will et al., 1997; Oey and Massey, 1995; Conti et al., 1986; Rousseau et al., 1978) and about 180 YSOs captured by observations with the Spitzer Space Telescope (Carlson et al., 2012; Chen et al., 2009). It is also possible that the stellar birth environment of N44 is an example for feedback-triggered star formation, as Oey and Massey (1995) have identified a 5 Myr age difference between the population of massive stars within N44's bubble and those found in the bubble's rim. Additionally, there is one supernova remnant (SNR 0523-679; Chu et al., 1993) located in the vicinity of the superbubble (with up to four supernovae estimated in the past; Oey and Massey, 1995) that exhibits characteristics of a core-collapse supernova (Jaskot et al., 2011). Jaskot et al. (2011) also find evidence for metallicity enhancement in the superbubble. Aside from the massive O-type main-sequence stars, the recent Zivkov et al. (2018) study (c.f. Section 1.6) also confirms the presence of PMS stars in N44. They determine a lower limit for the number of PMS sources in N44 of 1000 ± 38. In summary, with its complex hierarchical ISM structure and multiple centres of star formation at different evolutionary stages, N44 is an extraordinary example of an actively star-forming ecosystem.

The '*Measuring Young Stars in Space and Time*' program (MYSST, GO14689, P.I. D. Gouliermis), which we introduce in Ksoll et al. (2021a) in Section 3.3, conducts a deep, high spatial resolution photometric survey of N44 with the HST. With a FoV of $12.2 \times 14.7 \, \mathrm{arcmin}^2$, or about $180 \, \mathrm{pc} \times 215 \, \mathrm{pc}$ at the distance of the LMC, the survey captures N44's main superbubble, the system of star-forming centres south of the bubble and two smaller reference fields north-east of N44 (see green outline in Figure 2.3 for the position of the MYSST FoV in the greater LMC neighbourhood). Employing two broadband filters, F555W and F814W, from the ACS and WFC3 HST instruments, the MYSST observations reach down to 29 mag in F555W and 28 mag in F814W, making the MYSST survey one of the deepest views of extragalactic stars. Capturing e.g. unreddened 1 Myr old PMS stars with masses as low as $0.09 \, M_\odot$ (corresponding to the F555W detection limit), the MYSST program reveals even the lowest-mass constituents of N44. The final MYSST photometric catalogue, presented in Section 3.3 consists of 461,684 sources in total across the main FoV and the two offset reference fields. As extremely bright sources went into saturation, this catalogue does not entail objects brighter than 14 mag in F555W and 13 mag in F814W, therefore likely excluding the most massive stars of N44, i.e. early O-type stars.

3

Publications

In the following we present the publications that make up the core of this cumulative thesis. They appear unedited and in order of publication. The manuscripts 3.1 and 3.2 have been published in the journal 'Monthly Notices of the Royal Astronomical Society' (MNRAS), whereas 3.3 and 3.4 have been accepted for publication in the 'Astronomical Journal' (AJ), but are at the time of this writing still in production. Note that the page numbering for each manuscript follows that of the journal it appears in.

3.1 Hubble Tarantula Treasury Project – VI. Identification of pre-main-sequence stars using machine-learning techniques

Ksoll et al. (2018) presents a methodology to identify young PMS stars from HST photometry employing classical ML classification techniques with an explicit application to the HTTP survey (cf. 2.4). Note that the analysis presented in this publication has been carried out primarily over the course of my master thesis, but has been extended into the first part of the dissertation. It is included here mainly for *completeness* as it introduces a methodological approach that is applied and further tested in our subsequent study Ksoll et al. (2021b), presented in Section 3.4. Beyond of what has been presented in the master thesis we have updated the training set by including additional highly-extincted RGB examples and subsequently repeated the ML classification analysis to produce the final catalogue of PMS candidate sources as shown in Ksoll et al. (2018). Furthermore, we have extended the

ML-related performance assessment of our approach by including the balanced accuracy measure and we have added an analysis of the 2D density distribution using kernel density estimation of the identified PMS candidates, both of which were not part of the original master thesis. The preparation of the publication and the corresponding repetition of the ML analysis also required a revision of the figures and diagrams compared to the master thesis.

Contributions: Dimitrios Gouliermis first proposed the project to investigate the application of ML classification techniques to identify PMS stars using HST photometry, in particular on the data from the HTTP collaboration of the Tarantula Nebula. Under the guidance of Dimitrios Gouliermis and Ralf Klessen I carried out the ML analysis of this project, devising the strategy for construction and labelling of the training data, selecting, training and testing several ML classification approaches, and applying the final models to the HTTP photometric catalogue. I created all figures for the publication and wrote most of the text, excluding the introduction, which was provided by Dimitrios Gouliermis. All co-authors provided feedback and suggestions for the manuscript and assisted with proofreading.

Hubble Tarantula Treasury Project – VI. Identification of pre-main-sequence stars using machine-learning techniques

Victor F. Ksoll,[1,2]★ Dimitrios A. Gouliermis,[1,3]★ Ralf S. Klessen,[1] Eva K. Grebel,[4] Elena Sabbi,[5] Jay Anderson,[5] Daniel J. Lennon,[6] Michele Cignoni,[7] Guido de Marchi,[8] Linda J. Smith,[9] Monica Tosi[10] and Roeland P. van der Marel[5]

[1]*Institut für Theoretische Astrophysik, Zentrum für Astronomie der Universität Heidelberg, Albert-Ueberle-Str 2, D-69120 Heidelberg, Germany*
[2]*Interdisciplinary Center for Scientific Computing, University of Heidelberg, Mathematikon, Im Neuenheimer Feld 205, D-69120 Heidelberg, Germany*
[3]*Max Planck Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany*
[4]*Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Mönchhofstr 12-14, D-69120 Heidelberg, Germany*
[5]*Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA*
[6]*ESA – European Space Astronomy Center, Apdo. de Correo 78, E-28691 Associate Villanueva de la Cada, Madrid, Spain*
[7]*Department of Physics, University of Pisa, Largo Pontecorvo 3, I-56127 Pisa, Italy*
[8]*European Space Research and Technology Centre, Keplerlaan 1, NL-2200 AG Noordwijk, the Netherlands*
[9]*European Space Agency and Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA*
[10]*INAF-Osservatorio Astronomico di Bologna, Via Ranzani 1, I-40127 Bologna, Italy*

ABSTRACT

The Hubble Tarantula Treasury Project (HTTP) has provided an unprecedented photometric coverage of the entire starburst region of 30 Doradus down to the half Solar mass limit. We use the deep stellar catalogue of HTTP to identify all the pre-main-sequence (PMS) stars of the region, i.e. stars that have not started their lives on the main-sequence yet. The photometric distinction of these stars from the more evolved populations is not a trivial task due to several factors that alter their colour–magnitude diagram positions. The identification of PMS stars requires, thus, sophisticated statistical methods. We employ machine-learning classification techniques on the HTTP survey of more than 800 000 sources to identify the PMS stellar content of the observed field. Our methodology consists of (1) carefully selecting the most probable low-mass PMS stellar population of the star-forming cluster NGC 2070, (2) using this sample to train classification algorithms to build a predictive model for PMS stars, and (3) applying this model in order to identify the most probable PMS content across the entire Tarantula Nebula. We employ decision tree, random forest (RF), and support vector machine (SVM) classifiers to categorize the stars as PMS and non-PMS. The RF and SVM provided the most accurate models, predicting about 20 000 sources with a candidateship probability higher than 50 per cent, and almost 10 000 PMS candidates with a probability higher than 95 per cent. This is the richest and most accurate photometric catalogue of extragalactic PMS candidates across the extent of a whole star-forming complex.

Key words: methods: data analysis – methods: statistical – Hertzsprung–Russell and colour–magnitude diagrams – stars: pre-main-sequence – Magellanic Clouds – galaxies: star clusters: individual: NGC2060, NGC2070.

1 INTRODUCTION

Giant star-forming regions, the signposts of star formation across whole giant molecular clouds (GMCs), are one of the major birth-places of stars in a galaxy. The youthfulness of these regions is shown by their blue massive stars, located at the bright part of the main sequence in the colour–magnitude diagram (CMD). However, also the stars of low and intermediate masses in these regions hold important information about the star formation process. These stars are not yet fully formed, and since they do not fuse hydrogen in their cores, they do not appear on the main sequence like their massive counterparts (Schulz 2012). These so-called *pre-*

★ E-mail: v.ksoll@stud.uni-heidelberg.de (VFK); gouliermis@uni-heidelberg.de (DAG)

Table 1. Filter wavelength coverages, used *HST* instruments, and available data in individual filters out of the 822 204 total stars of the HTTP catalogue. Sources with photometric flags (as described in Sabbi et al. 2016) higher than 2 in certain filters are considered as non-detections. The last column refers to the fraction of stars out of the total amount detected in the respective filter.

Filter	λ_{mean} (Å)	λ_{min} (Å)	λ_{max} (Å)	*HST* instrument	Available data
F275W (UV)	2377.6	1990	2980	WFC3_UVIS1	46 215 (5.6%)
	2363.9	1990	2968	WFC3_UVIS2	
F336W (U)	3358.5	3014	3707	WFC3_UVIS1	151 679 (18.5%)
	3358.6	3014	3707	WFC3_UVIS2	
F555W (V)	5396.7	4584	6209	ACS_WFC	409 042 (49.8 %)
	5397.5	4584	6209	ACS_WFC	
F658N (H$_\alpha$)	6584.1	6510	6659	ACS_WFC	132 496 (16.1 %)
F775W (R)	7729.7	6804	8632	ACS_WFC	657 279 (79.9 %)
	7730.6	6804	8632	ACS_WFC	
	7660.1	6869	8571	WFC3_UVIS1	
	7658.4	6869	8571	WFC3_UVIS2	
F110W (J)	11 623.8	8832	14 121	WFC3_IR	617 129 (75.1%)
F160W (H)	15 392.3	13 854	16 999	WFC3_IR	618 508 (75.2 %)

Figure 1. Overview of the spatial coverage in the individual HTTP filters. The grey points indicate the objects not covered in the respective filter, while the blue ones mark objects with a detection.

our understanding of clustered star formation by parametrizing its properties, such as the star formation efficiency, rate, duration, and the low-mass end of the resulting stellar IMF. The characterization of PMS stars, with T Tauri stars being prototypical examples, can be achieved with spectroscopic measurements of usually only a small number of objects (see e.g. Bodenheimer 2011, and references therein for the spectral features of these stars). However, characterizing whole ensembles of such stars with spectroscopy is not practically feasible for star-forming regions in the Galaxy, nor even possible for those outside the Galaxy. The study of rich samples of faint PMS stars has, thus, to rely on deep photometric measurements.

More than a decade ago several studies showed the exceptional ability of the *Hubble Space Telescope* (*HST*) to detect faint PMS stars in the Magellanic Clouds, the satellite galaxies of our Milky Way (Brandner et al. 2001; Gouliermis, Brandner & Henning 2006; Nota et al. 2006; Sabbi et al. 2007). These are the only extragalactic PMS stars we can resolve and they are extremely useful in understanding star formation at GMC length-scales because of their cospatial, rich samples, unobscured by the dusty Galactic disc. The Magellanic Clouds, due to their exceptional star formation activity at much lower extinction and stellar field contamination than our Milky Way are, thus, very attractive targets for the study of ensembles of PMS stars. Nevertheless, there are various observational and physical constraints that introduce difficulties to the identification and characterization of PMS stars from photometry alone. The main issue is the dislocation of the stars from their theoretical CMD positions, introducing a spread of the PMS stars in the CMD along both the brightness and colour axes. Among the effects that produce the PMS spread on the CMD, differential reddening, variability, and excess emission due to circumstellar discs are considered the most important (Gouliermis 2012). Moreover, observational limitations, such as unresolved binarity, stellar crowding, and photometric uncertainties introduce a 'mixing' of the PMS stars' CMD positions.

These issues can be mitigated in (almost) single-age individual young clusters and associations (Gouliermis et al. 2007, 2010), allowing the investigation of their star formation history (e.g. Massey & Hunter 1998; Da Rio, Gouliermis & Henning 2009; Cignoni et al. 2010), stellar IMF (Sabbi et al. 2008; Schmalzl et al. 2008; Da Rio, Gouliermis & Gennaro 2010), structure (Schmeja, Gouliermis & Klessen 2009; Gouliermis, Hony & Klessen 2014), and star formation rate (Hony et al. 2015) from their apparently co-

main-sequence (PMS) stars are still under formation through contraction and accretion and they occupy the faint red part of the CMD (Stahler & Palla 2005). Assuming a typical stellar initial mass function (IMF), PMS stars with masses up to a few M$_\odot$ account for almost half of the total stellar mass budget of a young star cluster. Consequently, identifying and studying these stars improves

Figure 2. The CMD of the data obtained in the optical *F555W* and *F775W* (*V*-, *R*-equivalent) filters. It is the richest CMD of the HTTP survey in terms of spatial coverage. For reference representative errorbars of the photometric errors are overplot in red and the coloured labels indicate the approximate locations of populations of interest.

eval PMS stellar populations. However, the PMS spread across the CMD introduces a significant difficulty in disentangling these stars from the evolved stellar populations of the galactic field or older clusters across giant star-forming complexes. These structures of gas and stars host multiple star-forming centres, still embedded in their gaseous natal environments, with a significant amount of evolved field populations being projected on to their field of view (FoV).

A typical example of such regions is the *Tarantula Nebula* in the Large Magellanic Cloud (LMC), the natal environment of 30 Doradus, the most impressive starburst in the Local Group. The complexity of the region, indicated in various previous investigations, is revealed at its full extent with the deep panchromatic *HST* imaging of the whole nebula from the *Hubble Tarantula Treasury Project* (HTTP; Sabbi et al. 2013, 2016). The HTTP field contains different populations of various ages, so that an overlap between the turn-on, i.e. the locus in the CMD where the PMS joins the ZAMS (Cignoni et al. 2010), and faint giant/subgiant regions of the CMD occurs. These factors become evident in the particularly rich (in stellar numbers) CMD of the survey, constructed from measurements in the *F555W* and *F775W* (*V*- and *R*-equivalent) filters. The wide broadening of the upper main-sequence (UMS) stars and the elongation of the red clump (RC) (which in theory should appear nearly circular) provide clear evidence for significant differential extinction (see also De Marchi et al. 2016). In this CMD it is also practically difficult to distinguish the faint lower main-sequence (LMS) field stars from the faint PMS stars, as these two popu-

lations strongly overlap in the low-brightness regime. Because of these effects that prevent the clear identification of the various populations based on their CMD positions, the distinction of the faint PMS stars requires the use of sophisticated statistical approaches. In this study we develop a classification methodology by employing *machine-learning* techniques in order to perform a robust identification of the most probable PMS stars in the Tarantula Nebula. The clear sample of faint PMS stars in combination with the young bright stellar population of the UMS will provide an unprecedented stellar data set to investigate the complexity of the star formation process along the most interesting star-forming region in the Local Group.

Machine learning, i.e. the study of algorithms that can learn from and make predictions on data, has introduced a variety of statistical tools to the astronomical research. These tools are designed to solve problems of regression, classification, and clustering (see e.g. Beaumont et al. 2014; Dieleman, Willett & Dambre 2015; Elorrieta et al. 2016, for recent applications). While the first two types of problems are addressed with *supervised* learning processes, the third requires *unsupervised* methods. We employ classical machine-learning classification techniques on the HTTP photometry in order to determine the most successful in identifying the most probable PMS stars across the nebula, based on prior information retrieved from our data of the most prominent faint PMS stellar sample in our FoV. This sample comprises the stellar members of the giant Hɪɪ region, where the young stellar cluster NGC 2070, host of the starburst cluster R136, resides.

This paper is structured as follows. In Section 2 we give a short description of the photometric data set of HTTP. In order to account for the differential reddening of the region in our identification, we apply a correction for extinction to the photometric measurements of all stars in Section 3. This correction is estimated in terms of the proximity of PMS candidates to UMS stars, for which reddening is being determined from their dislocation from the unreddened main sequence. In preparation of our experiments on various classification methods we build a so-called training data set, which is made with a careful selection of the LMS and PMS stars included in the area of NGC 2070, as well as the 'contaminating' evolved field populations (Section 4). This is necessary because of the partial overlap between the LMS and PMS stars in the CMD and the extended star formation history of R136 of several Myr (Hunter et al. 1995). The classification of the observed stellar populations in terms of supervised machine learning based on the training data set takes place in Section 5, where the most robust algorithms for the identification of faint PMS stars are established, and the final data set of the best PMS candidates across the whole Tarantula Nebula is constructed. A summary and future prospects concerning this study are given in Section 6.

2 DATA DESCRIPTION

HTTP is a high spatial resolution stellar photometric survey of the Tarantula Nebula (the nebula of the starburst of 30 Doradus). It has a high dynamic range in spectral coverage, extending from the near-ultraviolet (NUV) to the near-infrared (NIR) part of the spectrum (Sabbi et al. 2013, 2016). Its spatial coverage extends across the whole region of the nebula of $\sim 16 \times 13$ arcmin^2, corresponding to 240×190 pc^2 at the distance of the LMC $((m - M)_0 = 18.55 \pm 0.05$ mag). This region includes the clusters NGC 2070 (with the starburst cluster R 136 at its core), NGC 2060 and Hodge 301 (see e.g. fig. 1 in Sabbi et al. 2016). The observations were obtained with the *Advanced Camera for Surveys* (ACS) and
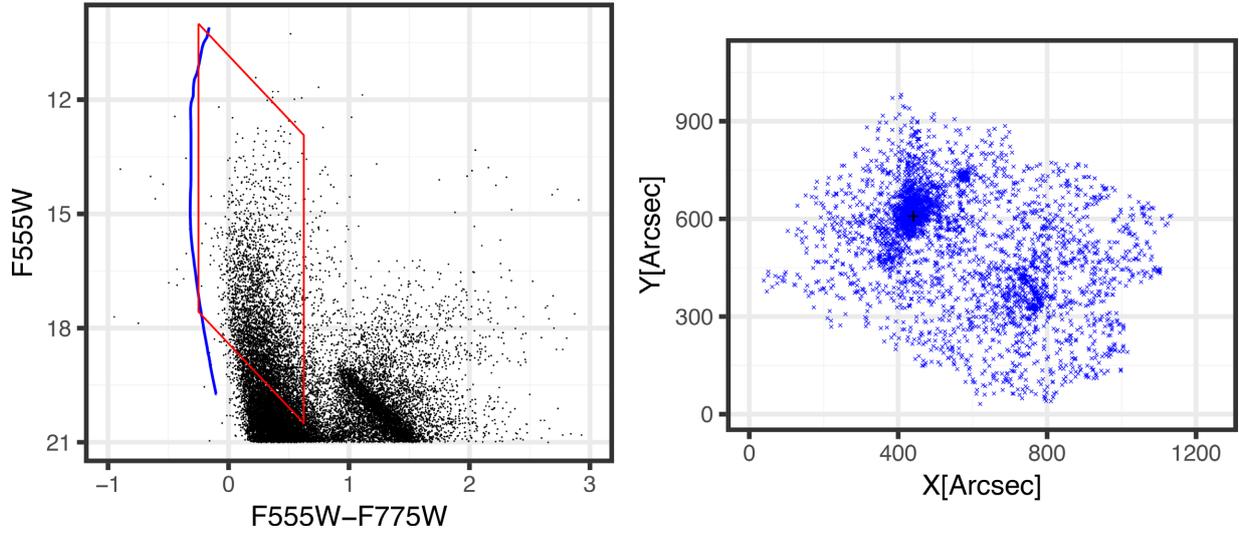
Figure 3. *Left:* The bright part of the optical CMD with the region, where the UMS stars are selected for the measurement of their extinction enclosed by the polygon in red. The blue line corresponds to the upper part of the *un-reddened* ZAMS, corrected for the distance of the LMC. The top and bottom sides of the polygon follow the direction of the reddening vector. *Right:* The spatial distribution of the UMS stars selected within the polygon for the determination of the extinction across the HTTP FoV. The black cross marks the centre of R 136 (RA$_{J2000}$ = 05^h 38^m 42^{s}3, Dec.$_{J2000}$ = −69° 06′ 03″3) to be used later in the selection of our training sample.

Wide Field Camera 3 (WFC3) on board the *HST* in filters corresponding to a wide range of broad and narrow bands of 0.27–1.6 μm (Table 1). An overview over the available data in the respective filters is given in Table 1, where it can be seen that there are significantly less data available in the *UV*, *U*, and *H$_\alpha$* filters than in the remaining four filters, with the most data being available in the *R*, *J*, and *H* bands. Fig. 1 shows the spatial coverage of the observations in each filter, indicating that filters *V* and *R* cover the largest area of the observed field, while especially the *UV* filter covers only a very sparse area and both infrared filters do not cover the South-Eastern region of the observed area. Fig. 2 shows the optical CMD of the HTTP survey, exhibiting the previously mentioned widened UMS and elongated RC caused by differential extinction (see Haschke,
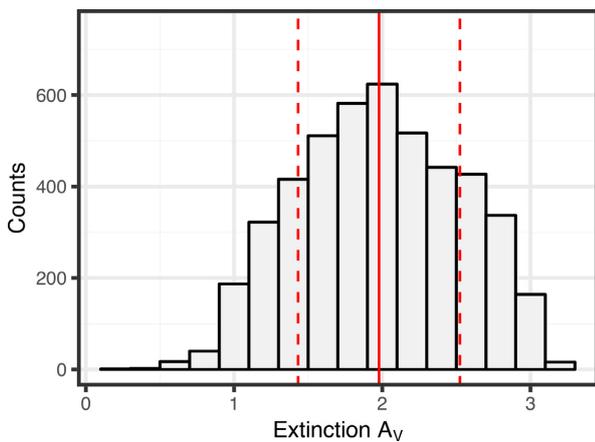
Grebel & Duffau 2011; De Marchi et al. 2016, for a discussion of the reddening in this area and maps).

3 EXTINCTION CORRECTION

Star-forming regions such as the Tarantula Nebula still include a significant amount of gas from the parental molecular cloud. It is, thus, expected that a considerable amount of interstellar extinction influences the photometric measurements of the observed stars and dislocates them accordingly in the CMD. This phenomenon contributes to the mixing of old and young stars in the CMD and it should be considered in the application of our classification procedure for the identification of PMS stars. We increase, thus, the number of variables by adding to the magnitudes and colour indexes of the stars their reddening, as determined by the extinction measurements of their close-by UMS stars (Panagia et al. 2000; Romaniello et al. 2002; De Marchi et al. 2016; De Marchi, Panagia & Beccari 2017). The use of UMS stars for the extinction correction is based on the fact that the young PMS candidates are more likely to be spatially correlated with the UMS population, than with the LMC field. In addition, earlier extinction studies showed that 'dust is highly localized near the hotter, younger stars' and the average extinction correction for older populations is lower than that for younger (Zaritsky et al. 2002). The use of RC stars for the extinction correction of PMS stars would, thus, compromise the accuracy of this correction and the reddening of the related star-forming regions.

We retrieve extinction measurements for the UMS stars in our catalogue by 'relocating' their CMD positions on the main sequence along the known reddening vector of the Tarantula Nebula with slope $R_{555} = A_{555}/E(m_{555} − m_{775}) = 3.35 \pm 0.15$, as determined by De Marchi et al. (2016), up to its intersection with the zero-age main sequence (ZAMS). We make use of PARSEC family of evolutionary models (Bressan et al. 2012) with a metallicity of $Z = 0.08$ for the LMC, corrected for a distance modulus of 18.55 mag (Panagia et al. 1991; Walborn & Blades 1997; De Marchi et al. 2016).

Figure 4. Histogram of the measured extinction values of the selected UMS extinction probes in the HTTP data, where the thick red line indicates the mean $A_{v,mean} = 1.98$ mag and the dotted lines the standard deviation $\sigma_{A_V} = 0.55$ mag.
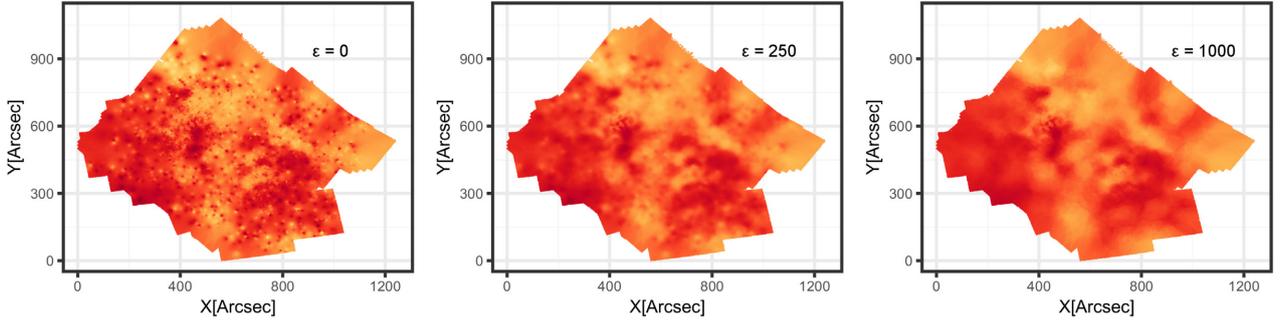
Figure 5. A series of extinction maps of the Tarantula Nebula, where each data point is colour-coded according to its assigned extinction value. From left to right, the series shows the influence of the smoothing parameter ϵ, introduced to the distance weight in equation (2) on the extinction map. The respective ϵ value is noted in the top right corner of each plot. Note that all plots share the colour gradient from Fig. 6. The adapted extinction law is described in De Marchi et al. (2016).

Fig. 3 (left) shows the selected UMS extinction probes. This selection is based on the CMD region occupied by these stars, as indicated by the red polygon, entailing 4605 stars distributed across the entire observed FoV, and identified in the V band. The selected sample of UMS stars exhibits higher concentrations in the regions corresponding to the clusters R136, Hodge 301, and NGC 2060, as demonstrated by their map also shown in Fig. 3 (right). We assign extinction values to each non-UMS star as the distance-weighted average of the extinctions $A_{V_n}^{(UMS)}$ of its N nearest UMS neighbours according to

$$A_V = \sum_{n=1}^{N} w_n A_{V_n}^{(UMS)} \tag{1}$$

with weights

$$w_i = \frac{1}{d_i^2 + \epsilon^2} \frac{1}{\sum_{n=1}^{N} \frac{1}{d_n^2 + \epsilon^2}}, \tag{2}$$

where d_i denotes the Euclidean distance in pixels to the ith nearest UMS neighbour and ϵ is a smoothing parameter (also given in pixels), which we introduce in order to reduce the dominance of close proximity to a single UMS star in the averaging process. The corresponding weighted standard deviation of the assigned average extinction value is given by

$$\delta A_V = \sqrt{\sum_{i=1}^{N} w_i (A_{V_i}^{(UMS)} - A_V)^2}. \tag{3}$$

This distance weighted A_V calculation eliminates the possibility of underestimation due to low-extinction foreground stars that may be projected by chance in the region of the PMS candidates. Nevertheless such stars represent an insignificant fraction of our stellar sample. The extinction measurements of the UMS probes are summarized in Fig. 4, where it is shown that these stars have a mean extinction of $A_{v, mean} = 1.98$ mag with a standard deviation of $\sigma_{A_V} = 0.55$ mag.

For the assignment of extinction values we use the $N = 20$ nearest UMS neighbours. Fig. 5 shows a series of A_V maps, as constructed for various values of ϵ. These maps are generated by colour-coding each data point in the spatial distribution plot according to its assigned (or measured for UMS stars) extinction value. It should be noted that the assigned extinction values in the regions without V-band coverage are biased towards the UMS stars at their borders and, thus, might not necessarily represent the true extinction within these regions. They are shown here only for visualization, and they

are not included in our further analysis. For the final estimation of the extinction corrections, and the construction of the final A_V of the region, we choose a smoothing parameter of $\epsilon = 500$, based on the natural appearance of the constructed extinction map, i.e. a map which is not oversmoothed and still provides spatially detailed A_V measurements (Fig. 6). It should be noted that the assigned A_V measurements of the majority of the stars are found to be insensitive to the chosen value for the smoothing factor, with the relative differences not exceeding 5–10 per cent for ϵ between 0 and 2000 px (in comparison with $\epsilon = 500$ px).

The extinction-corrected optical CMD of the Tarantula Nebula is shown in Fig. 7. In order to provide a more realistic appearance of this CMD, we applied artificial noise to the corrected positions of our UMS probes, based on small random dislocations along the reddening vector by amounts sampled from a Gaussian distribution with zero mean and the standard deviation of the colour indexes of all stars within the same magnitude range as the selected UMS probes. Our extinction correction demonstrates some over- or underestimation for the RC stars, as the remaining elongation of the RC shown in Fig. 7 indicates. However, this does not affect our classification, because the CMD positions of these stars do not overlap with those typically occupied by our target PMS stars.

4 BUILDING THE TRAINING SET

The machine-learning algorithms applied in this study are based on *supervised learning* techniques, i.e. they infer a function from labelled training data, which consist of a set of training examples (see Appendix A). These techniques require the construction of a labelled *training data set* in order to 'teach' the algorithms, in our case, how to identify PMS stars based on their positions in the CMD. Also, our study aims at the simple distinction between two classes of objects, namely PMS and non-PMS stars, i.e. we address a *binary classification problem*. Considering these, we build our training set so that each star has a label, which indicates whether it is a PMS example or not. We train our algorithms directly on the observational data of the HTTP data set, rather than on synthetic populations. The advantage of this approach is that modelled populations of stars would assume a specific behaviour of observable characteristics, such as binarity, circumstellar extinction, and variability, which affect the theoretical CMD positions of observed PMS stars. On the other hand, using real data allows us to account for these characteristics *intrinsically*, without modelling, and thus without possibly biasing their expected behaviour. The use of real
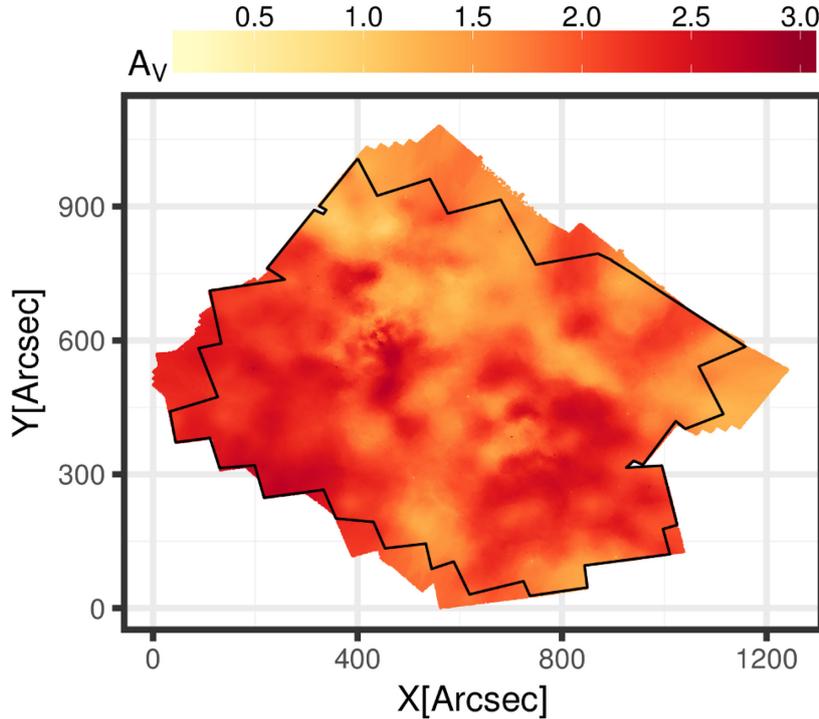
Figure 6. Extinction map of the Tarantula Nebula, where each data point is colour-coded according to its assigned extinction value, using a smoothing parameter of $\epsilon = 500$. The black outline indicates the area that is covered both in *V* and *R*, i.e. the area we considered for our further analysis.

data also allows for the unbiased assessment of various observational limitations, such as crowding and photometric uncertainties, that affect the identification of PMS stars.

4.1 Selection of the training stellar sample

Due to the aforementioned physical and observational constraints (see also Section 1), the identification of PMS stars in large data sets of multiple populations is not trivial, even if we take extinction into account. It is therefore important to train our identification algorithms on the most clear stellar training set possible, i.e. on a selected subset of the HTTP catalogue where PMS stars are clearly defined in the CMD. In this study we focus on low-mass PMS stellar populations (with masses up to few $M_\odot$), which are easily confused with low-mass MS (LMS) stars. Our training subset should, thus, comprise large numbers of PMS stars, as well as of LMS stars, and other evolved populations, which are easily distinguishable. Within large star-forming complexes, regions that comprise such stellar samples are those where high concentrations of easily identifiable young stars exist, i.e. young star clusters. We select, thus, our training subsample from the most densely populated areas of the nebula, i.e. the starburst-cluster R136 and its surroundings. Specifically, we define a squared area centred on R136 ($RA_{J2000} = 05^h 38^m 42\overset{s}{.}3$, $DEC_{J2000} = -69° 06' 03\overset{''}{.}3$) with a side length of 8 arcmin (~ 120 pc). We construct the surface density map of the region by applying a kernel density estimation with a two-dimensional normal kernel.

Fig. 8 shows the surface density map of the region with the statistically significant isopleths (contour lines) in steps of σ_n, the standard deviation of the map, overlaid. In our analysis we exclude grid points that fall outside the HTTP coverage (top left corner of the map), in order to avoid biases due to artificial zero measurements. As expected, the highest density peak corresponds to

the starburst-cluster R136, but we also identify the more evolved cluster Hodge 301 (Grebel & Chu 2000; Cignoni et al. 2016), northwest of R136, at density levels higher than $1\sigma_n$ above the average map density. A series of CMDs of stars in the region of NGC 2070 for three selected significance levels, namely $1\sigma_n$, $3\sigma_n$, and $7\sigma_n$, is shown in Fig. 9. This figure demonstrates that apart from the young UMS stars occupying the bright part of the CMDs, the high-density cluster region hosts indeed a prominent low-brightness PMS population, which is located at the red part of the CMD and well distinguishable from the LMS blue part of the CMD. Moving to higher densities within the cluster we find continuously less LMS stars, while the highest peak, corresponding to R136, contains almost exclusively PMS stars. The 4 Myr isochrone from the PARSEC evolutionary models (Bressan et al. 2012) is also shown in the figure for guidance of where young stellar populations are expected in the CMD.

A comparison of the CMD in Fig. 9 with that shown in Fig. 7 indicates that while it is relatively straightforward to identify the low-mass PMS stars inside the giant HII region NGC 2070, the identification of the same type of stars across the whole HTTP CMD requires a statistical modelling of their positions. Our classification focuses on the performance of this modelling across the whole extent of the HTTP survey. In order to include a fair number of non-PMS examples, i.e. LMS stars, in our training set we select the HTTP subset included within the isopleth of $1\sigma_n$ above background centred on R136, as shown in Fig. 9. The two distinct populations in the low-brightness CMD regime, i.e. LMS and PMS stars, are clearly demonstrated in the corresponding Hess diagram, shown in Fig. 10. For building the training data set we restrict the stars to be considered to the low-brightness regime, excluding most of the UMS stars, and we remove some of the very blue and red objects,
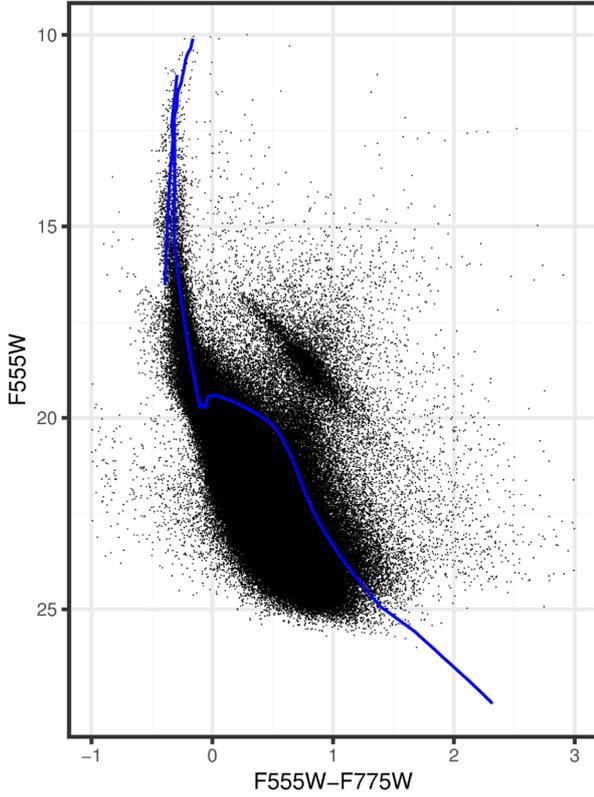
Figure 7. The extinction-corrected CMD of the HTTP data with additional artificial noise imposed on the corrected positions of the UMS stars in order to generate a more organic CMD. The blue line corresponds to the 4 Myr PARSEC isochrone, whose upper part is used as an approximation for the ZAMS to measure the extinction of the UMS stars.
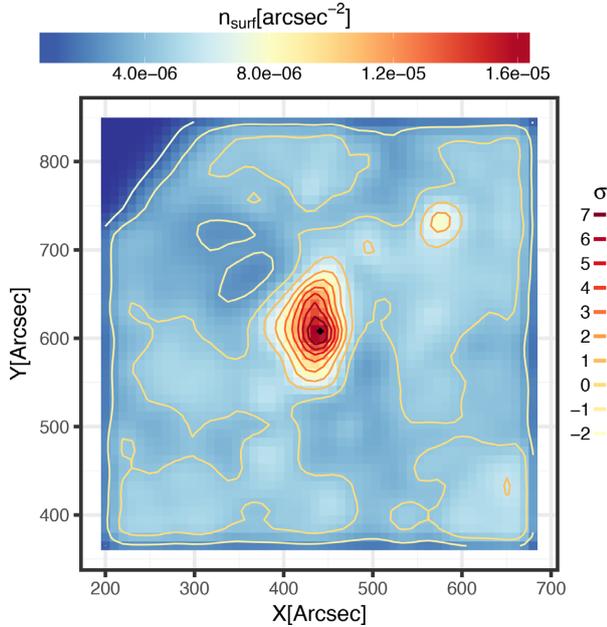
Figure 8. Surface density plot of NGC 2070 with overlaid density contours covering the range from $-2\sigma_n$ to $7\sigma_n$ in steps of σ_n. The black square marks the centre of R136.

as we suspect them to have poor photometry. The limits of the CMD region covered by this stellar sample are shown in Fig. 11. Considering that R136 is a young cluster, the LMS stars in the region are most probably field contaminants, not belonging to the cluster. Under these circumstances our selected subset is optimal in including good training examples of low-mass PMS and non-PMS stars. In the following section, in order to characterize each star as a positive or negative PMS example, we distinguish these two observed populations in a quantitative way.

4.2 Distinguishing PMS from LMS stars

After selecting the low-brightness stellar sample to be used for the training of our algorithms, we characterize its members as PMS or non-PMS (i.e. LMS) stars, according to their observed extinction-corrected CMD positions. A method based on the use of stellar number distributions along cross-sections of the faint CMD was proposed by Gouliermis et al. (2012) for separating the LMS from the PMS populations on the CMD. The distributions of well-separated populations show two distinct peaks, the width and the separation of which are found to depend on stellar brightness. We implement this method with one modification: We further introduce a reference line in the CMD, which has a slope roughly equal to the gap between the LMS and PMS, as observed in the Hess diagram of Fig. 10, and we calculate the distance of each star in the selected sample from this reference line.

With this modification we analyse the stellar distance distributions from the reference line, while avoiding to bin the stars, in contrast to Gouliermis et al. (2012), who analysed the binned stellar number distributions. We use a bimodal Gaussian mixture model to fit the calculated distances of the stars from the reference line and we quantify the fit via maximum-likelihood with the application of the *expectation maximization* (EM) *algorithm* (see Appendix B), an iterative method to derive maximum a posteriori estimates of parameters in statistical models, where the model depends on latent variables[1] (Dempster, Laird & Rubin 1977). We choose this method over a simple non-linear least-squares regression, because of its high reliability in converging to a successful fit even in cases where the regression could not.

The bimodal Gaussian distribution used to model the distances of stars from the LMS–PMS separating line has the form

$$G(x) = \sum_{m=1}^{2} \alpha_m \Phi(x; \mu_m, \sigma_m), \qquad (4)$$

where α_m denotes the mixing proportions, with the sum of all proportions (in this case two) equal to unity, and μ_m and σ_m are the mean and the standard deviation, respectively, of each of the individual components. Using the model fit by the EM algorithm, one can estimate the posterior probability p_{im} that a star i belongs to one of the components m of the Gaussian mixture model as

$$p_{im} = \frac{\alpha_m \Phi(x_i; \mu_m, \sigma_m)}{\sum_{k=1}^{2} \alpha_k \Phi(x_i, \mu_k, \sigma_k)}. \qquad (5)$$

With this measure we can distinguish the PMS from the LMS stars in our selected sample, on an individual-star basis by assigning a probability of PMS membership to each star. We can thus set a probability threshold above which all stars are considered as the best PMS examples. In Fig. 12 we show two examples of our test

---

[1] Unobserved, hidden variables, usually inferred from observables, like e.g. categories.
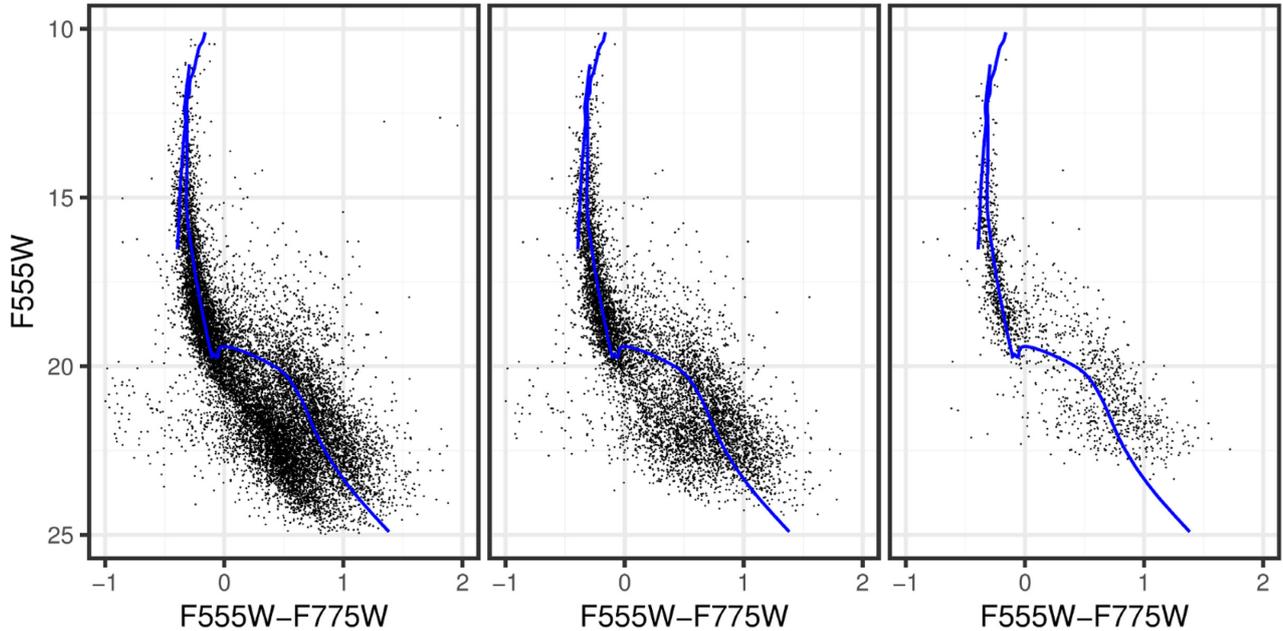
Figure 9. Optical extinction-corrected CMDs for the regions within NGC 2070 with surface density higher than $1\sigma_n$ (*left*), $3\sigma_n$ (*middle*), and $7\sigma_n$ (*right*) above the local background density (see Fig. 8 for the corresponding areas). The blue line marks the 4 Myr PARSEC isochrone.
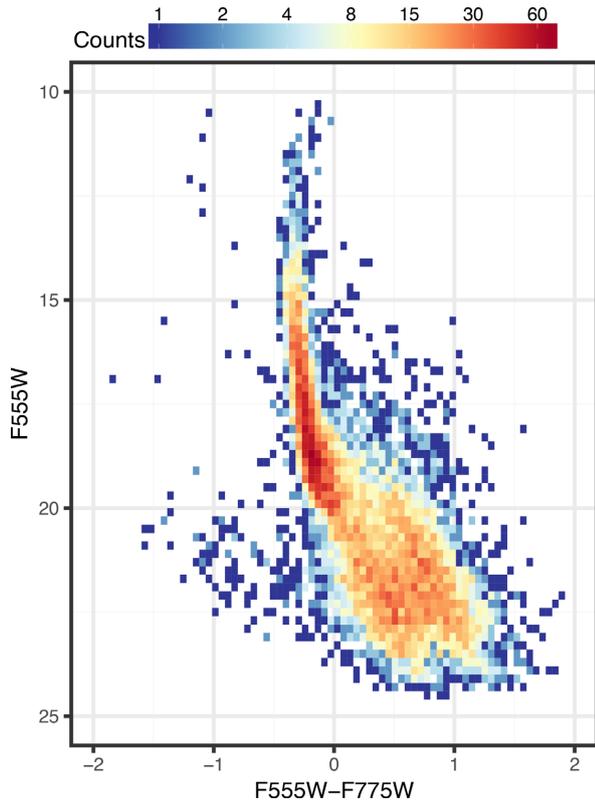
Figure 10. Optical extinction-corrected Hess diagram of the stars in the region of NGC 2070 included within the $1\sigma_n$ density significance level (see map in Fig. 8). The diagram uses 75 bins in both coordinate directions with limits $[-2, 2]$ in *F555W−F775W* and $[25,10]$ in *F555W*.

runs of the EM method for different reference lines. The histograms on the right-hand panel are only shown to provide a visualization of the result, since the fitting process itself requires no binning of the distance measure. The posterior probability p_{im} of each star being a PMS star is calculated from the Gaussian mixture model component with the larger mean μ_m, i.e. the PMS component of the model.

The examples of Fig. 12 show that the EM method is quite successful in distinguishing the two separate populations within the low-brightness regime of the CMD for a given reference line. Our tests also demonstrated that the outcome of the EM method is independent of the axis intercept of the reference line. However, as shown in the plots of Fig. 12 the result of the EM method does depend on the choice of the slope. It is thus important to accurately define the reference gap between the LMS and PMS populations in the observed CMD, in order to avoid any potential biases in the application of the EM method. Since representing this gap with a single straight line would provide an unrealistic boundary between LMS and PMS stars, for the application of the EM algorithm we do not consider a single LMS–PMS reference line (as in the examples of Fig. 12). Instead, we define a threshold curve using a series of PARSEC isochrones, ranging from 0.5 to 10 Myr, which approximates realistically the observed LMS–PMS gap. While there may be somewhat older PMS stars, we select 10 Myr as the oldest considered age based on the fact that this limit corresponds to the majority of the star formation history of the region, as specified by previous studies (Hunter et al. 1995; Cignoni et al. 2015). The corresponding isochrone model also nicely traces the observed gap in the low-brightness regime (see orange line in Fig. 13).

We consider the faint part of the 10 Myr isochrone up to one stage before its MS turn-on as the best representative line of the LMS–PMS gap, and we extend this line to brighter magnitudes by connecting the points corresponding to the same stage, i.e. to the red of the local minimum before the turn-on, for all the other isochrones. With this process we construct a threshold curve that
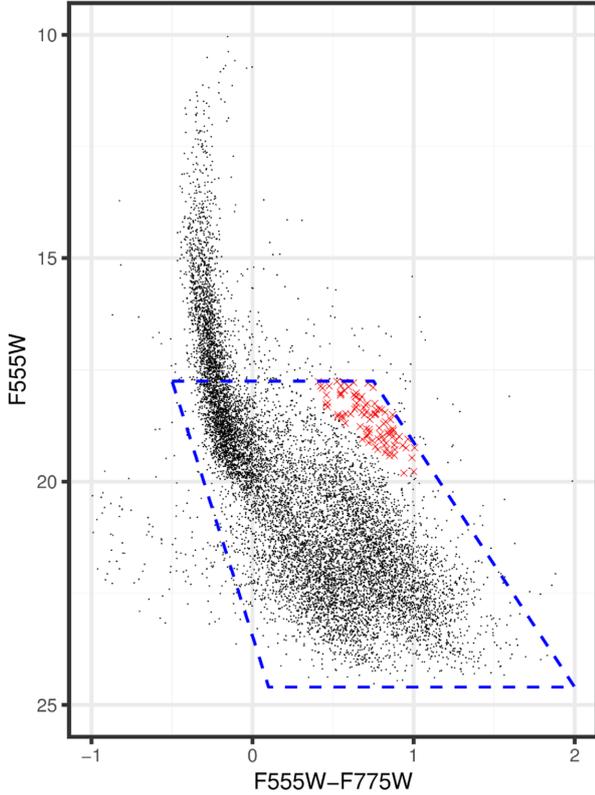
Figure 11. Optical extinction-corrected CMD of the data set used to create the training set. The blue dotted line indicates the data that are used for the EM fitting. The red crosses mark the data points that are excluded from the fitting, because they fall into the RC area.

Figure 12. *Left:* The optical CMD of our selected training stellar sample with the reference line (separating the observed LMS and PMS parts of the CMD), drawn with a dark red line. The blue box indicates the CMD region occupied by the selected sample. The points within the blue box are colour-coded according to their posterior probability of belonging to the right component of the mixture model displayed in the right-hand panel. *Right:* Histogram of the stellar (perpendicular) distances from the reference line with the bimodal Gaussian components of the fit solution overlaid. The panels show the resulting distributions of distances from the reference line for lines with the same intercepts but different slopes. While the outcome of the EM method is found to be independent of the actual position of the reference line, as demonstrated in these plots, it is very sensitive to its slope. It is thus important to identify a reference line the slope of which fits at the best possible degree that of the observed gap between LMS and PMS stars (see Section 4.2).

does not overlap with the UMS, while adequately tracing the gap between the LMS and PMS populations, which we want to quantify. This is demonstrated in Fig. 13, where each isochrone model is plotted with a different colour, and the thick orange line indicates the constructed threshold curve between PMS and LMS.

For the application of the EM algorithm, instead of using a single LMS–PMS reference line, we use the constructed LMS–PMS threshold curve to generate *a series of reference lines* by fitting a line to sequential sets of four points of the curve. We allow for some overlap between the point sets, with three of the brightest points in each set coinciding with the three faintest in the next. With this process we produce 46 different lines with different parameter sets of slopes and intercepts. The determination of the PMS membership probability for the stars in the selected sample is then made with the application of the following steps: (1) We calculate the distances of the stars from each of the 46 reference lines, (2) we fit the corresponding bimodal Gaussian distribution, (3) we estimate from each model the PMS component membership posterior probability for each star, and (4) we average the results from all Gaussian mixture models for each star. Our selection excludes all stars brighter than $m_{F555} = 17.75$ mag and the noisy observations to the right of the PMS and left of the LMS population, outside the CMD region limits shown in Fig. 11. In our treatment we also do not consider the stars marked with red crosses in this figure as these fall into the RC part of the CMD, and they are likely evolved stars. With the methodology described above we avoid contamination by UMS or other more evolved stars and objects with poor photometry during

the fitting process. The fit of the Gaussian distribution is repeated 100 times for each parameter set and the resulting PMS component membership probabilities are averaged to reduce the influence of the random initial model parameter guess in the EM algorithm.

The outcome of the application of the EM method is visualized in the CMD of Fig. 14, where each considered star is coloured according to its estimated PMS membership probability (p_{em}). The threshold curve used for the application of the EM algorithm and the curve corresponding to the limit of $p_{em} = 0.7$ are also shown in the figure (with orange continuous and red dashed lines, respectively). The comparison of these lines indicates that a minimum of PMS membership probability of about 70 per cent provides a reasonable separation between PMS and LMS stars in the training sample. We, thus, tested the construction of our training data set by using various probability thresholds, starting at 0.7 up to 0.9. We assigned as PMS stars those with p_{em} larger than the chosen threshold, while the remaining stars were characterized as non-PMS. Our subsequent
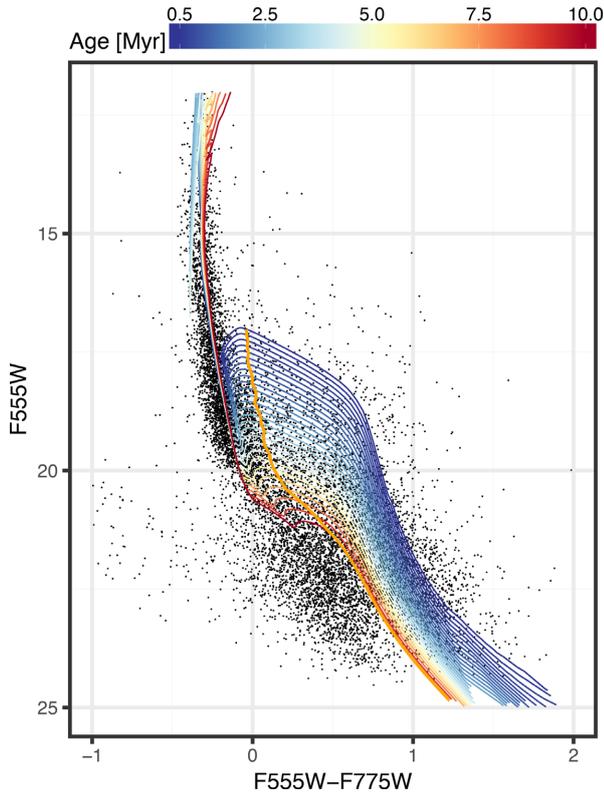
Figure 13. *Left:* Optical extinction-corrected CMD of our selected training stellar sample with an overlay of the PARSEC isochrones with ages between 0.5 and 10 Myr (coloured according to their age) used to approximate the LMS–PMS gap. The thick orange line marks our extrapolated threshold curve between the PMS and LMS.

investigation of the performance of the classification algorithms in dependence of the considered p_{em} thresholds showed that the best PMS candidates in the training set are all stars with $p_{em} \geq 0.85$ (see Section 5).

4.3 The final training data set

With the implementation of the EM method as described above we established a reasonable data set of low-brightness true PMS and LMS stars to be used for training the classification algorithms. However, while the CMD area covered by these stars in the region of NGC 2070 shows a clear distinction between these two stellar types, the bright part of this area in the complete HTTP CMD includes other types of evolved stars, such as RC and faint giant/subgiant stars. While the contamination of the PMS data set by these stellar types is not significant, they must be considered in our final training data set. We, thus, complete the compilation of the training set by 'artificially' adding examples of these evolved stars as *negative* (non-PMS) examples, so that the classification algorithms can treat them as such. For the RC stars we include the previously excluded examples, marked with red crosses in Fig. 11, and we assign a PMS membership probability of $p_{em} = 0$ to them. As we discuss in Section 4.1 we constrained our EM analysis for distinguishing PMS from LMS stars in a well-defined region in the CMD, where prominent members of both populations are located (blue dashed polygon in Fig. 11), excluding some sources at the extreme blue and red faint parts of the diagram. We now include these faint
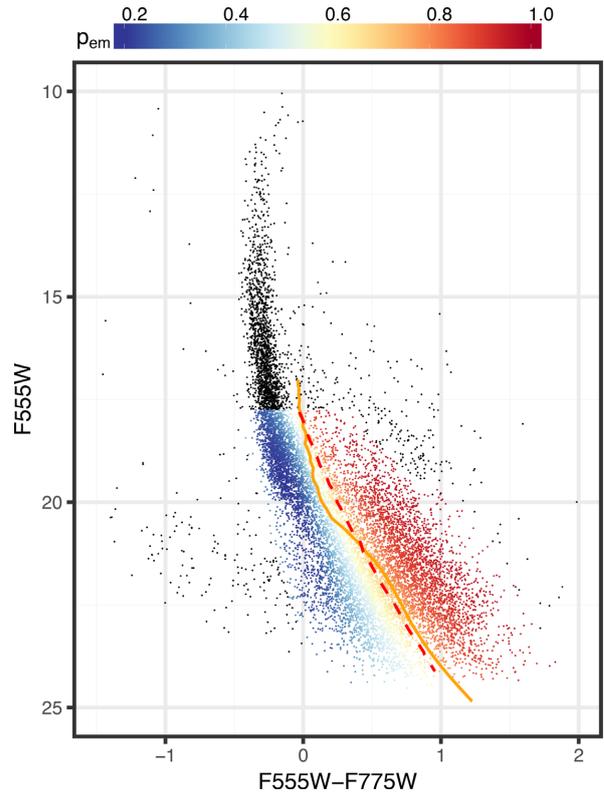
Figure 14. Optical extinction-corrected CMD of the stars in the region of NGC 2070 included within the $1\,\sigma_n$ density significance level (see map in Fig. 8). Stars that were considered during the application of the EM method are coloured according to their estimated PMS membership probability p_{em}. The orange line marks the previously established approximation curve of the LMS–PMS gap (see Fig. 13) for reference. The red dotted curve indicates a 70 per cent probability threshold of p_{em}.

uncertain sources in the final training set by assigning to them also zero PMS probability. The reason for this inclusion is to eliminate the danger of misclassifying objects with uncertain measurements as PMS stars.

4.3.1 Adding evolved field stars in the training set

Apart from the stellar sources discussed in the previous paragraph, an important contaminant of a PMS data set is the old stellar populations of the general LMC field, occupying the giant/subgiant branches of the CMD. The fainter giant and subgiant stars of the LMC field can roughly coincide, depending on age and reddening, with the bright part of the PMS population. In the case of variable extinction by gas and dust, giant and subgiant field stars are distributed along the reddening vector and can overlap with the CMD positions of PMS members of young clusters on the same line of sight. In order to use a training data set that accounts also for these contaminants we identify typical examples of faint, field giant and subgiant stars in regions of the observed HTTP FoV, which mostly cover the general LMC field. We select two such regions to account for both high- and low extinction of the field stars by the nebula.

In selecting these regions we were aided by a preliminary unrefined classification of PMS stars in the HTTP data set by employing a *support vector machine* (SVM) algorithm trained on the *V*- and
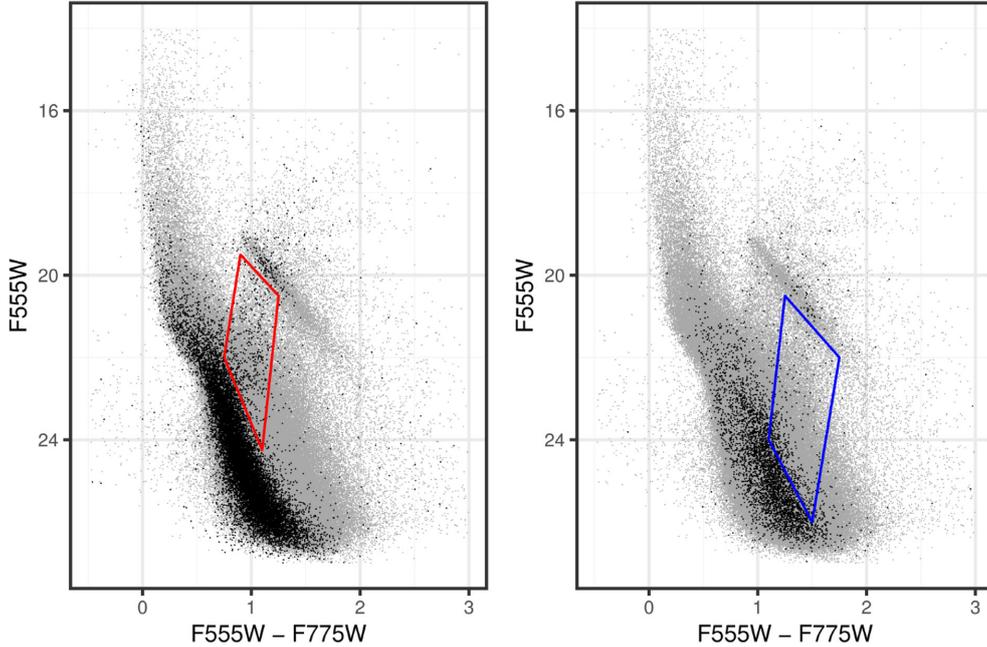
Figure 15. Optical CMD of the HTTP data in grey with an overlay of the CMD of our field star selections, where the low-extinction region is shown on the *left* and the high-extinction region on the *right*. The solid red and blue polygons indicate our selections for the low and high extinction, faint giant/subgiant candidates, respectively.

R-equivalent magnitudes of the stars in the preliminary training set constructed in the previous section. Details on the employment of this method are given in the final application of our classification (Section 5) and in Appendix A3. Here, it suffices to note that we applied the SVM method via a 10-fold cross-validation, repeated five times and labelling as PMS stars (label of 1) those with PMS membership probability, derived with the EM method, higher than the lowest reasonable limit of $p_{em} = 0.7$. We performed the classification of the HTTP stars with measurements in these two filters and we retrieved a *tentative set of PMS candidates*, i.e. stars with a classification probability ≥ 0.5, across the whole observed FoV. We constructed the surface density map of this stellar sample in order to identify the regions across the Tarantula Nebula that are mostly devoid of candidate PMS populations, i.e. the regions where the cleanest samples of field stars can be detected.

We combined the surface density map of the PMS candidates with the extinction map of Fig. 6 to identify the *field regions* within the observed area with both the lowest and highest extinction. We, thus, considered the field contaminants in the whole range of reddening conditions across the Tarantula Nebula. It is interesting to note that our selected low-extinction field region roughly coincides with that defined by Cignoni et al. (2015) as reference field in recovering the star formation history of NGC 2070. The identified field stellar populations are depicted in the CMD of Fig. 15. We select from this population the most prominent, faint giant/subgiant stellar candidates as enclosed by the red and blue polygons (one from the high reddening and one from the low-reddening region), and add a bit more than 900 objects to our training set as non-PMS examples, i.e. stars with zero PMS membership probability.

With the process described in this and previous sections, we have constructed a training data set of 10443 stars, containing the best possible examples of (1) evolved field stars, both LMS and potential giant/subgiant stars, (2) RC stars, (3) non-specified stars with poor photometry, and (4) young low-mass PMS stars, which we aim at identifying across the whole FoV. Fig. 16 shows the part of the CMD on which our training will take place, with an overview of the positions of these populations, coloured according to their p_{em} probability. As shown in this CMD, our classification is limited to the faint part of the CMD where the PMS stars reside, and therefore we do not include examples of the UMS stars, as their positions should not overlap with those of the PMS stars.

The available data in the training data set in each of the *HST* filters is summarized in the top part of Table 2. Apart from *V*- and *R*-equivalent wavebands, our training set includes substantial data in *J*- and *H*-equivalent wavebands (although with a smaller FoV), making the training of our algorithms on these variables also feasible. The remaining three (*UV*-, *U*-, and H_α-equivalent) filters are less helpful for the application of a machine-learning classification due to their significantly smaller coverage, drastically reducing the amount of data to predict on (see Table 1). As a consequence, we limit our tests to these four filters. Also, before feeding the learning process with the training examples, we need to define this limiting PMS membership probability threshold, p_{em}, to be considered for separating the positive (PMS) instances from the negative (non-PMS) instances in the training data set. The primary criterion for determining this threshold is the inclusion of the purest possible sample of PMS candidates, reducing, thus, the number of possible false positive examples. As we discuss later, we determined the best threshold choice to be 0.85 based on two additional factors. (1) The need for balance between the numbers of positive and negative examples in the training data set and (2) the classification performance of the algorithms.

Concerning the balance between positive–negative examples, the lowest reasonable threshold of $p_{em} = 0.7$ provides about 38 per cent of the training set as positive instances. At the even higher threshold
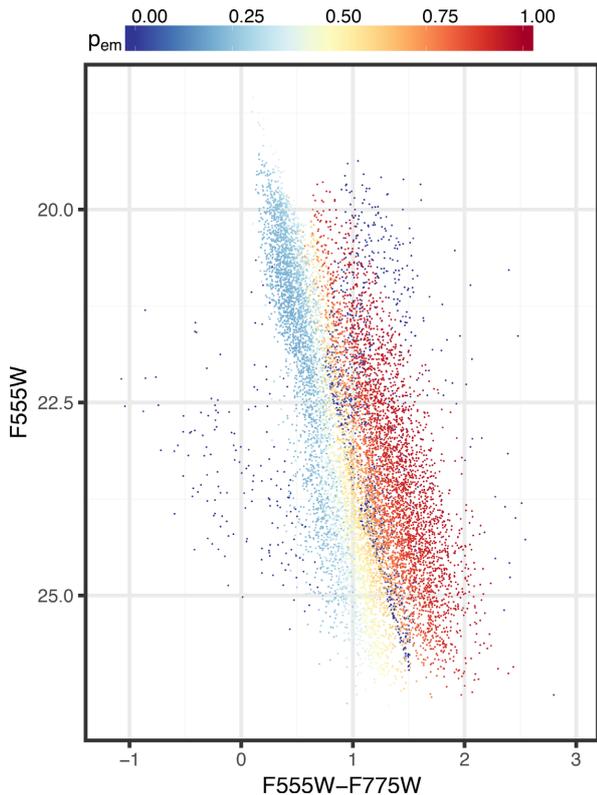
Figure 16. Optical CMD of the final training data set selection. The data points are colour-coded according to the PMS membership probability defined as described in the text (see Section 4).

of $p_{em} = 0.9$ we retain roughly 19 per cent of the data set in positive examples, which still provides a useful amount of observations. Using an even higher probability threshold would not be practical, because we would limit the training sample to a number of positive instances that would be unrealistically low, and that would introduce a strong imbalance between positive and negative examples, which is not ideal for machine learning. Concerning the performance of the classification algorithms, our experiments (Section 5) showed that we achieve the best trade-off between algorithm performance and training set balance by constraining the sample of PMS members in the training set to those with $p_{em} \geq 0.85$, corresponding to roughly 27 per cent of the total training data set.

5 CLASSIFICATION OF PRE-MAIN-SEQUENCE STARS

After constructing the training data set to be used for the learning process, we performed various experiments in order to identify the most efficient machine-learning algorithm for the classification of PMS stars in the Tarantula Nebula. Since our training set was constructed from the region of NGC 2070, our classification will identify the stellar *siblings* of the PMS members of this region, i.e. stars with similar characteristics and star formation history, spread across the whole nebula. In our experiments we tested three popular classification algorithms: Decision trees, random forests (RF), and SVM. Descriptions of the concepts behind these algorithms and references to the related literature are provided in Appendix A. During our early experimentation we also considered the application

of the Naive Bayes classifier, a simple probabilistic classifier based on Bayes' theorem, which assumes (naively) a strong independence between the features (see e.g. Russell, Norvig & Davis 2016).

The success of supervised machine-learning modelling is based on the availability of complete sets of observations with as many variables to model on as possible. As a consequence, in this study a limiting factor in our classification is the amount of available data, the algorithms can be trained on. This translates to both the number of available stars per filter and the number of stars observed in as many filters as possible, i.e. the size of the feature space (see Appendix A). The second aspect is particularly important, because most of the classification algorithms cannot perform a prediction on incomplete feature vectors, i.e. on missing data. With this in mind we optimized our classification for data sets with the best waveband and spatial coverage, i.e. for stars found in the *HST* wavebands equivalent to standard *V*, *R*, *J*, and *H* filters. Among the tested methods only *decision trees* can compensate for incomplete feature vectors using the so-called *surrogate splits* (see Appendix A1). It is, thus, the only algorithm that can predict on all available HTTP filters. For our tests with other algorithms, as mentioned above, we do not take into account measurements in the *UV*-, *U*-, and H_α-equivalent filters, due to the significantly large amount of non-detections in these wavebands (as shown in Table 1). Our experiments with the use of photometric flags as categorical classification variables (to compensate for non-detections) also performed very poorly.

The performance of the algorithms was measured using three metrics, the *accuracy*, the *balanced accuracy*, both estimated from the *confusion matrix*, and the area under the *receiver operating characteristic* (ROC) curve, or in short *area under the curve* (AUC), all described in Appendix A4. These metrics were calculated on the basis of a *train/test split*, i.e. by splitting the original training data set in two subsets, one to train the algorithm ('Train' subset), and the other to test its performance ('Test' subset). This method is very efficient when there are a sufficiently large number of records in the training data set, as in our case. Typically, a 70/30 split, i.e. ∼ 70 per cent of the training data set reserved for the Train subset and the rest for the Test, is the most efficient split for training the algorithm in order to avoid *overfitting*, i.e. constructing a general model that can fit a variety of data, and not exclusively those in the training data set. It is worth noting that the measurements of the predictive power of the classification model must be made on a *held-out* Test set, i.e. the records of the Test set must not be influenced in any way by the instances in the Train set. Therefore, splitting the Train/Test sets is an important aspect of the process.

For the training of the algorithms on the Train subset, we employed a 10-fold cross-validation (see Appendix A4.1) in all our experiments. Due to differences in the available data that depend on the availability of measurements in various wavebands, the Train/Test subsets may vary from one experiment to the other. For the sake of direct comparability of the presented results the SVM and RF algorithms are trained and tested on exactly the same subsets for any given experiment. This is of course not a necessity, as the algorithms' performance varies insignificantly for different partitions of the data.

The implementation of the algorithms was in the programming language *R*, an environment for statistical computing and graphics (R Core Team 2013). From early on the *Naive Bayes classifier* shows significantly low performance in our experiments, achieving accuracies of at most ∼60 per cent, providing thus classification models comparable to random guessing. As a consequence, we do not further discuss in our analysis this algorithm, and we focus on

Table 2. Overview of our machine-learning experiments for the identification of PMS stars in the Tarantula Nebula. The table lists (1) the investigated combinations of variables (features) for training and prediction, (2) the available training data for each individual feature, (3) the applied algorithms, (4) the available data in both the training set and the HTTP survey for each features set, (5) the performance of each algorithm on a held-out test set (the value in parenthesis gives the performance on the training set during cross-validation for comparison), and (6) the amount of resulting PMS candidates, i.e. stars identified with a predicted probability of being PMS of $p_{pms} \geq 0.5$. Note that the number of available records is given only for the RF and SVM algorithms, because the decision tree algorithm can predict and train on data sets with incomplete attributes, i.e. on all records in the HTTP survey. The performance on the test set is found to be comparable to that on the training set, which demonstrates that none of our models exhibit a case of overfitting.

Feature	Available training data (out of 10 443 stars)	1	2	3	4	5
F275W (UV)	2210 (21.2 %)	✓				
F336W (U)	4880 (46.7 %)	✓				
F555W (V)	10 443 (100 %)	✓	✓	✓	✓	✓
σ_V	10 443 (100 %)				✓	
F658N (H$_\alpha$)	4576 (43.8 %)	✓				
F775W (R)	10 443 (100 %)	✓	✓	✓	✓	✓
σ_R	10 443 (100 %)				✓	
F110W (J)	9600 (91.9 %)	✓				✓
F160W (H)	9597 (91.9 %)	✓				✓
Extinction A_v	10 443 (100 %)			✓		
Applied algorithms:						
Decision tree (DTRee)		✓	✓			
Random forest (RF)			✓	✓	✓	✓
Support vector machine (SVM)			✓	✓	✓	✓
Available data:						
Training total		10 443	10 443	10 443	10 443	9283
Reduced training		7310	7310	7310	7310	6498
Held-out test		3133	3133	3133	3133	2785
Prediction		822 204	403 018	403 018	400 229	287 434
Performance on test (training) set ($p_{em} \geq 0.85$):						
Accuracy (in %):						
DTree		86.11 (89.23)	87.80 (89.67)	–	–	–
RF		–	94.16 (94.36)	95.60 (95.85)	94.54 (94.77)	92.78 (93.00)
SVM		–	95.18 (95.36)	97.29 (97.47)	94.80 (94.26)	94.82 (94.64)
Balanced accuracy (in %):						
DTree		82.60 (87.15)	85.83 (87.92)	–	–	–
RF		–	92.06 (93.09)	94.24 (94.67)	92.71 (93.15)	90.88 (91.16)
SVM		–	93.39 (94.53)	96.61 (95.89)	93.06 (92.63)	93.05 (93.90)
ROC AUC:						
DTree		0.845 (0.899)	0.852 (0.885)	–	–	–
RF		–	0.983 (0.984)	0.990 (0.990)	0.986 (0.986)	0.977 (0.978)
SVM		–	0.988 (0.989)	0.994 (0.994)	0.985 (0.983)	0.989 (0.986)
Number of PMS candidates						
DTree		74 375	73 006	–	–	–
RF		–	21 306	20 996	20 923	15 898
SVM		–	21 550	19 487	21 554	16 655

the three remaining methods that proved to provide more accurate results. RF, operating by constructing a multitude of decision trees, correct for decision trees' occasional overfitting to the Train data set (Hastie, Tibshirani & Friedman 2009). The RF algorithm is thus a more efficient choice for our classification. Nevertheless, decision trees can handle missing values in the photometric variables of stars without imputation (by using surrogate splits). We apply, thus, a preliminary classification with decision trees only on the complete set of photometric variables (set no. 1), which includes a large number of missing values, in order to understand how these variables may influence our classification. Table 2 gives a detailed summary of our experiments, listing the sets of different variables (features) combinations used for training and classifying the HTTP stars, the algorithms applied, the available instances in both the training and

the whole HTTP data sets in dependence of the variables choice, the performance of the algorithms on the held-out Test data set in terms of the accuracy, balanced accuracy and the AUC, and the amount of the identified PMS candidates.[2]

As we discuss later in this section, for the final census of the PMS stars across the Tarantula Nebula we operate on the features set no. 3 (as described in Table 2), because it provides the largest stellar coverage across the whole extent of the observed field, thus the richest stellar sample, and the highest performance scores across the tested methods. This sample comprises observations in three

---

[2]In the following sections we only state the accuracy and AUC for simplicity. We refer the reader to Table 2 for the corresponding balanced accuracy.

variables (*V*-, *R*-equivalent magnitudes and extinction), providing the largest amount of HTTP data to classify.

Features set no. 1 includes all photometric measurements, while set no. 2 is equivalent to set no. 3, but without the inclusion of extinction. It is used in order to validate the influence of extinction to the accuracy of our models. Including the infrared measurements unfortunately limits not only the area that we can investigate (Fig. 1), but also drastically reduces the amount of data that can be classified, due to the lower number of stars observed in all four wavebands.[3] We nevertheless investigate in features set no. 5 their influence on the performance of the algorithms. (Note that extinction is not included in this features set, as it is intended to highlight the influence of the inclusion of the infrared bands alone.) Features set no. 4 is intended to give insights on whether the photometric errors in the respective filters could prove to be helpful in the classification approach, or not. Therefore, extinction was not included in this features set.

5.1 Classification with decision trees

Decision or classification trees use flowchart-like structures that break the process of a complex decision into a series of simpler decisions, made upon the input features of the available data (see Appendix A1). Beginning at the *root node* data flows through if-else *decision nodes* that split the data according to its features. The *branches* indicate the potential choices and the *leaf nodes* the final decisions. Given that the decision tree algorithm can compensate for incomplete feature vectors, our first experiments were made with the application of this algorithm in order to use the complete photometric variables space of the HTTP data set, which includes a large number of missing values for stars not identified in specific wavebands. We trained decision trees using various maximum tree depths up to 30, at most 5 surrogate splits per node and pre-pruning with complexity degrees of the order of 0.01, using the *Gini Index* as measure of node impurity (see Appendix A1 for explanations on these parameters). In order to compare with the other algorithms the results presented for this method are also based on PMS stars with $p_{em} \geq 0.85$. The final decision tree, with an accuracy of 86.11 per cent and an AUC of 0.845 on the held-out test set, did not achieve the expected performance, but it appeared quite promising in providing valuable insight on the importance of certain filters.

In this tree the two most important variables for primary splits appear to be the measurements in the *F555W* and *F160W* filters with minor contributions from those in *F775W*. Measurements in all remaining filters are only considered for surrogate splits. Comparing with Table 2 this is not surprising, since our training data set contains significantly less records in the *F275W*, *F336W*, and *F658N* filters, and therefore these variables are not considered for primary splits. The fact that the filters pair (*F555W*, *F160W*) was chosen over the (*F775W*, *F110W*) pair may indicate that measures in the *V*- and *H*-equivalent bands provide the intrinsically best combination of variables. This can be explained by the fact that the (*F555W*, *F160W*) filters pair provides a rich stellar sample across a dynamic range in colours, which is wider than those of any other combination of these four filters. The wide colour spread of the data allows a clearer distinction between PMS and LMS stars on the CMD. In any case these results provide strong indications that near-infrared measurements may be very important to

the identification of PMS stars with machine-learning classification techniques.

Considering that our primary data set is that with the best spatial coverage across the observed FoV, provided by stars observed in at least one of the *F555W* and *F775W* filters, we trained a decision tree also on this stellar sample. This tree performed decently on the held-out test set, with an accuracy of 87.8 per cent and AUC of 0.852, comparable to the tree model based on the whole stellar sample, and still below our expectations for successful classification. Moreover, we identified few issues related to the decision tree classification. Specifically, there was a large number of misclassifications of LMS and RC stars after applying the algorithm on the entirety of the HTTP data set, and the outcome of the classification itself appeared quite unrealistic, with the identified PMS candidates being aligned in prominent 'zigzag' patterns across the CMD. We explain this phenomenon as the result of the binary splits the decision tree algorithm performs in order to make a decision.

Another issue with the decision trees that were trained on the whole sample is that since the constructed trees use the observations in the *F555W* filter as one of the primary prediction parameters, they get heavily confused in the regions, where there is no coverage in this filter (see e.g. Fig. 1), by predicting an unrealistically large amount of PMS stars in these regions. This shows that the surrogate split method to compensate for incomplete feature vectors provides limited support to our classification goal. Based on these experiments, and due to the issues mentioned above, we assess that while the decision tree algorithm provides evidence for the importance of infrared measurements, in general it is not suited for the purpose of this study. As a consequence we did not proceed with any further tests of the decision tree, beyond these preliminary experiments. Our further tests were focused on the more sophisticated RF and SVM algorithms.

5.2 Classification with random forests

A number of classification trees can be combined into a collection known as *decision tree forest*, or simply RF, which is one of the most successful machine-learning classifiers. In contrast to a single decision tree that is grown in size and complexity as it is trained on the available data, the efficiency of the RF relies on the fact that the algorithm is a collection of smaller simpler trees that together reflect the data's complexity (see Appendix A2 for a detailed description). We applied our machine-learning method on the reduced training set in three steps: (1) We employed cross-validation to train 10 RF classification models, (2) we chose the best model based on its AUC during cross-validation, and (3) we evaluated the performance of the final model independently on the held-out test set. Two basic arguments in the implementation of the RF algorithm is the number of trees the 'forest' consists of, and the number of variables to be sampled in each node. Due to the unavoidably low number of available variables in each features set (Table 2), all of them were used in the training process for each considered set. The number of trees per forest, which should not be less than 200, was tested for values between 500 and 10 000. However, it appears that the algorithm's performance is not very sensitive to the number of trees, since all models provided AUC values with differences of the order of 0.0001. Nevertheless, the best trade-off between performance and computational demand was achieved with the models for 500 trees, with that for features set no. 2 having an AUC of 0.9660.

---

[3]We remind that while decisions trees can deal with missing measurements, the RF and SVM algorithms cannot.
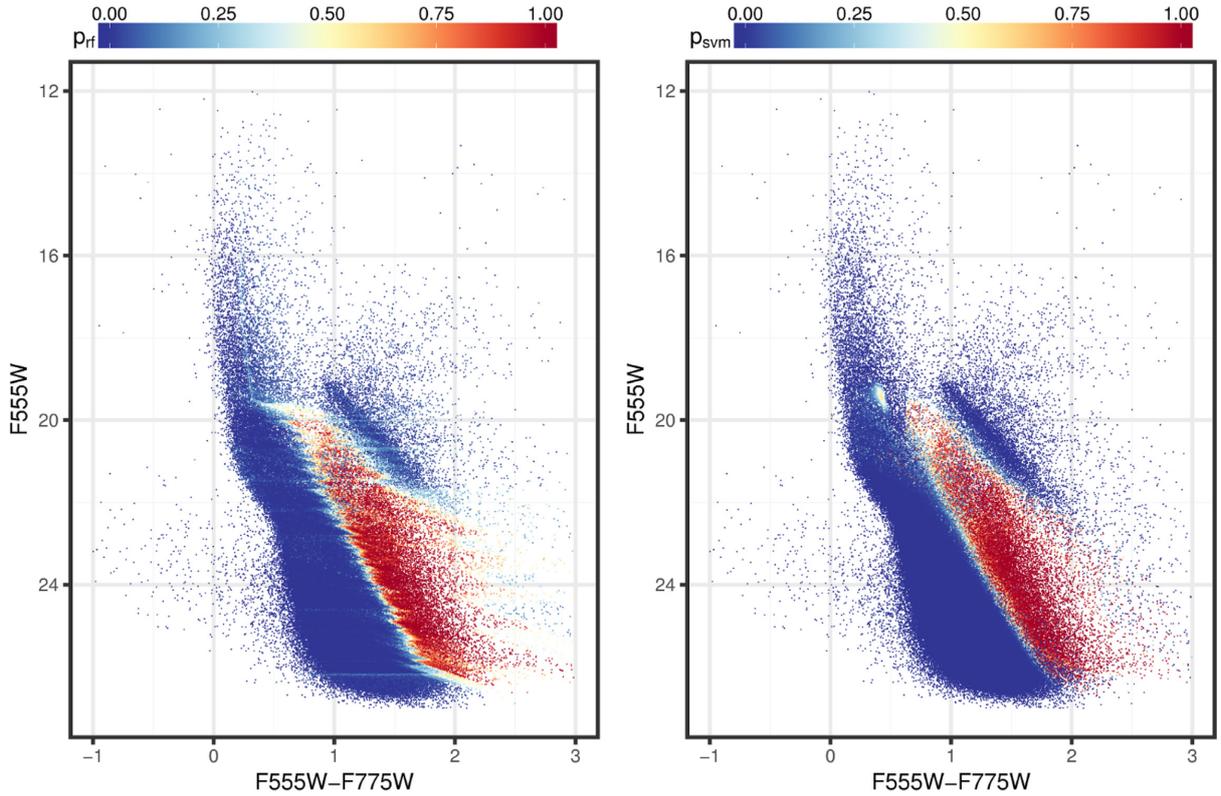
Figure 17. Optical CMDs of the HTTP stars, colour-coded according to their probability of being PMS stars, as predicted by the RF algorithm (*left*), and the SVM algorithm (*right*), trained on their measurements in the *F555W* and *F775W* bands and their extinction A_v.

We evaluated the influence of our selection for the EM-derived PMS probability threshold (p_{em}), introduced during the construction of our training data set, on the classification performance of the RF algorithm by testing its performance in a range of limiting values, varying from $p_{em} \geq 0.7$ to $p_{em} \geq 0.9$ for the prominent PMS candidates in the training data set. On our primary features set (set no. 3, i.e. *F555W*, *F775W* measurements and A_V), we found that apart from an overall excellent performance (AUC ≥ 0.984), there is a trend of increasing performance with higher limiting threshold. The best model was constructed for the highest considered threshold ($p_{em} \geq 0.9$), achieving an AUC of 0.991, with a difference of only 0.007 larger than that for $p_{em} \geq 0.7$. This indicates that the performance of the RF algorithm appears to be also not sensitive to the considered input sample of best PMS candidates. It should be noted, though, that choosing a threshold of $p_{em} = 0.9$ might already be critical in terms of maintaining a good balance between positive and negative examples in the training set, since this threshold accounts for only ~ 19 per cent of positive examples in the data set. Based on this, and in order to achieve a trade-off between correct training of the algorithms and reasonable selection of the best positive examples in the training set, we applied a threshold of $p_{em} \geq 0.85$ for the best PMS input sample.

With these settings the RF achieves an excellent accuracy and AUC of 95.6 per cent and 0.990 respectively for the primary features set (set no. 3), providing the best performance of the RF across all the feature combinations we have tested. In our implementation of RF, the predictions of the constructed models return, apart from the class of each star (PMS, non-PMS) a probability that this star is indeed a PMS star, p_{rf}, determined from the pro-

portion of votes of the trees in the ensemble. Classifying the whole available HTTP data on the features set no. 3 (403 018 stars, cf. Table 2), the model predicts 20 996 stars, with probabilities $p_{rf} \geq 0.5$. Fig. 17 (left) shows the optical CMD of the PMS candidates, coloured according to their probabilities p_{rf}. This CMD demonstrates that the RF drastically improves the zig-zag pattern of the decision tree, albeit a smoothed such pattern can be still observed, apparently the outcome of the underlying tree nature of the method. The classification result further exhibits a mixture of PMS and non-PMS classifications in the region where old, field MS turn-off and subgiant stars could potentially overlap with turn-on stars, clearly indicating that the algorithm distinguishes these two types of stars, in contrast to the decision tree, which in our tests tended to classify all stars in the region as PMS. The RF algorithm is also more successful than the decision trees in avoiding classification of RC as PMS stars, although there might still be a few misclassifications of the faintest RC stars. Fig. 18 (left) depicts the spatial distribution of the PMS candidates across the Tarantula Nebula, coloured according to their probabilities p_{rf}. This map demonstrates, in agreement with our expectation, a large abundance of PMS stars in the regions of NGC 2070 and NGC 2060, as well as in less prominent compact stellar clusters and in features that appear almost filamentary. In this map we also mark for guidance the positions of R136 (Sabbi et al. 2016), Hodge 301 (Glatt, Grebel & Koch 2010), and NGC 2060 (Cutri et al. 2012).

The classification of the RF models trained on the other three considered features sets (sets no. 2, 4, and 5 in Table 2) appear to be overall similar to that of set no. 3, both in terms of performance, returning only 1.06 per cent to 2.82 per cent less accuracy, and in
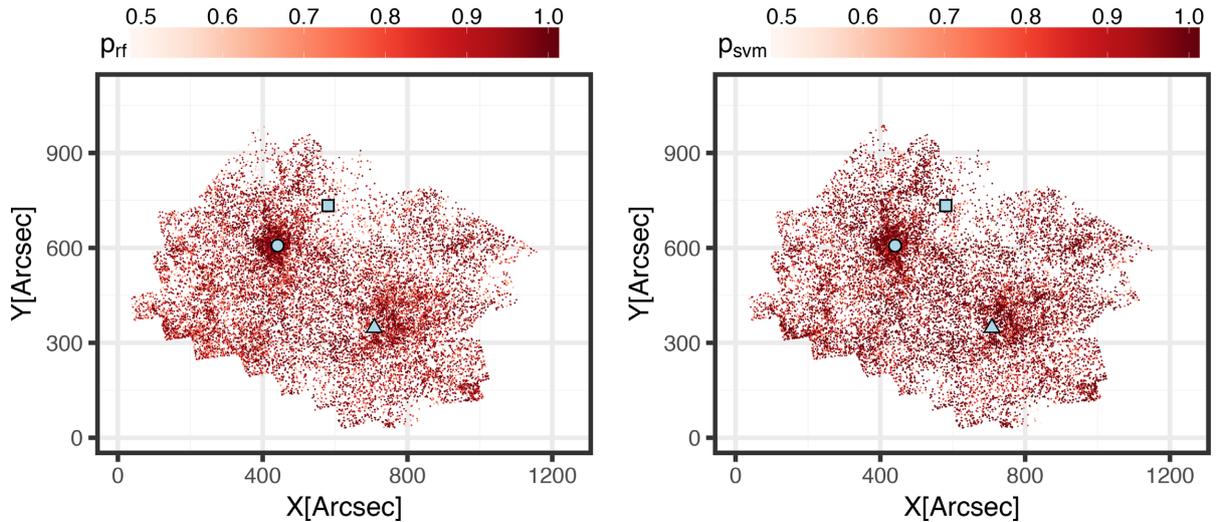
Figure 18. Spatial distribution of PMS candidates, i.e. all stars with $p_{rf/svm} \geq 0.5$, coloured according to their probability of being PMS stars, as predicted by the RF algorithm (*left*), and the SVM algorithm (*right*), trained on their measurements in the *F555W* and *F775W* bands and their extinction A_v. For guidance, the positions of R136, Hodge 301, and NGC 2060 are indicated by a large circle, square, and triangle, respectively.

terms of spatial distribution of the identified PMS stars. The spatial distribution of the RF model based on features set no. 5 was somewhat different than the rest, due to the drastically reduced amount of data to be classified and the smaller available spatial coverage of stars found in both optical and infrared bands. The PMS stellar samples identified with the models of sets no. 2 to 4 are essentially identical, with $\sim 20\,000$ common identifications. Interestingly, the RF model trained on set no. 3 identifies fewer candidate PMS stars than those predicted by the model of set no. 2 in regions of lower extinction. This indicates that the algorithm intrinsically assumes a spatial correlation between PMS stars and larger extinction, possibly due to the region it was trained on. The marginal differences between the models for the features sets no. 2 and 4 indicate that the RF method was not sensitive to the enlargement of the feature space with the addition of the photometric errors, possibly because the photometric errors add small decision power to the models.

5.3 Classification with support vector machines

The third classification algorithm we experimented with is the SVM (see Appendix A3 for a description). In our experiments the general purpose *Gaussian Radial Basis Kernel* was chosen as the SVM kernel:

$$K(x, x') = \exp(-\sigma ||x - x'||^2). \qquad (6)$$

The SVM model parameters, i.e. the cost C and the kernel width σ, are determined again via a 10-fold cross-validation, choosing the best model according to its AUC. The influence of the chosen p_{em} threshold for the best PMS candidates in the training sample was evaluated in the same way as for the RF. As in the case of the RF, we found that the performance of the SVM increases with a higher threshold, but only slightly, indicating that the performance of the modelling is not sensitive to this threshold. For the features set no. 3, the best AUC of 0.995 was achieved when training on PMS stars with $p_{em} \geq 0.9$, being only slightly larger though than the AUC values derived with other p_{em} thresholds, varying between 0.7 and 0.9. With the same reasoning as for the final RF classification we use

the condition of $p_{em} \geq 0.85$ for PMS candidacy here, also allowing a direct comparison between the results of the two methods. The corresponding accuracy of this SVM model is exceptionally good, equal to 97.29 per cent. The classification of the available HTTP data returns 19 487 PMS candidates with outcome probabilities $p_{svm} \geq 0.5$.

Fig. 17 (right) shows the optical CMD of the PMS candidates, coloured according to their probabilities p_{svm}. This CMD demonstrates that the SVM overall constructs a much smoother decision boundary than the RF and avoids successfully the misclassification of RC stars and objects with poor photometry as PMS stars. We also found that the SVM performs equally well as the RF in distinguishing possible turn-on and subgiant stars in their overlap CMD-region. There is, however, a small isolated patch of SVM-classified PMS candidates at ($F555W \sim 19.5$, $F555W$–$F775W \sim 0.5$), which could possibly be young stars still on the turn-on, but they fit mostly to the UMS. This patch is likely the result of the SVM being prone to overfit outliers, when there are very few records to train on, as is the case for this CMD region in our training data set (e.g. Fig. 16). Fig. 18 (right) shows the spatial distribution of the PMS candidates, as classified by the SVM. This map displays the same prominent spatial features as the PMS stars found with the RF, i.e. high concentrations of stars in the regions of NGC 2070 and NGC 2060, and indications of substructure in the space between the clusters.

We found somewhat larger variations in the SVM than in the RF across the various investigated features sets. The SVM model constructed from the features set no. 2 results in a very smooth two-dimensional decision boundary. As a consequence for this features set the algorithm does not distinguish possible turn-on and subgiant stars. Instead it classifies the entire CMD region up to the brightest limit of our training data set as PMS stars, albeit at a low overall probability. This certainly introduces a number of misclassifications of older subgiant stars and explains the larger number of identified PMS stars in comparison to the RF. Enlarging the features space with the inclusion of photometric errors in the *F555W* and *F775W* bands (i.e. features set no. 4) delivers a comparable performance and number of PMS candidates in comparison with features sets

no. 2. Similar to the RF the overall classification outcome does not appear to be overly sensitive to the inclusion of photometric errors, given the ~19 000 commonly identified PMS candidates, although the decision boundaries in the CMD appear slightly broadened. In an earlier test with a training set that did not account for highly extincted giant and subgiant stars, we found that the SVM trained on this features set misclassified a large amount of very faint blue LMS stars with larger photometric errors as PMS stars. Even though this does not happen to the same degree (there are still a few such misclassifications of faint LMS stars) with the more refined training set, described in this paper, we came to the conclusion, keeping those earlier tests in mind, that including the photometric errors is potentially compromising rather than helping the successful classification of PMS stars.

The classification result for features set no. 5, i.e. including the infrared measurements, is similar to the result for set no. 2, exhibiting a relatively smooth decision boundary in the CMD. There is a slight improvement, though, over features set no. 2 in the distinction between giant/subgiant and turn-on stars, indicating that infrared bands hold information that may be very useful in distinguishing possible MS turn-off and giant/subgiant stars from PMS turn-on stars, using machine-learning techniques. Nevertheless, we choose features set no. 3 as the primary set to base our final classification on due to the far more complete spatial coverage of the Tarantula Nebula over the other sets.

5.4 Comparison and combination

Both the RF and SVM methods scored an excellent performance on the held-out test data set for our primary training features set (set no. 3, $F555W$, $F775W$, A_V), with the SVM providing the best modelling across all our experiments in terms of accuracy. Despite their performance, we identified individual shortcomings in the classification results of both methods, as described in the previous two sections. The RF method inherited from its decision trees the trend to produce a zig-zag pattern (faint nevertheless) at the LMS–PMS CMD border, while the SVM includes in its sample of best PMS candidates the few members of an isolated patch at the faint part of the UMS. In order to overcome these shortcomings and identify the most accurate PMS stellar census across the entire Tarantula Nebula in terms of eliminated misclassifications, we combine the results of both methods.

A comparison between the results of the RF and the SVM methods is shown in Fig. 19, where each star in the optical CMD is colour-coded according to the difference of the probabilities derived from each method that this is indeed a PMS star, $\delta p = p_{svm} - p_{rf}$. An instance where the star is classified by the SVM as an excellent PMS candidate but not identified at all by the RF would have positive δp with a value close to 1, while in the opposite case this difference would be close to -1. On the other hand, instances in which both methods agree on the predicted probability of the stars being PMS stars would have differences δp close to zero. These are the records with the best prediction about their nature as PMS stars. Both methods provided the same classification and probabilities for the vast majority of the identified PMS stars, 17 728 stars in total. The optical CMD with the stars colour-coded according to their δp value is shown in Fig. 19. The red points in this CMD indicate the stars identified by the RF as PMS candidates but not by the SVM, while the blue points indicate stars classified by the SVM but not by RF. Discrepancies of the two methods are located around their respective decision boundaries between the PMS and non-PMS classes, but they are very few in comparison to the to-
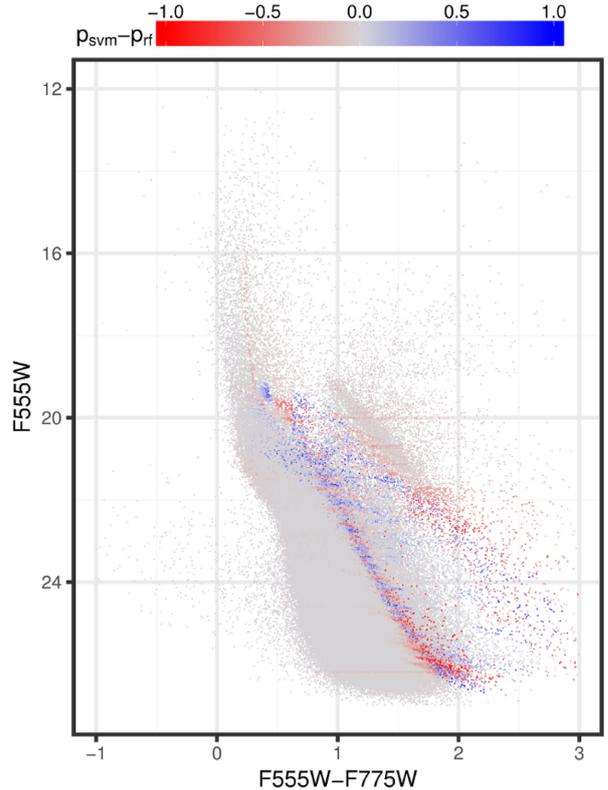
Figure 19. Optical CMD of the HTTP data, where each star is coloured according to the difference $p_{svm} - p_{rf}$ of the predicted probability to belong to the PMS from SVM and RF, respectively.

tal classified PMS stars. This CMD indicates that the classification of the two methods agrees well for the majority of the classified objects.

In general the SVM classification appears to treat the faintest part of the RC, falsely identified by the RF as PMS stars, better than the RF, while the RF classifies the lower UMS patch of stars, classified as PMS by the SVM, as negative instances. Each method, thus, 'corrects' for the shortcomings of the other. As a consequence, in order to achieve the most robust solution for the PMS stellar content of the Tarantula Nebula (in terms of producing the purest possible sample) we combine the individual classifications of each star by averaging the predicted PMS candidateship probabilities derived from both methods. This approach effectively compensates for the individual shortcomings of SVM and RF by assigning low mean PMS candidateship probabilities $\bar{p}$ to the likely misclassified objects, such as the UMS stars for the SVM and the RC stars for the RF. The CMD of Fig. 20 shows the improved final classification provided by this averaging. We construct the final catalogue of PMS candidate stars across the entire Tarantula Nebula, by providing the original HTTP photometric data of the stars, their predicted PMS candidateship probabilities, derived from both the SVM and RF methods, the difference of the probabilities δp, and the mean PMS candidateship probability $\bar{p}$. In this catalogue we include all stars, identified by at least one of the two methods as PMS candidates, delivering in total 22 755 possible PMS stars for the entire Tarantula Nebula. Imposing thresholds on the mean probability to distinguish the most probable PMS stellar population of the complex, this catalogue entails 19 831 candidates with $\bar{p} \geq 0.5$, 16 696 with $\bar{p} \geq 0.7$,
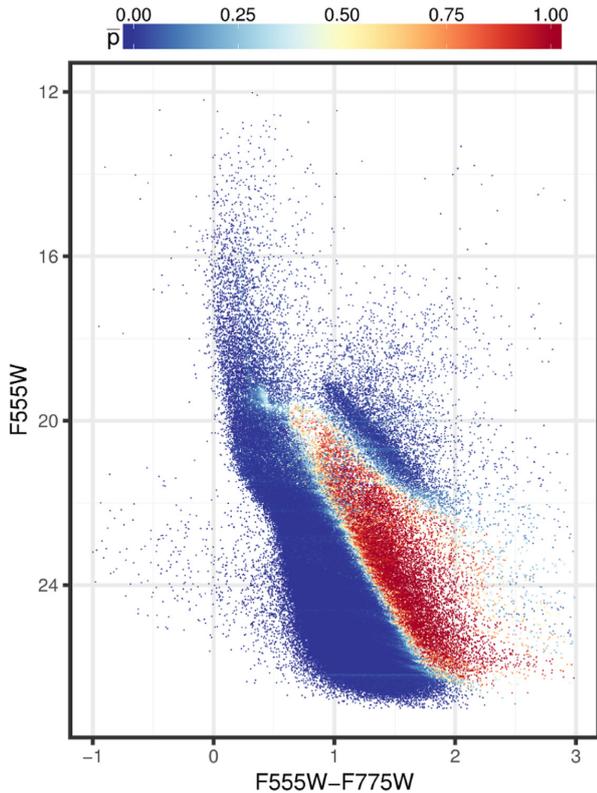
Figure 20. Optical CMD of the HTTP data, where each star is coloured according to the mean $\bar{p}$ of the predicted PMS candidateship probability of the SVM and RF.

13 526 for $\bar{p} \geq 0.85$, and 9636 with $\bar{p} \geq 0.95$. The corresponding spatial distributions are shown in Fig. 21. This series of thresholding the mean predicted PMS candidateship probability exhibits that the most probable PMS candidates, identified with our classification approach, also mark the most spatially confined structures across the entire Tarantula Nebula.

Fig. 22 shows the surface density map of the most probable PMS candidates ($\bar{p} \geq 0.95$). This map is not qualitatively different from those constructed for stars with different probability limits, so that the general clustering of PMS stars does not appear to be very sensitive to the threshold on $\bar{p}$. The remarkable coincidence of the maps of Fig. 21 (independently of the considered candidateship probability threshold) and the density map of Fig. 22 with the spatial distribution of the UMS stars (shown in Fig. 3) indicates that PMS stars are preferably clustered in regions of high concentrations of UMS stars. This is in agreement with results discussed in the literature concerning the clustering of PMS stars around massive young stars (Cignoni et al. 2015; Stephens et al. 2017), and provides an additional confirmation of the validity of our PMS identifications.

6 SUMMARY AND FUTURE PROSPECTS

In this paper we present our analysis with the employment of machine-learning classification techniques for the identification of PMS stars across the entire star-forming complex of the Tarantula Nebula in the LMC. For this classification we extracted a robust training subset from the observational data of the HTTP, which provides deep panchromatic Hubble imaging of the whole nebula, in

order to teach Naive Bayes classifier, decision tree, RF, and SVM classifiers to categorize the stars of the entire HTTP catalogue into the classes 'PMS' and 'non-PMS'. To construct this training data set we selected a high-surface-density region within the Tarantula Nebula, corresponding to the R136 starburst cluster at the heart of NGC 2070, based on the assumption that PMS stars are more likely to be located in the most clustered regions of the nebula. To account for differential extinction across the nebula, we used UMS stars as extinction probes and derived extinction measures for each individual star in the HTTP catalogue using a distance weighted average of the extinction of the 20 nearest UMS neighbours.

After attributing extinction to the NGC 2070 subset, improving upon the approach of Gouliermis et al. (2012), we developed a robust method to distinguish the cluster PMS stars from the field LMS stars in the training data set. This method is based on fitting bimodal Gaussian mixture models to the distance of all stars from the apparent gap on the CMD between these two populations via the maximum likelihood EM algorithm. From these mixture models we derived a probability p_{em} for each star in the training set to be PMS. We finalized the training set by adding further examples of evolved populations, such as RC stars, and subgiant stars in low- and high-extinction areas of the field of the Tarantula Nebula, as 'non-PMS'. We assigned the labels 'PMS' and 'non-PMS' to the stars depending on various selected thresholds of p_{em}, and after training the classification algorithms with this training set, we evaluated their performance for different variables (features) combinations. The findings of these experiments can be summarized as follows:

(i) During our preliminary tests neither the Naive Bayes nor the decision tree method were able to achieve adequate performance, providing accuracies not higher than ~ 60 and ~ 84 per cent, respectively. Consequently, both methods exhibited significant issues in classifying the entire HTTP catalogue, although the decision tree still provided valuable insights on the importance of specific features. It strongly suggests that near infrared measurements (e.g. in the $F160W$ filter) are very useful to a classification approach for PMS stars.

(ii) The best combination of features, in terms of stellar numbers, spatial coverage and algorithm performance, included the photometric measurements in the $F555W$ and $F775W$ filters in combination with the extinction values A_V, which we derived for each star using the UMS stars as extinction probes. Including the infrared wavebands resulted in a comparable performance of the classification algorithms, but, since the features set of the $F555W$, $F775W$, $F110W$, and $F160W$ filters suffered from the smaller spatial coverage of the infrared observations, it was not suited for finding the most complete PMS stellar census of the Tarantula Nebula. Extending the optical bands feature space by adding the photometric errors did not seem to provide any useful information for the RF, and it even compromised the classification ability of SVM.

(iii) The best performance of both the RF and SVM methods was achieved when stars with $p_{\mathrm{em}} \geq 0.9$ were selected as the best PMS examples in the training set. However, the best trade-off between algorithm performance and balance between the numbers of positive and negative examples in the training set was achieved with the use of a threshold $p_{\mathrm{em}} = 0.85$ for labelling the training stars.

Both the RF and SVM methods performed excellently on our primary features set ($F555W$, $F775W$, A_V), achieving accuracies of 95.6 and 97.3 per cent, and ROC AUCs of 0.990 and 0.994, respectively. The classification outcomes of both methods on the entire HTTP data also met the required expectations, except for minor shortcomings. Specifically, the RF algorithm misclassified a few
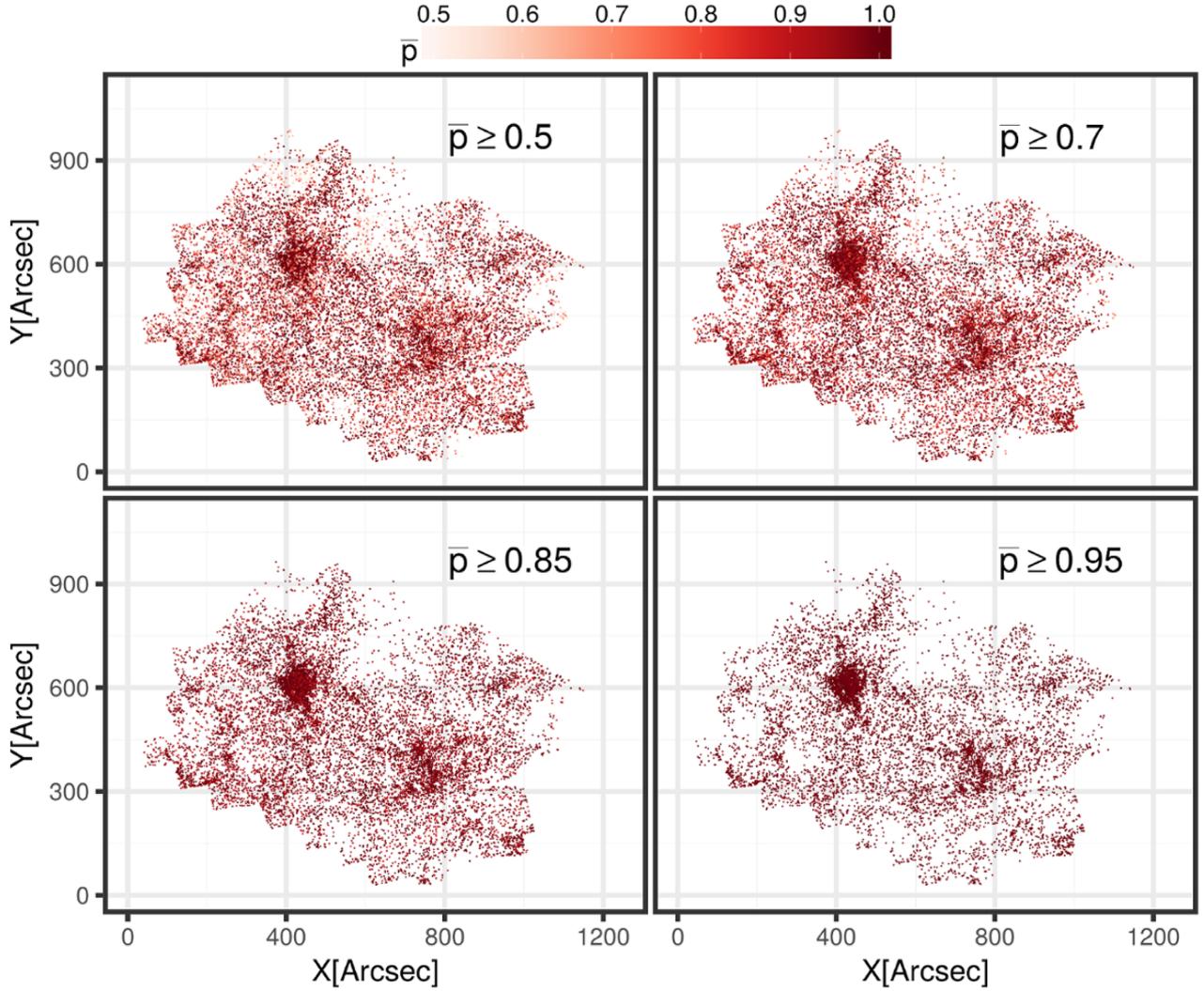
Figure 21. Series of spatial distribution plots of PMS candidates, determined by different thresholds of the mean predicted probability $\bar{p}$ of SVM and RF. The respective threshold is indicated in each plot, where all stars are coloured according to $\bar{p}$.

faint RC stars as PMS stars, and the SVM did so for a small patch of likely UMS stars. A direct comparison between the outcomes of the methods showed that they compensate each other's shortcomings. As a consequence, the most robust classification is achieved by combining the predicted PMS candidateship probability of each star derived from both methods.

The combination of the results of both RF and SVM methods resulted in 22 755 stars, identified as PMS by at least one of the methods. Among these sources, 19 831 stars have an average predicted PMS probability of $\bar{p} \geq 0.5$ and 9636 have $\bar{p} \geq 0.95$. There is a number of studies that can be performed with the use of this catalogue of the most probable low-mass PMS stars in the Tarantula Nebula region. We identify three science cases, each deserving its own independent investigation.

(1) *PMS stars with emission lines due to accretion.* The use of the H_α filter permits the identification of several types of young stellar sources, such as massive main-sequence and supergiant emission-line stars (Oe, Be, and B[e] stars), as well as PMS stars with strong emission lines, such as classical T Tauri stars (Appenzeller & Mundt 1989; Bertout 1989). The excess in these stars implies that the photospheric lines are not as deep as those of main-sequence stars of the same spectral type (e.g. Hartigan, Edwards & Ghandour 1995; Gullbring et al. 1998). With our classification we find that ~ 60 per cent of the PMS candidates are also detected in the H_α band, while only ~ 30 per cent of the stars classified as non-PMS are found in this band. Our preliminary study of these stars indicates that most of the PMS stars do not show strong accretion, which is expected for Weak Line T Tauri stars that show very weak, if any emission lines (e.g. Montmerle et al. 1993). A more detailed analysis of the H_α excess of all these stellar types in order to determine their mass accretion rates and investigate their variations across the nebula will be the topic of a separate study.

(2) *Physical characteristics of PMS stars.* The use of the HTTP multiband photometry will allow us to construct the spectral energy distributions (SEDs) of our low-mass PMS candidates in order to establish their masses and ages through dedicated SED-fitting techniques. This study for the determination of physical parameters for the identified PMS stars is currently under development with
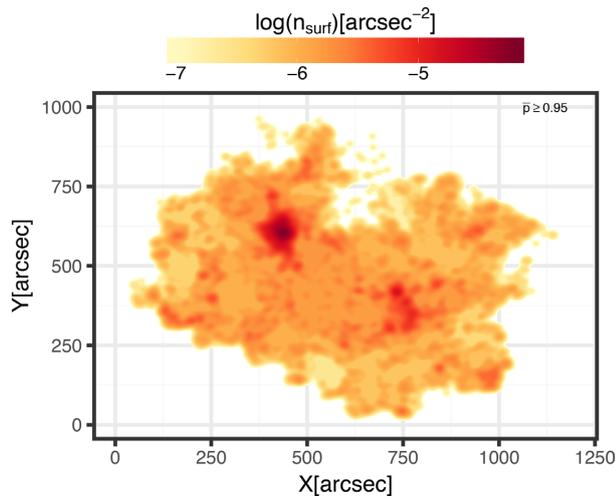
Figure 22. Surface density plot of the PMS candidate stars with $\bar{p} \geq 0.95$, i.e. the population in the bottom right panel in Fig. 21.

the use of the *Bayesian Extinction and Stellar Tool* (Gordon et al. 2016), appropriately tailored to our photometric data (Ksoll et al., in preparation). These results will further allow the characterization of the stellar IMF across its whole dynamic range and its variability across the HTTP FoV by combining mass estimates of our PMS stars with those of the UMS populations (Evans et al. 2011). We will be further able to investigate the propagation of star formation in time with the investigation of spatial distributions of ages of the PMS stars.

(3) *The clustering pattern of star formation*. The spatial distribution of the classified PMS stars shows well defined sub-structures within the regions of the clusters NGC 2070 and NGC 2060, as well as compact and loose – occasionally filamentary – clusterings across the whole observed FoV. An elaborate investigation of the clustering behaviour of the PMS stars in the Tarantula Nebula, based on the results of this study, and the quantification of the spatial cross-correlation between PMS and UMS stars is currently being performed in a separate forthcoming study (Gouliermis et al., in preparation).

REFERENCES

Appenzeller I., Mundt R., 1989, A&AR, 1, 291
Beaumont C. N., Goodman A. A., Kendrew S., Williams J. P., Simpson R., 2014, ApJS, 214, 3
Benaglia T., Chauveau D., Hunter D., Young D., 2009, J. Stat. Softw., 32, 1
Bertout C., 1989, ARA&A, 27, 351
Bodenheimer P. H., 2011, Principles of Star Formation. Springer-Verlag, Berlin
Brandner W., Grebel E. K., Barbá R. H., Walborn N. R., Moneti A., 2001, AJ, 122, 858
Breiman L., 2001, Mach. Learn., 45, 5
Breiman L., Friedman J. H., Olshen R. A., Stone C. J., 1984, Classification and Regression Trees. Statistics/Probability Series. Wadsworth Publishing Company, Belmont, California
Bressan A., Marigo P., Girardi L., Salasnich B., Dal Cero C., Rubele S., Nanni A., 2012, MNRAS, 427, 127
Cignoni M. et al., 2015, ApJ, 811, 76
Cignoni M. et al., 2016, ApJ, 833, 154
Cignoni M., Tosi M., Sabbi E., Nota A., Degl'Innocenti S., Prada Moroni P. G., Gallagher J. S., 2010, ApJ, 712, L63
Cortes C., Vapnik V., 1995, Mach. Learn., 20, 273
Cutri R. M. et al., 2012, American Astronomical Society, AAS Meeting #219, 401.06
Da Rio N., Gouliermis D. A., Henning T., 2009, ApJ, 696, 528
Da Rio N., Gouliermis D. A., Gennaro M., 2010, ApJ, 723, 166
De Marchi G. et al., 2016, MNRAS, 455, 4373
De Marchi G., Panagia N., Beccari G., 2017, ApJ, 846, 110
Dempster A. P., Laird N. M., Rubin D. B., 1977, J. R. Stat. Soc. B, 39, 1
Dieleman S., Willett K. W., Dambre J., 2015, MNRAS, 450, 1441
Elorrieta F. et al., 2016, A&A, 595, A82
Evans C. J. et al., 2011, A&A, 530, A108
Glatt K., Grebel E. K., Koch A., 2010, A&A, 517, A50
Gordon K. D. et al., 2016, ApJ, 826, 104
Gouliermis D. A., 2012, Space Sci. Rev., 169, 1
Gouliermis D., Brandner W., Henning T., 2006, ApJ, 636, L133
Gouliermis D. A., Henning T., Brandner W., Dolphin A. E., Rosa M., Brandl B., 2007, ApJ, 665, L27
Gouliermis D. A. et al., 2010, Astrophysics and Space Science Proceedings, Vol. 15, The Impact of HST on European Astronomy. Springer-Verlag, Berlin, p. 71
Gouliermis D. A., Schmeja S., Dolphin A. E., Gennaro M., Tognelli E., Prada Moroni P. G., 2012, ApJ, 748, 64
Gouliermis D. A., Hony S., Klessen R. S., 2014, MNRAS, 439, 3775
Grebel E. K., Chu Y.-H., 2000, AJ, 119, 787
Gullbring E., Hartmann L., Briceño C., Calvet N., 1998, ApJ, 492, 323
Hartigan P., Edwards S., Ghandour L., 1995, ApJ, 452, 736
Haschke R., Grebel E. K., Duffau S., 2011, AJ, 141, 158
Hastie T., Tibshirani R., Friedman J. H., 2009, The Elements of Statistical Learning. Springer Series in Statistics, 2nd edn. Springer -Verlag, New York, NY
Hony S. et al., 2015, MNRAS, 448, 1847
Hunter D. A., Shaya E. J., Holtzman J. A., Light R. M., O'Neil E. J.,, Jr, Lynds R., 1995, ApJ, 448, 179

James G., Witten D., Hastie T., Tibshirani R., 2014, An Introduction to Statistical Learning: With Applications in R. Springer, Heidelberg, Dordrecht

Karatzoglou A., Smola A., Hornik K., Zeileis A., 2004, J. Stat. Softw., 11, 1

Kuhn M., 2017, caret: Classification and Regression Training, available at: https://CRAN.R-project.org/package=caret

Liaw A., Wiener M., 2002, R News, 2, 18

Massey P., Hunter D. A., 1998, ApJ, 493, 180

McLachlan G. J., Peel D. A., 2000, Finite Mixture Models. Wiley Series in Probability and Statistics: Applied Probability and Statistics Section. Wiley, New York; Weinheim

Meyer D., Dimitriadou E., Hornik K., Weingessel A., Leisch F., 2015, e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien, available at: https://CRAN.R-project.org/package=e1071

Montmerle T., Feigelson E. D., Bouvier J., Andre P., 1993, in Levy E. H., Lunine J. I., eds, Protostars and Planets III. Univ. Arizona Press, Tucson, p. 689

Nota A. et al., 2006, ApJ, 640, L29

Panagia N., Gilmozzi R., Macchetto F., Adorf H.-M., Kirshner R. P., 1991, ApJ, 380, L23

Panagia N., Romaniello M., Scuderi S., Kirshner R. P., 2000, ApJ, 539, 197

Platt J. C., 1999, Advances in Large Margin Classifiers. MIT Press, Cambridge, MA, p. 61

R Core Team 2013, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, available at: http://www.R-project.org/

Romaniello M., Panagia N., Scuderi S., Kirshner R. P., 2002, AJ, 123, 915

Russell S. J., Norvig P., Davis E., 2016, Artificial Intelligence, 3rd edn, global edition. Always Learning, Pearson, Boston; Munich

Sabbi E. et al., 2007, AJ, 133, 44

Sabbi E. et al., 2008, AJ, 135, 173

Sabbi E. et al., 2013, AJ, 146, 53

Sabbi E. et al., 2016, ApJS, 222, 11

Schmalzl M., Gouliermis D. A., Dolphin A. E., Henning T., 2008, ApJ, 681, 290

Schmeja S., Gouliermis D. A., Klessen R. S., 2009, ApJ, 694, 367

Schulz N. S., 2012, Astronomy and Astrophysics Library, The Formation and Early Evolution of Stars: From Dust to Stars and Planets. Springer-Verlag, Berlin

Stahler S. W., Palla F., 2005, The Formation of Stars. Wiley, New York

Stephens I. W. et al., 2017, ApJ, 834, 94

Szeliski R., 2011, Computer Vision: Algorithms and Applications. Springer-Verlag, Berlin

Taylor M. B., 2005, in Shopbell P., Ebert R., eds, ASP Conf. Ser. Vol. 347, Astronomical Data Analysis Software and Systems XIV. Astron. Soc. Pac., San Francisco, p. 29

Therneau T., Atkinson B., Ripley B., 2017, rpart: Recursive Partitioning and Regression Trees, available at: https://CRAN.R-project.org/package=rpart

Venables W. N., Ripley B. D., 2002, Modern Applied Statistics with S. Statistics and Computing, 4th edn. Springer -Verlag, New York; Berlin; Heidelberg

Walborn N. R., Blades J. C., 1997, ApJS, 112, 457

Zaritsky D., Harris J., Thompson I. B., Grebel E. K., Massey P., 2002, AJ, 123, 855

APPENDIX A: CLASSIFICATION ALGORITHMS

The following sections will give an overview over the three classification algorithms; decision tree, RF classifier and SVM; which are used in Section 5 for the identification of PMS stars in the Tarantula Nebula, as well as two performance measures – the confusion matrix and the ROC curve – for these algorithms. All of the described methods are so-called *supervised learning techniques*, because they require a labelled data set to be trained on in order to perform a classification of new data. Here, 'labelled' means that this training set consists of example observations, for which the class is known. In a two class scenario the examples of the class of primary interest, in our case the PMS stars, are often called '*positives*', and examples that do not belong to that class are called '*negative*' (i.e. LMS, UMS, red clump stars, etc., for us). In this context the space spanned by all possible values of the attributes of an object is usually referred to as the *feature space*, where an individual instance is represented by a feature vector, a vector that contains all its respective attributes.

The algorithms are applied with implementations in *R*, a popular language and environment for statistical computing and graphics (R Core Team 2013). Specifically, packages e1071 (Meyer et al. 2015) and RPART (Therneau, Atkinson & Ripley 2017) are used for the application of *Naive Bayes*, and *decision trees*, respectively. For the application of *RF* and *SVM* we used the *R* package CARET (Kuhn 2017), which invokes package RANDOMFOREST (Liaw & Wiener 2002) for the former and package KERNLAB (Karatzoglou et al. 2004) for the latter method.

A1 Decision trees

The general idea of a *decision tree classifier* or *classification tree* (Breiman et al. 1984) is that, ideally, it is possible to partition the feature space such that all object instances will be correctly classified. Thus, the result corresponds to a hierarchical partition of the feature space. This partition is represented by the end point (called *leaves* or *terminal nodes*) of a tree, where each node of the tree splits the feature space according to the value of a certain attribute. Interpreting the tree as a probability model, each node i of the tree possesses a probability distribution p_{ik} over the classes k. After building the tree, each case in the training set is assigned to one leaf, so that each leaf has a random sample N_{ik} from the distribution p_{ik} (Venables & Ripley 2002).

The decision tree is constructed by recursively splitting the feature space until a stopping criterion is reached. At a certain node a split (usually a binary split, separating a continuous variable x_j into $x_j < t$ and $x_j \geq t$) is chosen according to a measure of its value. Most commonly a measure of *impurity* is defined for each node. A node is considered to be pure if it only contains instances of a single class. A widely used impurity measure is the *Gini index*:

$$G = \sum_{j \neq k} p_{ij} p_{ik} = 1 - \sum_k p_{ik}^2, \tag{A1}$$

which measures the training error rate of classifying object instances in a node to class k instead of the majority class of the node. For a pure node the Gini index is zero. Based on this impurity measure, at each node the split that reduces most the average impurity is performed. Stopping criteria can be when all nodes become pure, the tree reaches a maximum predefined depth or further splitting cannot reduce the average impurity more than a given minimal threshold. If the tree construction is stopped before all nodes become pure, the terminal nodes assume the majority class of their assigned training instances (Venables & Ripley 2002). Fig. A1 shows an example decision tree model, constructed on our training set using only *V* and *R* magnitudes (see Section 5.1).

A1.1 Prediction

To classify a new object instance, the decision tree propagates it according to its attributes along the tree, starting from the root,

Figure A1. Schematic representation of a pruned decision tree model, trained on the *V*- and *R*-equivalent magnitudes of the stars. Each node is labelled according to the majority class of observations in the node and shows furthermore the prediction probability for PMS, as well as the percentage of total observations assigned to the node.

until a corresponding terminal node is reached, assigning the respective nodes class to the new object instance (Hastie et al. 2009). To some degree the decision tree can compensate for missing attributes of an object, by either assigning the majority class of the deepest non-terminal node reached with the available attributes or using *surrogate splits*. In the latter method each non-terminal node keeps a list of surrogate splits during the tree construction. During classification, if the primary split attribute is missing, one of these surrogate attributes is used to propagate the object instance further along the tree. Surrogate splits are constructed as follows: During the construction of the tree when considering a certain attribute for a split only those training instances are considered, which are not missing that attribute. Afterwards a list of surrogate attributes and split points is generated, sorted according to how well this surrogate split approximates the split by the primary attribute. During prediction, surrogate splits are considered in that same order (Hastie et al. 2009).

A1.2 Tree pruning

If the training data are noisy, i.e. the class distributions overlap in feature space, a decision tree might overfit and thus perform badly on a set of new object instances. In order to avoid this one employs a method called *cost-complexity pruning* (James et al. 2014). In this approach rooted subtrees of the decision tree are constructed by removing terminal subtrees. Then each of these subtrees is assigned a value R, which is the sum of some measure R_i of the leaves of the tree. The size of these trees is equal to their number of leaves. One can now show that a set of rooted subtrees of tree T, which minimizes the cost-complexity measure:

$$R_\alpha = R + \alpha \times \text{size} \tag{A2}$$

is a nested tree. With increasing α one can find the optimal trees by a series of snip (i.e. cutting terminal subtrees) operations on the current tree, producing a sequence of trees with sizes of T down to just the root node. To choose the desired degree of pruning one computes an impurity measure versus α for the pruned tree and finds the smallest tree close to the minimum of the impurity measure when predicting on a separate validation set or using cross-validation (Venables & Ripley 2002).

A2 Random forest

The basic concept of the *RF classifier* (Breiman 2001) is the so-called *bagging*, a general purpose procedure for variance reduction of statistical models through averaging many models of high variance and low bias (James et al. 2014). Decision tree models suffer from high variance, e.g. trees that fit to randomly determined halves of the same training data could vary significantly from one another, but provide low bias, if grown deep enough. Deep un-pruned decision trees are the underlying model of the *RF classifier*. Following the principles of bagging, an RF is constructed by building B individual trees, which are grown by *bootstrapping* from the training data, i.e. taking repeated samples from the single training set, generating B different bootstrapped training subsets (Hastie et al. 2009; James et al. 2014).

Improving upon a simple bagging of decision trees, the RF further increases the variance reduction by a modification of the tree construction procedure. Instead of choosing the split attribute that reduces the impurity measure the most, m random attributes out of the available p are selected and the best variable and split point are determined out of those. A small value for m is typically helpful if a large number of the attributes are correlated. For classification purposes a general choice is $m = \sqrt{p}$ (see e.g. Hastie et al. 2009). The

tree growth then proceeds until a minimum node size is reached. This procedure *decorrelates* the trees by preventing strong predictor attributes to dominate the split selection in all trees grown,[7] thus increasing the overall variance reduction of the bagging approach by averaging many *uncorrelated* models (James et al. 2014). In order to classify a new object instance the RF classifier casts a *majority vote* over all the trees it has grown, i.e. each individual tree classifies the object, counting the results and assigning the class most voted for (Hastie et al. 2009).

A3 Support vector machine

The SVM (Cortes & Vapnik 1995) is a classifier that produces a non-linear decision boundary in feature space by constructing a linear boundary in a transformed version of the feature space. It is a generalization of the *support vector classifier*, which itself is based on the *maximal margin classifier* (Hastie et al. 2009; James et al. 2014).

A3.1 Maximal margin and support vector classifier

The underlying concept of the maximal margin classifier is the *optimal separating hyperplane*, a hyperplane being a flat affine $p - 1$ dimensional subspace of a p-dimensional space (e.g. a line in 2D or a plane in 3D space), describing the solution space to a set of linear equations

$$x^T \beta + \beta_0 = 0, \tag{A3}$$

where β and β_0 denote a vector of coefficients and a constant vector, respectively. Given a set of n p-dimensional training instances x_i, which fall into two classes with labels $y_i \in \{-1, 1\}$, in a classification context a *separating hyperplane* describes a hyperplane constructed such that it perfectly separates all training instances according to their class labels, i.e. having the property

$$y_i(\beta_0 + x_i^T \beta) > 0 \tag{A4}$$

for all $i = 1, \ldots, n$ (James et al. 2014). Such a hyperplane induces a natural classification rule for a new test instance x_* by assigning a class depending on which side of the hyperplane it is located, i.e. classifying x_* based on the sign

$$G(x_*) = \text{sign}[x_*^T \beta + \beta_0]. \tag{A5}$$

The *margin* is defined as the minimum of the (perpendicular) distances of all training instances to a given hyperplane (e.g. the distance from the plane of points 1, 2, and 7, ignoring points 3–6, as shown in Fig. A2, left-hand panel). The *optimal separating* or *maximal margin hyperplane* is the separating hyperplane for which the margin is the largest, i.e. the hyperplane that has the farthest minimum distance to the training instances. A classifier equation (A5) based on this hyperplane is a *maximal margin classifier*. Training instances that are equidistant from the maximal margin hyperplane and lie on the margin are called *support vectors*, as they 'support' the hyperplane in the sense that a variation of their position would change the hyperplane as well (James et al. 2014). To build this

classifier one has to find the maximal margin hyperplane as the solution to the optimization problem

$$\max_{\beta, \beta_0, ||\beta||=1} M, \tag{A6}$$

subject to

$$y_i(\beta_0 + x_i^T \beta) \geq M \ \forall i = 1, \ldots, n, \tag{A7}$$

where M, $M > 0$, represents the width of the margin.

The *support vector classifier* is a generalization of the maximal margin classifier for the case in which the training data is not linearly separable, i.e. when there is no solution to the optimization problem with $M > 0$ (Fig. A2, left-hand panel). The basic concept behind this method is a *soft margin*, which means that instead of constructing a hyperplane that *perfectly* separates the training instances, a hyperplane is built that allows some instances to be on the incorrect side of the margin or even of the hyperplane, i.e. a hyperplane that *almost* separates the classes. Such a hyperplane is the solution to the optimization problem

$$\max_{\beta, \beta_0, \epsilon_1, \ldots, \epsilon_n} M, \quad ||\beta|| = 1 \tag{A8}$$

subject to

$$y_i(\beta_0 + x_i^T \beta) \geq M(1 - \epsilon_i) \quad \epsilon_i \geq 0,$$
$$\sum_{i=1}^{n} \epsilon_i \leq C, \tag{A9}$$

where C, $C > 0$, is a tuning parameter, M is again the width of the margin, which is to be made as large as possible, and $\epsilon_1, \ldots, \epsilon_n$ are *slack variables*, allowing individual instances to fall on the wrong side of the margin or hyperplane. A value $\epsilon_i = 0$ signifies that the ith training instance is on the correct side of the margin, while $\epsilon_i > 0$ indicates that it is on the wrong side of the margin (*violating* the margin, e.g. points 3 and 4 in Fig. A2, left-hand panel). A value of $\epsilon_i > 1$ indicates that the instance is on the wrong side of the hyperplane (e.g. points 5 and 6 in the figure). The tuning parameter C (often called *cost*), bounding the sum $\sum \epsilon_i$, signifies a *budget* that determines the number and severity of tolerated margin violations. Consequently a large C, tolerating many margin violations, results in a wider margin, while a smaller C narrows it. Both this hyperplane, built by the support vector classifier, and the classifier itself are only dependent on training instances that lie directly on the margin (points 1, 2, 7 in Fig. A2, left-hand panel) or are violating it, i.e. its *support vectors* (James et al. 2014).

A3.2 Support vector machine

The support vector classifier is an effective tool for a two-class setting if the two classes can be divided by a linear boundary. In the scenario of non-linear class boundaries, however, it will perform poorly without modification, as indicated by the example shown in Fig. A2 (middle panel). To create such non-linear class boundaries with a support vector classifier one has to enlarge the feature space, by e.g. adding quadratic functions of the features. While the classifier is still linear within the enlarged feature space it was built in, it corresponds to a non-linear class boundary in the original feature space. This is the basic concept behind the SVM. The support vector classifier, i.e. the solution to the optimization problem of equations (A8), (A9), can be written as

$$f(x) = \beta_0 + \sum_{i=1}^{n} \alpha_i \langle x, x_i \rangle \tag{A10}$$

---

[7]Suppose there is a single strong predictor attribute along with a number of moderately strong ones. If we grow B decision trees with the standard procedure this strong predictor attribute would always be considered for the splits. With the random sampling procedure, however, on average $(p - m)/p$ splits will not even consider the dominant predictor for the split.
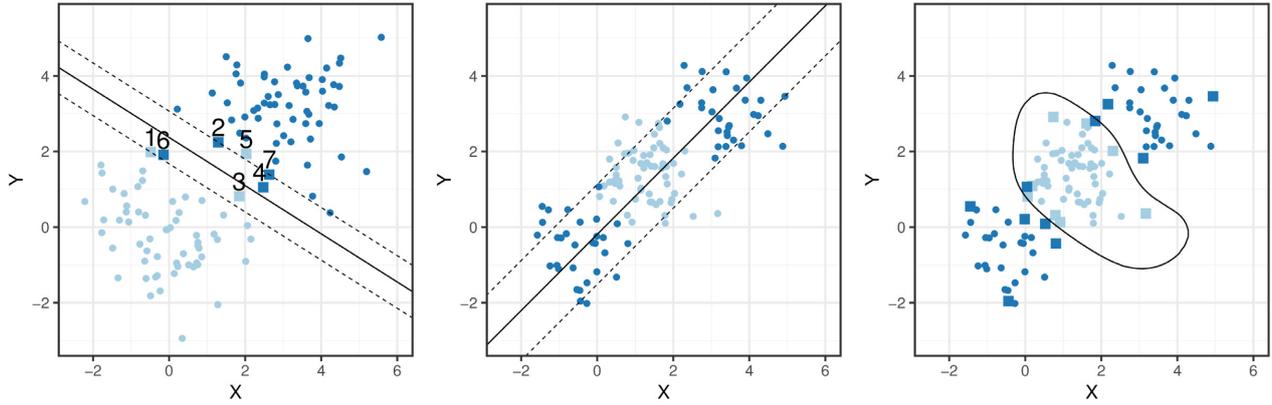
Figure A2. *Left:* Example of a support vector classifier fit to a small data set, distinguishing two classes of points in a 2D space. The black solid and dashed lines mark the constructed separating hyperplane and boundaries of the margin, respectively. The coloured squares indicate the support vectors of the classifier. The instances 1 and 2 are support vectors lying *on* the margin, while 3 and 4 are examples for instances being on the wrong side of the margin of their respective class. 5 and 6 are instances that are on the wrong side of the margin *and* the separating hyperplane. *Middle:* Support vector classifier fit to a small data set, where the class boundaries are non-linear. Here, the linear support vector classifier performs poorly. *Right:* SVM using a radial basis kernel fit to the same data set. The solid black line indicates the non-linear decision boundary of the constructed SVM, while the coloured squares mark the support vectors.

where $\langle x_i, x_{i'} \rangle = \sum_{j=1}^{p} x_{ij} x_{i'j}$ is the inner product and α_i ($i = 1, \ldots, n$) are n parameters, one per training instance, which in the solution are only nonzero for support vectors. One can now *generalize* the inner product with a *kernel function $K(x, x_i)$*. By choosing a *linear kernel* $\left(K(x_i, x_{i'}) = \sum_{j=1}^{p} x_{ij} x_{i'j} \right)$ we retrieve the normal support vector classifier, but if instead a *polynomial* or *radial* kernel function is chosen, we essentially fit a support vector classifier in a higher dimensional space, constructing a non-linear class boundary in the original feature space (see e.g. Hastie et al. 2009, for a full presentation of the calculation).

This combination of a support vector classifier with a non-linear kernel function is an *SVM*. The kernel 'trick' has the advantage of not only being computationally efficient, but also avoiding the necessity for an explicit transformation to the enlarged feature space and even allowing the latter to become infinite-dimensional, as e.g. is the case for the radial kernel (Hastie et al. 2009; James et al. 2014). Fig. A2 (right-hand panel) shows an example of the non-linear class boundary constructed by an SVM with a radial kernel on the data set, where the linear support vector classifier failed to construct a meaningful class boundary. To provide class probabilities instead of class labels, when using an SVM, one can use Platt's posterior probabilities, which fit a sigmoid function to the decision value *f* of the SVM

$$P(y = 1|f) = \frac{1}{1 + \exp(Af + B)}, \tag{A11}$$

where *A* and *B* are estimated by minimizing the negative log-likelihood function (Karatzoglou et al. 2004; Platt 1999).

A4 Training and performance measures

A4.1 Training with cross-validation

Cross-validation is the most commonly used method for training classification models and estimating their prediction error. Typically, a *k*-fold cross-validation is applied by (1) partitioning the data set into *k* equal-sized subsets, (2) training the algorithm on the total data of the *k* − 1 subsets, while holding out the remaining subset

to test its performance on, and (3) repeating step (2) *k* times, while holding out each of the subsamples for testing in each iteration. As a result, none of the *k* produced models has made predictions on its own training data. The model that predicts best among them is considered as the final classification model (see e.g. Hastie et al. 2009). While cross-validation is usually applied to test the modelling process, the evaluation of the performance of the *final model* is done via a Train/Test split of the training data set.

A4.2 Confusion matrix

The *confusion matrix* is a way to summarize the performance of a classification algorithm when predicting on a test set with known labels. It contains the following quantities:

(i) *True Positives (TP):* number of instances that are correctly predicted to be positives.

(ii) *False Positives (FP):* number of instances that are incorrectly predicted to be positives.

(iii) *True Negatives (TN):* number of instances that are correctly predicted to be negatives.

(iv) *False Negatives (FN):* number of instances that are incorrectly predicted to be negatives.

An example of the confusion matrix is given in Table A1. From the confusion matrix one can derive the *accuracy* performance measure by dividing the trace by the sum of all entries or calculating:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{P} + \text{N}}, \tag{A12}$$

where P and N denote the number of positive and negative instances in the training set, respectively. Further diagnostics that can be derived from the confusion matrix are the *true positive rate* (TPR),

Table A1. Example of a confusion matrix.

	Actual positive	Actual negative
Predicted positive	TP	FP
Predicted negative	FN	TN
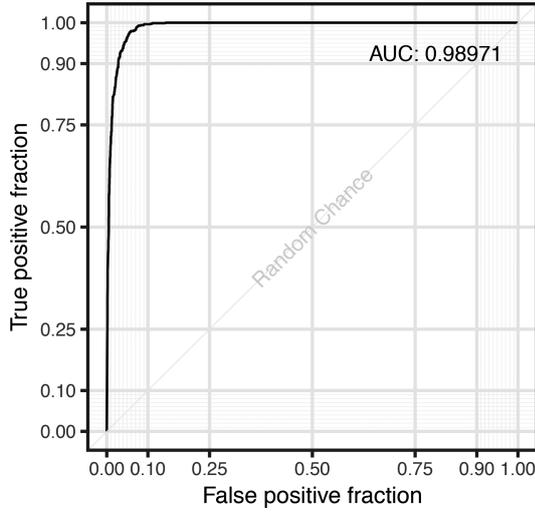
Figure A3. Example ROC curve of the RF on our features set no. 3 (see Section 5).

also called *Sensitivity*, and the *false positive rate* (FPR), *Specificity*, which are defined by

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{\text{TP}}{\text{P}}$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} = \frac{\text{FP}}{\text{N}}. \tag{A13}$$

The The *Balanced Accuracy* is an accuracy measurement that accounts for an imbalance in the number of positive and negative instances in the training set. Therefore it allows for an assessment of the class-specific accuracy and is defined as the mean of Sensitivity and Specificity (see e.g. Szeliski 2011)

$$\text{BACC} = \frac{1}{2}\left(\frac{\text{TP}}{\text{P}} + \frac{\text{FP}}{\text{N}}\right). \tag{A14}$$

A4.3 Receiver operating characteristic curve

The *ROC Curve* relates the *true positive rate* to the *false positive rate* for different parameters of the classification rule (such as the decision threshold). The closer the curve is to the top left corner, i.e. the larger the AUC, the better the algorithm performs. The ROC curve of a randomly guessing algorithm corresponds to a straight line with unit slope. Consequently the AUC is a commonly used quantitative summary of performance of an algorithms (Hastie et al. 2009; Szeliski 2011). Fig. A3 shows an example ROC curve from our analysis (see Section 5.2).

APPENDIX B: MIXTURE MODELS AND THE EM ALGORITHM

Mixture Models are a useful method for density estimation, with the most popular being the Gaussian mixture model of the form

$$f(x) = \sum_{m=1}^{M} \alpha_m \Phi(x; \mu_m, \Sigma_m), \tag{B1}$$

where α_m denotes the mixing proportions, subject to $\sum_m \alpha_m = 1$, M marks the total number of components and the individual Gaussian densities have a mean μ_m and covariance matrix Σ_m.

These parameters are usually fit by maximum likelihood with, for instance, the *Expectation Maximisation (EM) algorithm* (Dempster et al. 1977). A mixture model can then be used to provide an estimate of the posterior probability that a certain observation i belongs to a component m, given by

$$r_{im} = \frac{\alpha_m \Phi(x_i; \mu_m, \Sigma_m)}{\sum_{k=1}^{M} \alpha_k \Phi(x_i; \mu_k, \Sigma_k)}. \tag{B2}$$

When fitting a finite mixture model like equation (B1) to an observed random sample $\boldsymbol{x} = (x_1, \ldots, x_n)$ the log-likelihood from the data, which is to be maximised to retrieve the parameters of the model, takes the form

$$l(\theta; \boldsymbol{x}) = \sum_{j=1}^{n} \log \left(\sum_{i=1}^{M} \alpha_i \Phi_\theta(x_j; \theta_i) \right) \tag{B3}$$

where $\theta = (\alpha_1, ..., \alpha_M, \theta_1, \ldots, \theta_M)$ denotes all parameters of the model and $\theta_i = (\mu_i, \Sigma_i)$ the parameters of mixture component i. In practice, maximizing equation (B3) can be complicated numerically due to the sum in the logarithm. To alleviate this problem the *EM algorithm* treats it as an *incomplete data* problem. The observed data vector x is assumed to be incomplete, missing a set of associated component-label vectors $z = (z_1, \ldots, z_n)$, where each z_j is a M-dimensional vector with $z_{ij} = 1$ or 0, according to whether x_j belongs to component i. Thus, the complete data vector is $\boldsymbol{x}_c = (\boldsymbol{x}, \boldsymbol{z})$ with log-likelihood

$$l(\theta; \boldsymbol{x_c}) = \sum_{i=1}^{M} \sum_{j=1}^{n} z_{ij}[\log \alpha_i + \log \Phi_\theta(x_j; \theta_i)]. \tag{B4}$$

Based on this incomplete data assumption the EM algorithm proceeds iteratively, alternating between the *expectation* (*E*) and the *maximisation* (*M*) step (see e.g. McLachlan & Peel 2000; Benaglia et al. 2009; Hastie et al. 2009).

In the *E-step* the conditional expectation of the complete-data log-likelihood, based on the observed data **x** and the current fit $\theta^{(k)}$, expressed as the operator

$$Q(\theta'; \theta^{(k)}) = E(l(\theta', \boldsymbol{x}_c)|\boldsymbol{x}, \theta^{(k)}) \tag{B5}$$

is computed. For the finite mixture model equation (B4) shows that the complete-data log-likelihood is linear in the latent data z_{ij}, so that the E-step in iteration k+1 only requires to calculate the current conditional expectation of Z_{ij} (the random variable corresponding to z_{ij}) given the observations $\boldsymbol{x}$

$$E(Z_{ij}|\boldsymbol{x}, \theta^{(k)}) = p(Z_{ij} = 1|\boldsymbol{x}; \theta^{(k)}) = r_i(x_j; \theta^{(k)}), \tag{B6}$$

where following (B2)

$$r_i(x_j; \theta^{(k)}) = \frac{\alpha_i^{(k)} \Phi_\theta(x_j; \theta_i^{(k)})}{\sum_{h=1}^{M} \alpha_h^{(k)} \Phi_\theta(x_j; \theta_h^{(k)})}. \tag{B7}$$

With (B7) the operator (B5) becomes

$$Q(\theta'; \theta^{(k)}) = \sum_{i=1}^{M} \sum_{j=1}^{n} r_i(x_j; \theta^{(k)})[\log \alpha_i + \log \Phi_\theta(x_j; \theta_i)]. \tag{B8}$$

In the *M*-step in iteration k+1, $Q(\theta'; \theta^{(k)})$ is globally maximised with respect to θ' to update the estimate of the parameters:

$$\theta^{(k+1)} = \arg\max_{\theta'} Q(\theta'; \theta^{(k)}). \tag{B9}$$

For *finite* mixture models the updated estimates of the mixing proportions $\alpha_i^{(k+1)}$ and the component parameters θ_j^{k+1} can be deter-

mined independently. The maximum likelihood estimate of the mixing proportions takes the form

$$\alpha_i^{(k+1)} = \sum_{j=1}^{n} \frac{r_i(x_j; \theta^{(k)})}{n}, \tag{B10}$$

while the update for the component parameters can be deduced by solving

$$\sum_{i=1}^{M} \sum_{j=1}^{n} r_i(x_j; \theta^{(k)}) \partial \log \Phi_\theta(x_j; \theta_i) / \partial \Theta = 0, \tag{B11}$$

where $\Theta = (\theta_i, \ldots, \theta_M)$.

The algorithm stops once the difference $l(\theta^{(k+1)}; \boldsymbol{x}) - l(\theta^{(k)}; \boldsymbol{x})$ is smaller than some threshold provided that the sequence of likelihood values of the incomplete data $\{l(\theta^{(k)}; \boldsymbol{x})\}$ converges. The EM-algorithm works because the EM-iteration does not decrease the log-likelihood of the incomplete data, i.e.

$$l(\theta^{(k+1)}; \boldsymbol{x}) \geq l(\theta^{(k)}; \boldsymbol{x}). \tag{B12}$$

In Section 4.2 the EM algorithm was employed to fit a *mixture of two Gaussian normal distributions*. In this case the EM algorithm operates as follows (see e.g. McLachlan & Peel 2000; Hastie et al. 2009; Benaglia et al. 2009):

(i) Initially guess the parameters $\theta^{(0)}$.

(ii) *E-step*: Compute equation (B8) via (B7).

(iii) *M-step*: Update parameters according to (B9), i.e. calculate (B10) and (B13). In this case the solutions to equation (B11) have the closed forms:

$$\mu_i^{(k+1)} = \frac{\sum_{j=1}^{n} r_{ij}^{(k)} x_j}{\sum_{j=1}^{n} r_{ij}^{(k)}}$$

$$\sigma_i^{(k+1)} = \frac{\sum_{j=1}^{n} r_{ij}^{(k)} (x_j - \mu_i^{(k+1)})^2}{\sum_{j=1}^{n} r_{ij}^{(k)}}, \tag{B13}$$

where $r_{ij}^{(k)} = r_i(x_j; \theta^{(k)})$.

(iv) Repeat steps 2 and 3 until convergence is reached.

This paper has been typeset from a TeX/LaTeX file prepared by the author.

3.2 Stellar parameter determination from photometry using invertible neural networks

(Ksoll et al., 2020) presents the first application of a cINN (Ardizzone et al., 2018, 2019) to the astronomical inverse problem of predicting physical parameters of stars from photometric observations. In this project the cINN is trained and tested on synthetic data from the PARSEC stellar evolution models, and applied to observational HST data from the two clusters Wd2 and NGC 6397 (see Sections 2.1 and 2.2).

Contributions: To follow-up upon Ksoll et al. (2018), Dimitrios Gouliermis and Ralf Klessen suggested to extend the ML analysis of photometric data to solving the inverse problem of predicting stellar parameters from photometry. Ullrich Koethe proposed the cINN approach, developed in his group at the Visual Learning Lab, for this project. An initial implementation for the cINN was provided by Lynton Ardizzone. Elena Sabbi suggested Westerlund 2 and NGC 6397 as the real observational benchmarks for the approach and provided the HST data for Wd2. Peter Zeidler contributed his gas extinction map of Wd2 for the extinction analysis. I performed the remaining analysis for this project, constructing training sets, adapting the cINN implementation, compiling evaluation scripts, training and testing the cINN models, and predicting on the real data. I also composed all the text for the manuscript, bar the introduction, mostly provided by Dimitrios Gouliermis, and created all figures. All co-authors additionally provided feedback to the manuscript during the draft stage and assisted with proofreading.

Stellar parameter determination from photometry using invertible neural networks

Victor F. Ksoll,[1,2]★ Lynton Ardizzone,[3] Ralf Klessen,[1,2] Ullrich Koethe,[3] Elena Sabbi,[4]
Massimo Robberto,[4,5] Dimitrios Gouliermis [iD],[1,6] Carsten Rother,[3] Peter Zeidler [iD][4,5]
and Mario Gennaro [iD][4,5]

[1]*Zentrum für Astronomie, Institut für Theoretische Astrophysik, Universität Heidelberg, Albert-Ueberle-Str. 2, D-69120 Heidelberg, Germany*
[2]*Interdisziplinäres Zentrum für Wissenschaftliches Rechnen, Universität Heidelberg, Im Neuenheimer Feld 205, D-69120 Heidelberg, Germany*
[3]*Heidelberg Collaboratory for Image Processing, Visual Learning Lab, Universität Heidelberg, Berliner Str. 43, D-69120 Heidelberg, Germany*
[4]*Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA*
[5]*Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD 21218, USA*
[6]*Max Planck Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany*

ABSTRACT

Photometric surveys with the *Hubble Space Telescope* (*HST*) allow us to study stellar populations with high-resolution and deep coverage, with estimates of the physical parameters of the constituent stars being typically obtained by comparing the survey data with adequate stellar evolutionary models. This is a highly non-trivial task due to effects such as differential extinction, photometric errors, low filter coverage, or uncertainties in the stellar evolution calculations. These introduce degeneracies that are difficult to detect and break. To improve this situation, we introduce a novel deep learning approach, called conditional invertible neural network (cINN), to solve the inverse problem of predicting physical parameters from photometry on an individual star basis and to obtain the full posterior distributions. We build a carefully curated synthetic training data set derived from the PARSEC stellar evolution models to predict stellar age, initial/current mass, luminosity, effective temperature, and surface gravity. We perform tests on synthetic data from the MIST and Dartmouth models, and benchmark our approach on *HST* data of two well-studied stellar clusters, Westerlund 2 and NGC 6397. For the synthetic data, we find overall excellent performance, and note that age is the most difficult parameter to constrain. For the benchmark clusters, we retrieve reasonable results and confirm previous findings for Westerlund 2 on cluster age ($1.04^{+8.48}_{-0.90}$ Myr), mass segregation, and the stellar initial mass function. For NGC 6397, we recover plausible estimates for masses, luminosities, and temperatures, however, discrepancies between stellar evolution models and observations prevent an acceptable recovery of age for old stars.

Key words: methods: data analysis – methods: statistical – stars: formation – stars: fundamental parameters – stars: pre-main-sequence – galaxies: clusters: individual: Westerlund 2, NGC 6397 .

1 INTRODUCTION

Machine learning (ML) employs statistical models to predict the characteristics of a data set using samples of previously collected data without relying on physical models of the system. The introduction of ML for solving regression, classification, and clustering problems has revolutionized scientific research, and in particular has provided effective methods for analysing big astronomical data (Feigelson & Babu 2012; Ivezic et al. 2014). In order to construct a model from observed data, ML methods rely on human-defined classifiers or 'feature extractors' (Hastie, Tibshirani & Friedman 2009). However, complex problems require algorithms that automate the creation of feature extractors using large amounts of data. These algorithms represent a family of ML techniques, named deep learning, and they are based on the construction of artificial neural networks (NNs, Goodfellow, Bengio & Courville 2016). While training NNs requires

significant computational power, they achieve far higher levels of accuracy than classic ML for many non-linear problems. In this pilot study, we employ invertible NNs to infer stellar ages and masses from *Hubble Space Telescope* (*HST*) imaging of two well-studied stellar clusters. Our aim is to explore the efficiency of NNs in extracting stellar physical parameters from photometry alone. We train our networks using modelled-observable properties relations provided by theoretical evolutionary models.

Star clusters, the building blocks of galaxies, are the signposts guiding our understanding of the formation and evolution of stars. This understanding stems from the physical properties of stars in clusters, being deduced from detailed comparisons of photometric observations to theoretical evolutionary models. The interface where observations meet theory is often provided by the observational colour–magnitude diagram (CMD) and its theoretical counterpart, the Hertzsprung–Russell diagram (HRD). In the HRD two physical properties of stars, the effective temperature and the luminosity, are compared to stellar evolutionary models to determine fundamental stellar parameters, the initial mass and the age of the star, which are

★ E-mail: v.ksoll@stud.uni-heidelberg.de

not directly accessible by observations alone. This comparison can be directly performed through fitting of isochronal evolutionary models to the observed CMDs. This method, however, lacks proper statistical basis because the relations between observables and physical properties may present degeneracies that need to be accounted for. More advanced methods, based on Bayes statistics, derive probabilistically the cumulative properties of stellar populations, such as the mean age, in terms of posterior probability distribution functions of the properties of individual stars, for example, the age (see Valls-Gabaud 2014, and references therein). These methods provide a significant improvement by tackling the intrinsic model degeneracies through priors on the stellar initial mass function (IMF), binary fraction, or extinction distribution (e.g. Jørgensen & Lindegren 2005; Da Rio, Gouliermis & Gennaro 2010).

Bayesian inference encompasses a specific class of ML models, that is, those based on strong prior intuitions. However, these priors do not add significant value in the case of big data, and are computationally expensive and slow. As a consequence, other ML methods are employed to infer stellar physical parameters from photometry. The most successful techniques developed so far are generally based on time-domain observations, such as light curves using photometric-brightness variations (e.g. Miller et al. 2015) or time-series asteroseismic observations (e.g. Bellinger et al. 2016). These methods make use of various instances of each specific target star in time, a data set which cannot be easily obtained for rich stellar samples in compact clusters. Investigations of stars in clusters normally rely only on 'static', rather than time-dependent imaging, which cannot be addressed by classic ML methods. Moreover, it is now well understood that parameter degeneracies encoded in the evolutionary models make the problem of inferring stellar masses and ages from photometric measurements a non-linear problem. The solution of such problems calls for the employment of artificial NNs.

There have been several recent studies that employ NN approaches to solve prediction tasks in astronomy similar to the problem that we analyse in this paper. Sharma et al. (2020) train a convolutional NN on a suite of spectral libraries in order to classify stellar spectra according to the Harvard scheme and successfully apply their approach to data from the Sloan Digital Sky Survey (SDSS) data base. Kounkel, Covey & Stassun (2020) leverage *Gaia* DR2 photometry and parallaxes to construct an NN that predicts age, extinction, and distance of stellar clusters in the Milky Way, allowing them to study the star formation activity in the spiral arms. Cantat-Gaudin et al. (2020) use a similar NN approach, also predicting physical parameters of stellar clusters from *Gaia* data, but use 2D histograms of the observed CMDs as inputs. Olney et al. (2020) use a deep convolutional NN to predict surface temperature, metallicity, and surface gravity of young stellar objects (YSOs) based on spectra from the Apache Point Observatory Galactic Evolution Experiment (APOGEE). Within their training set construction they employ another convolutional NN to infer physical parameters of YSOs, that is, ages, masses, extinction, surface temperature/gravity, from photometry in nine bands of the *Gaia* system, as well as distance, stellar radius, and luminosity. This auxiliary network is trained on synthetic isochrone data and successfully recovers surface temperatures for YSOs on real *Gaia* observations.

For many applications in natural sciences, the forward process of determining measurements from a set of underlying physical parameters is well defined, whereas the inverse problem is ambiguous because multiple parameter sets can result in the same observation (i.e. degeneracies). Classical NNs attempt to address this ambiguity by solving the inverse problem directly. However, to fully characterize degeneracies, the full posterior parameter distribution,

conditioned on an observed measurement, must be determined. A particular class of NNs, so-called invertible neural networks (INNs), is well suited for this task (e.g. Ardizzone et al. 2019a). Unlike classical NNs, INNs learn the forward process, using additional latent output variables to capture the information otherwise lost. This invertibility allows a model of the corresponding inverse process to be learned implicitly, providing the full parameter posterior distribution for a given observation and corresponding distribution of the latent variables. INNs are therefore a powerful tool in identifying multi-modalities, parameter correlations, and unrecoverable parameters.

In this paper, we present the application of INNs to the regression problem of predicting physical parameters of individual stars based on observed photometry. Note that we do not perform an exhaustive analysis of the approach, but rather aim to provide an introduction to the method, highlighting our first successes. This paper is the first in a series, in which we adapt and develop the approach, as well as explore its limitations.

As mentioned above, in general this regression task is prone to errors due to the many sources of degeneracy in the mapping from physical to observable space, such as metallicity, extinction, variability, binarity, and the intrinsic overlap of certain phases in stellar evolution in the observable space, for example, the red giant branch (RGB) and the pre-main sequence. Since our primary goal is to test the viability of the method, in this paper we neglect some of these factors, adopting the following simplifying assumptions: (1) we only deal with single metallicity populations, (2) we obtain an estimate of the individual stellar extinction of the query stars, (3) we assume perfect observations, so we do not include photometric errors, and (4) we exclude effects from variability or binarity.

We train and test our method on synthetic data from the PARSEC stellar evolutionary models (Bressan et al. 2012). Furthermore, we conduct additional synthetic tests on data from the MIST (Dotter 2016) and Dartmouth (Dotter et al. 2008) models. Lastly, we perform a benchmark study on real observational data from the *HST* of the young star forming cluster Westerlund 2 and the old globular cluster NGC 6397. These clusters are chosen for our pilot study due to their well-defined single ages (Zeidler et al. 2016; Brown et al. 2018), allowing for an accurate evaluation of our results.

In Section 2, we summarize the physical properties of our benchmark targets and the reduction of the observational data from their respective surveys. Furthermore, we outline the construction of our training sets from the synthetic data provided by the PARSEC models. In the following Section 3, we elaborate the background of the INN approach and provide details of the final architecture of our models as well as the performance measures used to evaluate their success. Section 4 summarizes the performance of the cINN on the PARSEC synthetic test data for each of our four training sets and details the results of the application to the MIST and Dartmouth data. In Section 5, we present the prediction outcome on the real observational data for both Westerlund 2 and NGC 6397. We discuss possible future extensions of our approach beyond the simplifications assumed for this work in Section 6. The final Section 7 summarizes our key findings.

2 DATA SELECTION AND PREPARATION

2.1 Observational data

To test our neural-network-based approach to predicting physical parameters of stars on real observational data, we use two 'well behaved', supposedly single age (or close to) stellar clusters for which very high spatial resolution *HST* observations are available, namely
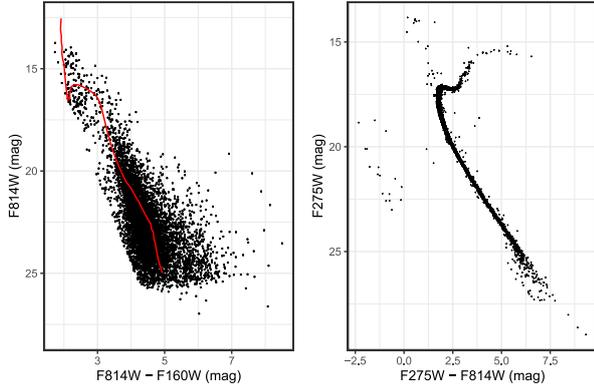
Figure 1. Left: CMD of our Wd2 data set. The red line represents a 1 Myr PARSEC (Bressan et al. 2012) isochrone corrected for the median stellar extinction and distance modulus of Wd2. Right: UV-I CMD of the NGC 6397 photometric catalogue from the HUGS project.

the young massive star-forming region Westerlund 2 (hereafter referred to as Wd2) located within the Milky Way and the old globular cluster NGC 6397 belonging to the galactic halo. Since this paper serves only as an introduction to the INN approach to gain initial insights into the systematics of the method we do not conduct an exhaustive study of the full range of the cluster mass, age, and metallicity distribution, but we consider only the two extremes in age (i.e. very young and very old).

2.1.1 Westerlund 2

Wd2 is one of the most massive star-forming clusters in the Milky Way, harbouring a total stellar mass larger than $10^4 \, \mathrm{M_\odot}$ (Ascenso et al. 2007). It is located in the Carina-Sagittarius arm at a distance of 4.16 ± 0.33 kpc (Zeidler et al. 2015) from the Sun. At an age of 1.04 ± 0.72 Myr (Zeidler et al. 2016), Wd2 makes for an excellent example of a young massive cluster at solar metallicity still in its early star formation stages within close proximity to the Sun. While Wd2 exhibits an average total-to-selective extinction of $R_V = 3.95 \pm 0.135$ that is larger than the galactic average $R_V = 3.1$, the cluster is only affected by relatively low differential reddening with $E(B-V)_g = 1.87$ mag (median colour excess of the gas, Zeidler et al. 2015). For our following considerations, we adopt $R_V = 3.8$ to be both in agreement with the findings of Zeidler et al. (2015) as well as the spectroscopic observations of Carraro et al. (2013) and Vargas Álvarez et al. (2013) who suggest $R_V = 3.85 \pm 0.07$ and 3.77 ± 0.09, respectively. Thus, the corresponding median gas extinction of Wd2 lies at $A_{V,g} = 7.1$ mag.

Combining multi-epoch *HST* images taken with the Wide Field Camera 3 (WFC3) in F814W with previously obtained UVIS-IR data in F160W (PI: Nota, GO-13038) Sabbi et al. (2020) compile the photometric catalogue that we employ for this study. Due to the long 350 s exposure times in F814W, this photometric catalogue does unfortunately not contain the brightest objects of Wd2, that is, the most massive upper-main-sequence (UMS) constituents, as they were saturated. Disregarding these missing UMS sources, the Sabbi et al. (2020) photometric catalogue consists of 9267 stars, of which 6268 are thought to belong to Wd2. The remaining stars in the sample can be tentatively classified as lower-main-sequence (LMS) fore- or background contaminants that fall into the line of sight. The left-hand panel in Fig. 1 shows the CMD of the 6268 cluster stars.

Adopting the Zeidler et al. (2015) gas extinction map of Wd2, we can derive individual stellar colour excesses $E(B-V)_*$ for 8939 stars that fall within the border of the map following their prescription:

$$E(B-V)_* = 0.4314 \cdot E(B-V)_g + 0.7400. \tag{1}$$

The individual stellar extinctions then follow as

$$A_V = R_V \cdot E(B-V)_*. \tag{2}$$

2.1.2 NGC 6397

NGC 6397 is the nearest metal poor globular cluster, with a distance of $d = 2.39 \pm 0.17$ kpc (distance modulus, DM $= 11.89 \pm 0.16$ mag) derived from parallax measurements with high precision *HST* astrometry (Brown et al. 2018). Spectroscopic measurements indicate a metallicity of [Fe/H] $= -2.02$ (Kraft & Ivans 2003; Vulic, Barmby & Gallagher 2018), making it a prime example of an ancient metal-poor stellar population. Fitting of the main-sequence turn-off suggests a cluster age of 13.4 ± 1.9 Gyr (Brown et al. 2018). Several extinction studies indicate a moderate reddening, constraining $E(B-V)$ to a value between 0.183 mag (Gratton et al. 2003), 0.186 mag (Schlegel, Finkbeiner & Davis 1998), and 0.187mag (Anthony-Twarog, Twarog & Suntzeff 1992). In this work we adopt $E(B-V) = 0.185 \pm 0.002$ mag from (Brown et al. 2018), corresponding to an average extinction of $A_V = E(B-V) \cdot R_V = 0.5735 \pm 0.0062$ mag with $R_V = 3.1$. To derive individual stellar extinctions here, we simply sample from a Gaussian distribution with this mean and standard deviation.

We use the photometric catalogue of NGC 6397 from the *HST* legacy survey '*HST* UV Globular Cluster Survey (HUGS)' (Piotto et al. 2015; Nardiello et al. 2018), which provides coverage in the F275W, F336W, and F438W filters, observed with the WFC3/UVIS channel, as well as in F606W and F814W, imaged with the Advanced Camera for Surveys (ACS/WFC) (Nardiello et al. 2018). To pre-process this data, we follow the prescription in Section 3 of Nardiello et al. (2018). We divide the photometric error and quality of fit distributions of each filter into 12 mag bins and find the 3.5σ clipped average of the magnitude and parameter in each bin. Here, σ refers to the standard deviation of the distribution in the given bin. In each bin, 3.5σ is then added to the mean value and a linear interpolation is performed between these points. For the photometric errors, we then reject all observations that lie above this interpolated line while for the quality of fit parameter we reject all instances below the line. Finally, we limit the catalogue to observations with a sharpness value between -0.15 and 0.15 in all five filters. Following these selection criteria we obtain a photometric catalogue containing 4831 stars. The right-hand panel of Fig. 1 shows the corresponding UV-I CMD.

2.2 Synthetic training data

In order to train the NN for the purpose of predicting physical parameters given photometric observations of individual stars, a large training set is required that contains both the physical parameters and the corresponding photometric observations of each star. Since at present such a training data set is not readily available, we build it from theoretical stellar evolutionary models. In particular, we use version 1.2s of the PARSEC stellar evolutionary tracks (Bressan et al. 2012; Chen et al. 2014, 2015; Tang et al. 2014) and more specifically the isochrone tables derived for the *HST* photometric systems 'WFC3 wide' and 'ACS WFC'. Since our observational test cases Wd2 and NGC 6397 differ both in metallicity and *HST* filter coverage, we have to construct individual training sets for each cluster. This is

consistent with the fact that our NN structure can only deal with single metallicity cases. An artificial training set is also appropriate since our NN cannot deal with missing observational features.

At this point, it is important to note that using synthetic training data comes with caveats. In particular it is known that the photometry interpolated from the stellar evolution models may show minor discrepancies in colors as the approximation to the real bandpasses may be imprecise. Consequently, the synthetic photometry can never perfectly match real observations. Additionally, the models themselves may be discrepant, for example, for very low-mass stars (Jackson, Deliyannis & Jeffries 2018) or YSOs (Olney et al. 2020), or may exhibit physically questionable properties such as the large gap in surface temperature for pre-main-sequence stars at 4000 K in the PARSEC models. Nevertheless, for example, Olney et al. (2020) find that an NN approach, trained on synthetic data, can recover realistic physical properties for YSOs on real data, where traditional isochrone fitting approaches fail due to the model discrepancies. In any case, given the task we aim to solve here, the use of synthetic training data is simply unavoidable. Therefore, we proceed keeping these caveats in mind for our real data benchmarks.

For both clusters, we construct two training sets, one agnostic to prior knowledge of the stellar ages and one where we constrain the stellar ages to a range close to the supposed cluster ages derived in previous studies.

The first two training sets 'Wd2_I' and 'NGC 6397_I' thus consider isochrones with $\log(\text{age/yr})$ in the range 5–10.1 in steps of 0.05 dex. The 'NGC 6397_I' set also specifically entails the $\log(\text{age/yr}) = 10.13$ isochrone to include the supposed age of NGC 6397 of 13.4 Gyr (Brown et al. 2018). For the other two training sets 'Wd2_II' and 'NGC 6397_II', we restrict the isochrones to $\log(\text{age/yr})$ ranges of 5–8 in 0.025 dex, and 9.0–10.13 in 0.01 dex, respectively. Fig. 2 shows the HRD corresponding to these training sets. We do not impose a restriction on the range of initial stellar mass M_{ini} so that the full mass range of the PARSEC models (0.09–$350\,\text{M}_\odot$) is available in all but the 'NGC 6397_II' training set, where the range has been reduced to 0.09–$1.837\,\text{M}_\odot$ due to the fact that the more massive stars have already died at these ages. The other physical parameters that we consider for prediction are current mass M_{curr}, luminosity L, effective temperature T_{eff}, and surface gravity g. Again, we do not limit these parameters so that the respective ranges depend on the isochrones included in each training set.

For these training sets, we do not perform population synthesis based on the isochrone tables, but instead we consider each point of the isochrones as an individual example star, aiming at performing parameter prediction on a star-by-star basis. To this purpose, we need to populate the physical parameter space in the training set as evenly as possible, since overpopulated regions could introduce biases in the training process, so that our trained model might in the end generalize poorly when predicting parameters for a star that falls in a less-populated area in parameter space. We face this problem with the isochrone models. While the PARSEC models provide perfectly evenly spaced isochrones in $\log(\text{age})$, Bressan et al. (2012) perform an interpolation when generating the isochrones from the stellar evolutionary tracks that aims to produce smooth isochrone curves, resulting in a severe oversampling of certain masses due to the fact that very small mass variations can cause a significant change of position in the HRD on the post-main-sequence part of the isochrones. Fig. 3 shows an example of this mass oversampling for the 'Wd2_I' case. The left-hand diagram highlights how the interpolation strategy of the PARSEC isochrone tables results in a severe oversampling of masses along the ridge where the models stop, because stars of a given mass die away.
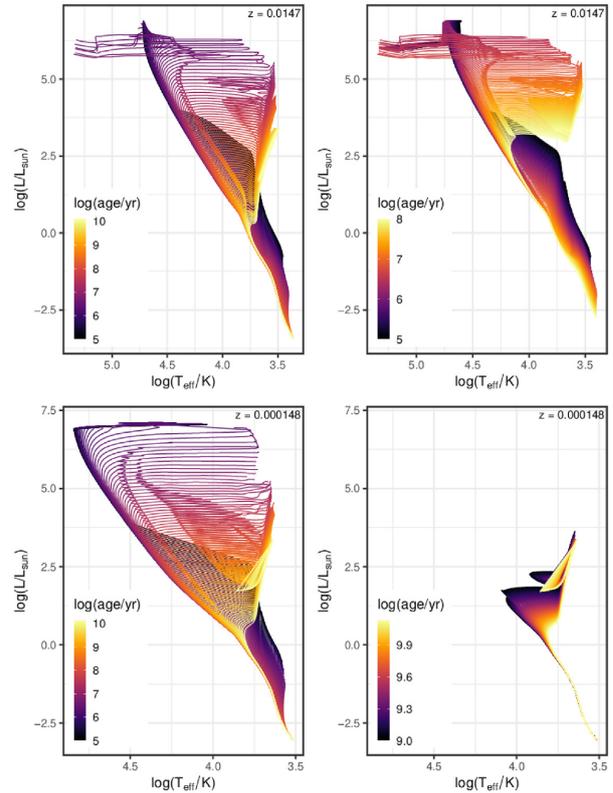
Figure 2. HRDs for the PARSEC 1.2 s isochrone tables used as basis for our training sets. The top row shows set 'Wd2_I' with isochrones from 5 to 10.1 in $\log(\text{age/yr})$ in steps of 0.05 dex (left) and set 'Wd2_II' with isochrones in the range of 5–8 in 0.025 dex (right). In the bottom row are the corresponding HRDs of the sets 'NGC 6397_I'(left), containing isochrones from 5 to 10.13 in 0.05 dex, and 'NGC 6397_II' with isochrones from 9 to 10.13 in steps of 0.01 dex. All isochrones are colour coded according to their $\log(\text{age})$.

Consequently, there are several regions, for example, the old low-mass and young supermassive stars, where the age–mass space is strongly underpopulated.

To remedy this problem, we have devised a procedure to augment the isochrone tables so that the density differences between the over- and underpopulated regions in age–mass space are reduced. We begin by oversampling each isochrone in M_{ini} space, first performing a linear spline interpolation in the M_{ini}–L–T_{eff} space to determine its arc length, that is, the length of the path along the isochrone from the lowest mass model point to the most massive one. Then we find 10 000 equidistant (in terms of the logarithm of the arc length) M_{ini} points along each isochrone. For these points, we determine the remaining parameters (L, T_{eff}, g, and magnitudes) by performing a linear interpolation between the nearest lower and nearest higher initial mass neighbour on the original isochrone.

The resulting age–mass distribution of these oversampled isochrones is shown in the middle diagram of Fig. 3. The plot indicates that this procedure does not solve the issue of oversampled mass bins directly, in fact, it further highlights those regions. But at the same time it manages to populate previously sparsely sampled regions. To finally produce an evenly sampled training set we then augment the original isochrone tables by adding random samples from our oversampled isochrones until every age–mass bin contains at least 30 example stars (this value is chosen to roughly represent
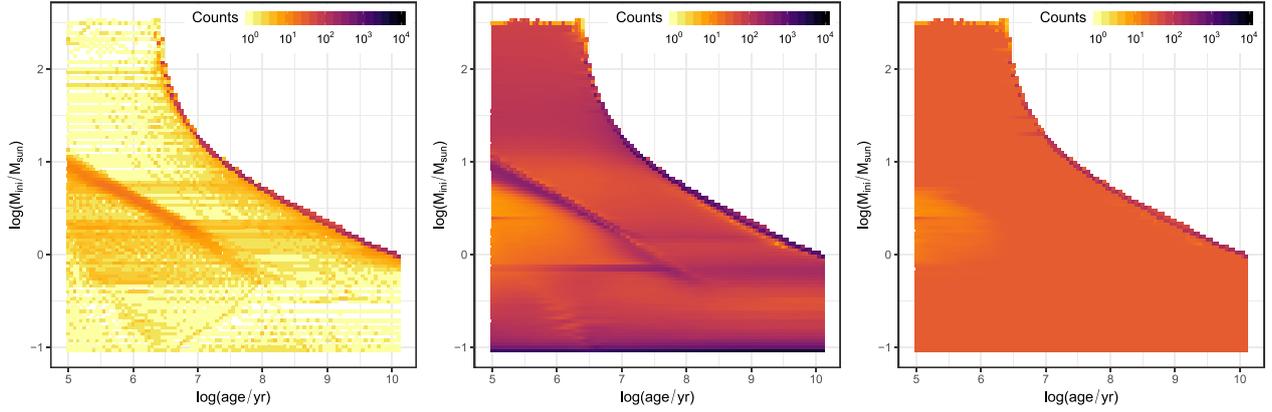
Figure 3. 2D histograms of the age versus initial mass distributions for the original PARSEC isochrone tables (left), the case where each isochrone is oversampled 10 000 times in regular intervals in terms of the logarithm of its arc length in M_{ini}–L–T_{eff} space (middle) and our final training set base (right). In the latter, each age–mass bin that contains less than 30 model points in the original table is reinforced with additional samples from the oversampled isochrones to reach 30 examples. This example is from the 'Wd2_I' training set.

the number of the least populated bins in the oversampled data set). If the oversampled table does not contain enough additional examples to augmented the original isochrones to 30 examples in a given bin, we simply include all available additional examples. We also ensure to only augment with examples that do not appear in the original tables. The resulting distribution in age–mass space is depicted in the right-hand panel of Fig. 3, showing that this approach achieves a mostly even sampling across the whole parameter range.

There are two reasons why we do not achieve a perfectly even sampling. First, subsampling the overpopulated bins would result in a significant information loss in the HRD and CMD as several post-main-sequence evolutionary tracks fall into these bins. Second, oversampling the isochrones and then augmenting the original tables to a degree that all bins reach the level of the originally most populated bin would result in a data set so large that it becomes not manageable for our remaining processing.

The last step in our training set construction procedure is to augment the data taking extinction into account. We do so for each star in the training set by including additional copies of it at different amounts of extinction A_V and altering their observable features, that is, magnitudes in *HST* filters, accordingly. For Wd2, we consider an extinction range from 0 to 12 in steps of 0.2 mag and for NGC 6397 from 0 to 3 in steps of 0.05 mag in accordance with the Wd2 gas extinction map from Zeidler et al. (2015) and the suggested average extinction of NGC 6397 by Brown et al. (2018). For the extinction law, we use the diffuse Milky Way extinction curve by Cardelli, Clayton & Mathis (1989), deriving the A_λ/A_V values in dependence of R_V for the *HST* filters according to

$$\frac{A_\lambda}{A_V} = a_\lambda + \frac{b_\lambda}{R_V}, \tag{3}$$

where a_λ and b_λ denote wavelength dependent coefficients defined by Cardelli et al. (1989). Table A1 in Appendix A provides the derived A_λ/A_V values for all filters.

In conclusion, each training set contains the six physical parameters: age, initial mass M_{ini}, current mass M_{curr}, luminosity L, effective temperature T_{eff}, surface gravity g, extinction A_V, and magnitudes in filter combinations corresponding to our real observations. These are F814W$_{WFC3}$ and F160W$_{WFC3}$ for Wd2, and F275W$_{WFC3}$, F336W$_{WFC3}$, F438W$_{WFC3}$, F606W$_{ACS}$, and F814W$_{ACS}$
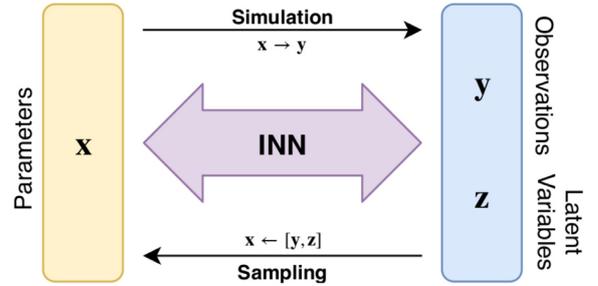
Figure 4. Schematic overview of the INN approach for solving an inverse problem. Adapted from Ardizzone et al. (2019a).

for NGC 6397. In total our training sets contain 124 81 881, 209 03 602, 123 56 282, and 168 17 090 example stars for 'Wd2_I', 'Wd2_II', 'NGC 6397_I' and 'NGC 6397_II', respectively. Fig. A1 in Appendix A shows the corresponding prior distributions of all physical parameters for these training sets.

3 NEURAL NETWORK SETUP

3.1 INN and cINN

In this paper, we solve the inverse problem of predicting physical parameters of stars from *HST* photometry employing an INN as described in Ardizzone et al. (2019a, b). This INN approach provides an inverse solver that estimates the *complete posterior* distribution of physical parameters *conditioned* on an observation. Fig. 4 outlines the concept of the INN methodology. Given a well-understood simulation that maps physical parameters **x** to observations **y**, we assume that this forward process entails an inherent information loss, such that **y** does *not* explain all variance of **x** and *degeneracies* occur in the mapping. To retain this information that would be otherwise lost additional latent variables **z** are introduced to encode all the variance of **x** that is not captured in **y**.

A benefit of a network with an invertible architecture with regard to our current regression problem is that once it has been trained to approximate the known forward process *f*, it provides a solution for the inverse process f^{-1} for free. In the application outlined here, the
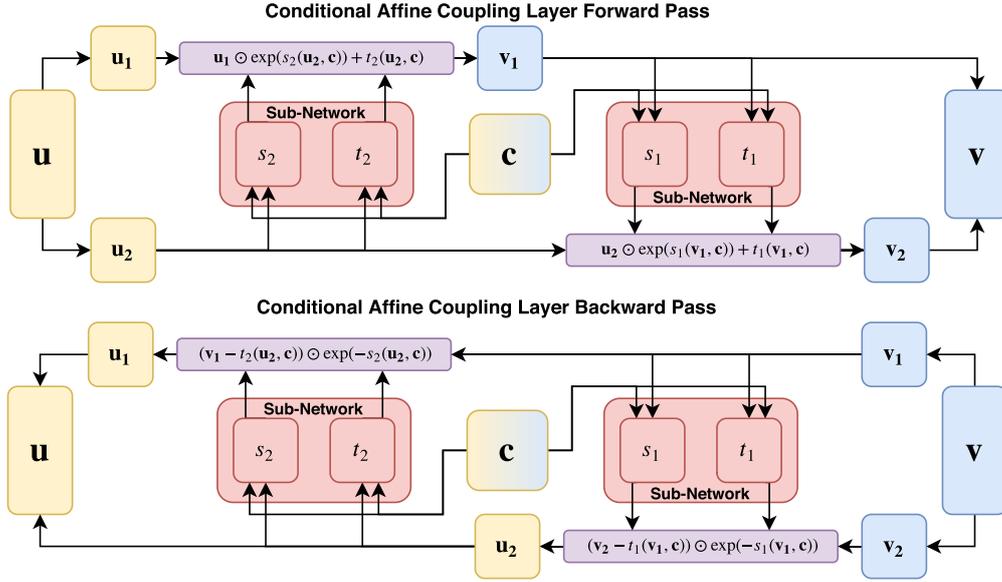
Figure 5. Schematic overview of the architecture of the conditional affine coupling blocks used in the cINN. In particular, we show the GLOW (Kingma & Dhariwal 2018) configuration, where the outputs $s_i()$ and $t_i()$ are computed by a single subnetwork (for each i). The top panel shows how data are passed through the block in the forward direction (from $\boldsymbol{x}$ to $\boldsymbol{z}$), while the bottom panel displays the inverted case following the affine transformations in equations (4) and (5).

INN will thus learn how to associate physical parameter values $\boldsymbol{x}$ to unique pairs $[\boldsymbol{y}, \boldsymbol{z}]$ of observations and latent variables, as it trains to optimize the forward mapping $f(\boldsymbol{x}) = [\boldsymbol{y}, \boldsymbol{z}]$ and then implicitly finds the inverse $\boldsymbol{x} = f^{-1}(\boldsymbol{y}, \boldsymbol{z}) = g(\boldsymbol{y}, \boldsymbol{z})$ (Ardizzone et al. 2019a). For simplicity, the prior distribution of the latent variables $p(\boldsymbol{z})$ is assumed (and enforced during training) to be Gaussian. The desired posterior distribution $p(\boldsymbol{x}|\boldsymbol{y})$ is represented by the function $g(\boldsymbol{y}, \boldsymbol{z}) = \boldsymbol{x}$, which, given the condition $\boldsymbol{y}$, transforms the known distribution $p(\boldsymbol{z})$ to $\boldsymbol{x}$-space (Ardizzone et al. 2019a). In practice, this means that for a given observation $\boldsymbol{y}$ the posterior distribution $p(\boldsymbol{x}|\boldsymbol{y})$ is determined by sampling the latent variables.

In Ardizzone et al. (2019a), the invertibility of the network is achieved by a series of reversible blocks based on the architecture proposed by Dinh, Sohl-Dickstein & Bengio (2016). These blocks split their input vector $\boldsymbol{u}$ into two halves $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ and then apply two complementary affine transformations with element-wise multiplication $\odot$ and addition $+$,

$$\boldsymbol{v}_1 = \boldsymbol{u}_1 \odot \exp(s_2(\boldsymbol{u}_2)) + t_2(\boldsymbol{u}_2),$$

$$\boldsymbol{v}_2 = \boldsymbol{u}_2 \odot \exp(s_1(\boldsymbol{v}_1)) + t_1(\boldsymbol{v}_1), \qquad (4)$$

where s_i and t_i are mappings that can be arbitrarily complex functions of $\boldsymbol{u}_2$ and $\boldsymbol{v}_1$ that do not need to be invertible themselves and can even be represented by NNs. These affine transformations are easily inverted given the output $\boldsymbol{v} = [\boldsymbol{v}_1, \boldsymbol{v}_2]$,

$$\boldsymbol{u}_2 = (\boldsymbol{v}_2 - t_1(\boldsymbol{v}_1)) \odot \exp(-s_1(\boldsymbol{v}_1)),$$

$$\boldsymbol{u}_1 = (\boldsymbol{v}_1 - t_2(\boldsymbol{u}_2)) \odot \exp(-s_2(\boldsymbol{u}_2)). \qquad (5)$$

Based on the Ardizzone et al. (2019a) method, Ardizzone et al. (2019b) present an extension to their original INN approach, the cINN. Here, they adapt the affine coupling block architecture to accept additional conditioning inputs $\boldsymbol{c}$. Since the mappings s_i and t_i, also when represented by NNs, are only evaluated in the forward direction, even when inverting the network, it is possible to concatenate these conditioning inputs with the regular inputs of

the subnetworks without compromising the INNs invertibility, for example, by replacing $s_2(\boldsymbol{u}_2)$ with $s_2(\boldsymbol{u}_2, \boldsymbol{c})$, etc., in equations (4) and (5). Fig. 5 shows an illustration for the forward (top) and backward (bottom) pass of this conditional affine coupling layer design in the GLOW (Generative Flow; proposed by Kingma & Dhariwal 2018) configuration (see Section 3.2 for details). In this setting the forward mapping is modified to $f(\boldsymbol{x}; \boldsymbol{c}) = \boldsymbol{z}$ and the inverse to $\boldsymbol{x} = g(\boldsymbol{z}; \boldsymbol{c})$. The invertibility is given for *fixed* condition $\boldsymbol{c}$ as

$$f(\,\cdot\,; \boldsymbol{c})^{-1} = g(\,\cdot\,; \boldsymbol{c}). \qquad (6)$$

In our regression problem, the conditioning is given by the observations. Therefore, as for the standard INN, during training given an observation the network will learn to encode all information about the physical parameters in the latent variables that was not contained in the observation. Also analogous to the standard INN, we retrieve the desired posterior distribution $p(\boldsymbol{x}|\boldsymbol{y})$ for a given observation $\boldsymbol{y}$ by sampling the latent variables according to their Gaussian priors and using the inverted network g:

$$\boldsymbol{x}_{\text{posterior}} = g(\boldsymbol{z}; \boldsymbol{c} = \boldsymbol{y}), \ \text{ with } z \sim p_Z(z) = \mathcal{N}(z, 0, \mathsf{I}), \qquad (7)$$

where I is the $K \times K$ unity matrix with $K = \dim(\boldsymbol{z})$.

One of the cINN benefits over the standard INN architecture is that no zero padding (as described in Ardizzone et al. 2019a) is necessary if the dimension of $[\boldsymbol{y}, \boldsymbol{z}]$ were to exceed that of $\boldsymbol{x}$, as the conditioning input $\boldsymbol{c}$ can be arbitrarily large in this approach and the dimension of $\boldsymbol{z}$ simply matches that of $\boldsymbol{x}$.

3.2 Architecture details

To implement the cINN for our purposes, we use the 'Framework for Easily Invertible Architectures' (FrEIA) for PYTHON (Ardizzone et al. 2019a, b) based on the 'pytorch' library (Paszke et al. 2019).

In our problem the input $\boldsymbol{x}$ is given by the six physical parameters of the isochrone tables, so that, following the cINN architecture, we also have six latent variables $\boldsymbol{z}$. Our cINN is conditioned on
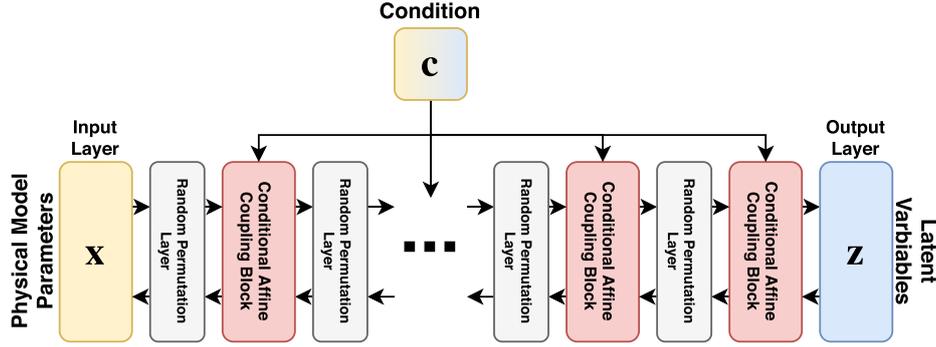
Figure 6. Schematic representation of the cINN architecture used for the physical parameter prediction based on photometry. In total, we use 16 conditional affine coupling blocks interchanged with random permutation layers.

the observables, 2 and 5 mag for Wd2 and NGC 6397, respectively, and the individual stellar extinctions, so that the condition c has the dimension 3 in the Wd2 cases and 6 for NGC 6397. Ardizzone et al. (2019b) also introduce a 'conditioning' network which transforms the input condition into some intermediate representation and is trained jointly with the cINN. We do not use this additional network in our setup, as we find that given the few observables in our problem the cINN tends to overfit to the synthetic training data when employing a feature extraction network, resulting in poor performance on the real benchmark data.

Our cINN consists of 16 conditional affine coupling blocks, each in the GLOW configuration (Kingma & Dhariwal 2018), which reduces computational cost and speeds up learning by jointly predicting the subnetwork outputs $s_i()$ and $t_i()$ using a single subnetwork. As in Ardizzone et al. (2019b), we introduce an additional nonlinear transformation of the scale coefficients s,

$$s_{\text{clamp}} = \frac{2\alpha}{\pi} \arctan\left(\frac{s}{\alpha}\right), \quad (8)$$

where $\alpha = 1.9$, so that $s_{\text{clamp}} \approx s$ for $|s| \ll \alpha$ and $s_{\text{clamp}} \approx \pm\alpha$ for $|s| \gg \alpha$, in order to avoid instabilities induced by large magnitudes of the exponential $\exp(s_{\text{clamp}})$.

We alternate the conditional affine coupling blocks with random permutation layers. The latter consist of random orthogonal matrices which mix the information between the two streams u_1 and u_2 in the coupling blocks. Following Ardizzone et al. (2019b), these matrices are fixed during training and cheaply invertible. The combination of these permutation layers with the interlocked affine transformations of the affine coupling blocks ensures that the network cannot ignore the conditioning input when learning the forward mapping. The subnetworks in the conditional affine coupling layers are simple fully connected feed-forward networks with three hidden layers of width 512 with rectified linear units (ReLU) as activation functions. Fig. 6 provides a schematic overview of our setup for the cINN.

We train the cINN models as described in Ardizzone et al. (2019b) by minimization of the *maximum likelihood loss*

$$\mathcal{L} = \mathbb{E}_i \left[\frac{||f(x_i; c_i, \theta)||_2^2}{2} - \log |J_i| \right], \quad (9)$$

where x_i is a training example with its corresponding condition c_i and J_i denotes the determinant of the Jacobi matrix $J_i = \det\left(\frac{\partial f}{\partial x}\big|_{x_i}\right)$ evaluated at x_i.

For each training set, the cINN is trained until the loss curve converges, but at least long enough that the model has seen each training example multiple times.

3.3 Data pre-processing

In preparation for training, the cINN we split our training data into physical parameters x (age, M_{ini}, M_{curr}, L, T_{eff}, g) and observables y (magnitudes $+ A_V$). To avoid issues in the training process that can occur due to their broad range of values, the physical parameters are transformed to logarithmic space. This serves not only to even out magnitude differences, but it has the general benefit of implicitly enforcing that these quantities can only be positive. Since all our observables are photometric magnitudes and thus already a logarithmic quantity, this step is not necessary there. On top of that we add a small amount of Gaussian noise (standard deviation of 1×10^{-5}) to the strongly discretized log(age) parameter. This form of data augmentation through a small amount of noise serves to smooth out discretization artefacts of the input (Ardizzone et al. 2019b). The remaining parameters are sampled unevenly enough that augmentation with noise is unnecessary.

After that we re-scale each parameter so that their resulting distribution has zero mean and unit standard deviation, following the *linear* transformation

$$\hat{x}_i = (x_i - \mu_{x_i}) \cdot \frac{1}{\sigma_{x_i}}, \quad (10)$$

where μ_{x_i} and σ_{x_i} are the mean and standard deviation of the distribution of the physical parameter x_i. At prediction time these linear re-scaling operations are easily inverted in order to retrieve the correct predicted physical parameters $x_{i,\text{pred}}$ from the predicted $\hat{x}_{i,\text{pred}}$ as

$$x_{i,\text{pred}} = \hat{x}_{i,\text{pred}} \cdot \sigma_{x_i} + \mu_{x_i}. \quad (11)$$

For the observables, after first centring the data ($\tilde{y}_i = y_i - \mu_{y_i}$), we perform a matrix whitening procedure (Hyvärinen & Oja 2000) on the $N \times M$ matrix $\tilde{Y}$, where N is the total number of examples in the training set and M the number of observables. The resulting *linearly* transformed matrix $\hat{Y}$ has the properties that all its columns $\hat{y}_i$ have unit variance and that its covariance matrix $\Sigma_{\hat{Y}}$ is equal to the unity matrix. $\hat{Y}$ is calculated as follows:

$$\hat{Y} = W_{\tilde{Y}} \tilde{Y} = ED^{-\frac{1}{2}} E^T \tilde{Y}, \quad (12)$$

where E is the orthogonal matrix of eigenvectors of the covariance matrix $\Sigma_{\tilde{Y}}$ of $\tilde{Y}$ and $D^{-\frac{1}{2}} = \text{diag}(d_1^{-\frac{1}{2}}, ..., d_m^{-\frac{1}{2}})$ with d_i being the ith eigenvalue of $\Sigma_{\tilde{Y}}$. In practice, we add a fudge factor $\epsilon = 1 \times 10^{-7}$ in the calculation of $D^{-\frac{1}{2}}$ to avoid overamplification of eigenvectors associated with small eigenvalues

$$D^{-\frac{1}{2}} = \text{diag}\left(\frac{1}{\sqrt{d_1 + \epsilon}}, ..., \frac{1}{\sqrt{d_m + \epsilon}}\right). \quad (13)$$

The scaling parameters μ_{x_i}, σ_{x_i}, μ_{y_i}, and $\mathbf{W}_{\tilde{Y}}$ are calculated from our entire synthetic data set, before we perform the split in training and test set. At prediction time of the real data from Wd2 and NGC 6397, the observational data are scaled using the *same* scaling parameters derived from the synthetic data the respective models were trained on (e.g. if we train the cINN on the synthetic data set 'Wd2_I', the real observations are scaled using the scaling parameters derived from that data set).

3.4 Evaluating training success

After training our models until the maximum likelihood loss converges, we evaluate the performance of these trained models on a held-out subset of the training data. In all our cases, these randomly chosen test subsets contain 20 000 observations. On a given test set we begin verifying if the cINN has converged to a good solution by confirming that the predicted distribution $\mathbf{Z}_{\text{test}}$ of the latent variables actually follows the multivariate normal distribution we prescribed as the target. This is easily checked by calculating the covariance matrix $\Sigma_{\mathbf{Z}_{\text{test}}}$ of $\mathbf{Z}_{\text{test}}$ and determining if its close enough to the unity matrix, as well as checking that all columns follow a normal distribution with zero mean.

To ascertain the quality of the predicted posterior distributions for each of the physical parameters, we compute the median *calibration error* $e_{\text{cal}}^{\text{median}}$. For a given confidence interval q, the calibration error over a set of N observations is defined as

$$e_{\text{cal}} = q_{\text{inliers}} - q, \tag{14}$$

where $q_{\text{inliers}} = \frac{N_{\text{inliers}}}{N}$ indicates the fraction of observations for which the true value falls within the q-confidence interval of the corresponding predicted posterior distribution. Negative values of e_{cal} indicate that the model is overconfident, predicting too narrow posterior distributions, while a positive e_{cal} describes an underconfident model that predicts posteriors that are too broad (Ardizzone et al. 2019a). We calculate $e_{\text{cal}}^{\text{median}}$ as the median of the absolute values of the calibration errors over a range of confidence intervals from 0.01 to 0.99 in steps of 0.01.

Apart from the calibration error, we also measure the cINN model accuracy for point estimations $\hat{x}$, that is, maximum a posteriori (MAP) estimates, of each physical parameter by computing the root mean square error (rmse) with respect to the ground truth x^* over the entire test set

$$\text{rmse} = \sqrt{\frac{\sum_{i=1}^{N} \left(\hat{x}_i - x_i^* \right)^2}{N}}. \tag{15}$$

In order to better compare the rmses of the four different models we train, we also compute a normalized rmse (nrmse). We derive this quantity for each physical parameter x by dividing the rmse by the range $\bar{x} = x_{\text{max}}^{\text{ts}} - x_{\text{min}}^{\text{ts}}$ covered in the training set, that is,

$$\text{nrmse} = \frac{\text{rmse}}{\bar{x}}. \tag{16}$$

To derive these performance measures for all of the 20 000 observations in the test sets for each posterior we sample 4096 times from the latent space $\mathbf{Z}$.

3.5 Determining MAP estimates

In order to assess the point estimate accuracy (Section 3.4) on our test set, as well as on the predicted physical parameters for the real observations presented in Section 4, we compute MAP estimates. To this purpose, given a posterior distribution for a physical parameter,

we first perform a kernel density estimation on the posterior using a Gaussian kernel function and then we find the parameter value at which this density estimate has a maximum. In practice, we evaluate the density on a regularly spaced grid of 1024 points ranging from the minimum to the maximum of the given posterior. To derive a suitable bandwidth h for this kernel density estimation, we use Silverman's rule of thumb,

$$h = 1.06 \cdot \min \left(\sigma, \frac{\text{IQR}}{1.34} \right) \cdot n^{-\frac{1}{5}}, \tag{17}$$

where IQR denotes the interquartile range, σ the standard deviation of the data, and n the number of data points (Silverman 1986). We choose this bandwidth estimator for its computational efficiency in order to quickly derive MAP estimates for our test observations, keeping in mind that this estimator is prone to suggest suboptimal bandwidths for density distributions that differ strongly from unimodal Gaussians.

3.6 Resimulation error

To verify whether the predicted posterior distributions are correct and not just cINN artefacts one usually performs a resimulation. Here, either the MAP estimates of the physical parameters or individual samples of the predicted posteriors are put into the simulation, that maps the physical to the observable space, to derive the associated resimulated observables $\mathbf{y}_{i,\text{resim}}$. They are then compared with the cINN input condition $\mathbf{y}_{i,\text{input}}$ of the given star. Using the MAP estimates one can compute an MAP re-simulation error over the test set following

$$\text{rmse}_{\text{resim}}^{\text{MAP}} = \sqrt{\frac{\sum_{i=1}^{N} (\mathbf{y}_{i,\text{resim}} - \mathbf{y}_{i,\text{input}})^2}{N}}. \tag{18}$$

Unfortunately, we do not have direct access to the stellar evolution code that our training data is based on, just the publicly available isochrone tables. Therefore, we cannot perform a full resimulation for our predictions.

To still get an idea of the resimulation error of our approach, we adopt a simple approximation instead. For a given MAP estimate or sample prediction of the physical parameters we do a nearest neighbour search in the $\mathbf{x} + A_{\text{V}}$ space on the training data (*after* the test split). Even though we do not predict the extinction, we have to include it in this nearest neighbour search to select the correct copy of the data point closest to our query. We note that also in a full resimulation we would have to input extinction to correctly retrieve the magnitudes. This approach allows us to report the approximate MAP resimulation error $\text{rmse}_{\text{resim}}^{\text{MAP}}$ on our synthetic training data (see Table 1).

It is important to keep in mind though that this is only an *approximation*, so that in cases where the distance to the nearest training data point is large in this 7D parameter space, the associated magnitudes might not necessarily represent the true resimulated observables of a given prediction. It is therefore likely that this approximation tends to overestimate the resimulation error.

4 TRAINING RESULTS

For all four of our models the cINN training process converges quickly, the training time being usually within 1–2 h when making use of GPU acceleration with a NVIDIA GTX 1080 graphics card. Once trained the prediction of posterior distributions is very rapid.

Table 1. Overview of the performance on a test set of 20 000 cases for the four cINN models we have trained. Reported are the calibration error, median uncertainty at 68 per cent confidence (width of the 68 per cent confidence interval), the standard (rmse) and nrmse of the MAP estimates, as well as the total MAP resimulation error $\mathrm{rmse}_{\mathrm{resim}}^{\mathrm{MAP}}$ from our nearest neighbour approximation.

Performance measure	Training set			
	Wd2_I	Wd2_II	NGC 6397_I	NGC 6397_II
Calibration error				
$\log$ age	0.005	0.001	0.005	0.011
$\log M_{\mathrm{ini}}$	0.009	0.006	0.006	0.004
$\log M_{\mathrm{curr}}$	0.009	0.004	0.007	0.007
$\log L$	0.068	0.048	0.003	0.007
$\log T_{\mathrm{eff}}$	0.028	0.020	0.007	0.003
$\log g$	0.013	0.003	0.006	0.007
Median uncertainty at 68 per cent confidence				
$\log$ age	0.199	0.049	0.065	0.120
$\log M_{\mathrm{ini}}$	0.004	0.002	0.004	0.001
$\log M_{\mathrm{curr}}$	0.004	0.003	0.004	0.002
$\log L$	0.002	0.002	0.005	0.001
$\log T_{\mathrm{eff}}$	0.001	0.001	0.001	0.001
$\log g$	0.006	0.004	0.004	0.002
rmse				
$\log$ age	0.572	0.379	0.481	0.1659
$\log M_{\mathrm{ini}}$	0.065	0.120	0.018	0.0036
$\log M_{\mathrm{curr}}$	0.064	0.074	0.019	0.0036
$\log L$	0.093	0.154	0.008	0.0011
$\log T_{\mathrm{eff}}$	0.041	0.071	0.003	0.0002
$\log g$	0.131	0.200	0.021	0.0034
nrmse				
$\log$ age	0.1122	0.1263	0.0938	0.1468
$\log M_{\mathrm{ini}}$	0.0180	0.0334	0.0050	0.0028
$\log M_{\mathrm{curr}}$	0.0179	0.0207	0.0053	0.0028
$\log L$	0.0091	0.0160	0.0008	0.0002
$\log T_{\mathrm{eff}}$	0.0207	0.0366	0.0023	0.0003
$\log g$	0.0191	0.0291	0.0038	0.0007
$\mathrm{rmse}_{\mathrm{resim}}^{\mathrm{MAP}}$	0.071	0.123	0.078	0.043

For the 20 000 observations in our test sets generating the posterior distributions, sampling each 4096 times, takes in total about 10 min, averaging around 35 predicted posterior distributions per second. This makes the cINN approach a very time efficient predictor.

4.1 Performance overview

Across all four cINN models we were able to achieve well converged model solutions. Both the covariance of the latent variables, as well as their distributions, evaluated on the respective test sets, reach their targets of unity and standard normal distribution, respectively. Fig. B1 in Appendix B shows an example of the achieved covariance matrix and latent variable distributions for the 'Wd2_I' cINN model. Table 1 gives an overview of our remaining performance measures, namely the median calibration error, the median uncertainty at 68 per cent confidence, the rmse and nrmse of the MAP point estimate (see equations 15 and 16), as well as our approximation of the total resimulation error across all four trained models.

In terms of the median calibration error, we find that all four models reach calibrated solutions for their predicted posterior distributions, as the largest error across all parameters and models is only about 6.8 per cent. Given the similar magnitude of the errors for all four models, there is no clear influence of the training set size or feature

abundance on the cINN's ability to converge to a well-calibrated solution. In particular, there is no significant difference between the models trained on the full training sets 'Wd2_I' and 'NGC 6397_I' versus their counterparts 'Wd2_II' and 'NGC 6397_II'. As the latter include prior knowledge about the age of the clusters, they should theoretically allow for more accurate solutions of the regression problems (i.e. less degenerate mappings). The only notable difference between the Wd2 and NGC 6397 models in terms of the median calibration error is that we find slightly better calibrated solutions for the luminosity and effective temperature prediction for the NGC 6397 models.

Concerning the median uncertainty at the 68 per cent confidence level, an indicator of the average width of the predicted posterior distributions, we find that all four trained cINN models can constrain all physical parameters, except for the age, remarkably well with uncertainties on the order of only a few 0.001 dex on average. Again, the availability of more features or the prospect of less degenerate mappings by including prior knowledge does not significantly improve the result.

Judging by the uncertainty values, the stellar age appears to be the most difficult parameter to constrain. Of the six parameters, age is also the only one where the prediction is influenced by the amount of available features. The 'NGC 6397_I' cINN model constrains the age to distributions that are about 0.1 dex narrower than the

similar one trained on 'Wd2_I', despite the fact that both training sets cover basically the same physical parameter space (albeit at different metallicity). For the age prediction, we also observe a difference between the 'Wd2_I' and 'Wd2_II' models, as the cINN trained on the data set including prior knowledge returns narrower age posterior distributions. The lower uncertainty is likely influenced by the overall smaller range in possible predicted ages, but could also be a result of the missing degeneracies in 'Wd2_II'. Interestingly, we do not observe the same effect between 'NGC 6397_I' and 'NGC 6397_II', where in fact the median uncertainty increases for the model trained on the much narrower age range. This could indicate that constraining the age distribution for these old stars (above 1 Gyr) may not facilitate the regression problem, while the reverse may be true for the young stars.

The point estimate accuracy, as measured by the rmse between the MAP prediction and the true values, confirms that age is the most difficult parameter to predict for all our models. With rmses of a few 0.01–0.1 dex, the cINN predicts the remaining five physical parameters very well, while the rmse for the age prediction, on the order of 0.5 dex, is about a magnitude larger. For comparison, a predictor that returns a random value drawn from a uniform distribution within the age range of 'Wd2_I' achieves an rmse of about 2.1 (nrmse of 0.41). The $\sim$0.1 dex differences in the rmses between the models trained on 'Wd2_I' and 'NGC 6397_I' suggest that an increased feature abundance (i.e. number of observables) improves the point estimation accuracy of the model. Interestingly, while the 'Wd2_II' model decreases the age rmse by about 0.2 dex, the error of the point estimate for all remaining physical parameters increases. Comparing the nrmses between 'Wd2_I' and 'Wd2_II', however, we find that both models perform evenly well and all flat rmse differences are likely effects of the different parameter ranges. We find a similar behaviour between 'NGC 6397_I' and 'NGC 6397_II' for all parameters except the age again, where the 'NGC 6397_II' model actually performs the worst across all models. As previously indicated by the uncertainty, this supports the finding that the age prediction within the range from 1 to 13 Gyr is the most difficult task on the synthetic data.

Finally, for our approximation of the total MAP resimulation error we find excellent results for all of our models, with values on the order of only 0.1 mag and below. Considering that our approximation likely overestimates this error because we have to rely on the observables of a nearest neighbour proxy, errors this small are more than satisfactory. The corresponding comparisons of the 'resimulated' and observed magnitudes show almost perfect 1-to-1 correlations with very few outliers in both the MAP and entire posterior resimulations. Therefore, we are very confident that, even though we could not perform a true resimulation, our predicted posterior distributions are true and not just numerical artefacts. Importantly, this also indicates that the overall broader age posteriors are generally not caused by an underperforming cINN but rather due to actual intrinsic degeneracies in the age prediction, correctly captured by the cINN.

4.2 Wd2_I and Wd2_II

As indicated by the summary statistics in Table 1 the 'Wd2_I' and 'Wd2_II' cINN models perform very well. We look at this in more detail. Fig. 7 shows example posterior distributions for all six physical parameters for three held-out test observations predicted by the 'Wd2_I' model. This plot exhibits some of the typical posterior distributions that the cINN returns on the synthetic data in this regression problem.

The first case, shown in the left-hand column, is an example where the cINN constrains all physical parameters of the star extremely well with very narrow posterior distributions centred around the known true value. As the low median uncertainties at 68 per cent confidence for all parameters except age already suggest, this kind of prediction is among the most common results for the synthetic test set. The left-hand panel in Fig. 8 presents the approximate resimulation for the full posterior of this example. Evidently we match the input observation almost exactly (note the small axis range and error), confirming the validity of the predicted posterior. The observed deviation is a direct result of the nearest neighbour approximation and the discreteness of the training set. The latter is also the reason why the two resimulation solutions appear so 'far' apart, as there simply are no models in between. The samples with a greater discrepancy to the true observation (bottom left corner) have a larger distance to the nearest neighbour than the others. The resimulation approximation is therefore less precise for these samples as the distance is a direct measure of similarity between the nearest neighbour proxy and the given query samples.

In contrast, the middle column exhibits an example of the kind of degeneracy that we frequently find in this regression problem, with bimodal solutions within the predicted posterior distributions. The age and mass distributions indicate that this observation could be explained by a $\sim$2.9 M$_\odot$ star that is $\sim$425 Myr old, so likely well within its post-main-sequence phase. Or it could be a very young ($\sim$0.1 Myr) more massive $\sim$4.75 M$_\odot$ pre-main-sequence star. Due to the overlap of the post-main-sequence and pre-main-sequence evolution in observable space, especially so in the presence of extinction, this is one of the major degeneracies that make the prediction of stellar physical parameters from photometry such a difficult regression problem. In this example the cINN prediction reveals that this degeneracy is not broken with only two passbands, but also finds that the young 4.75 M$_\odot$ star is the most likely solution as indicated by the MAP estimates. Therefore, the cINN successfully recovers the true solution for this synthetic star.

The middle panel of Fig. 8 shows the approximate resimulated magnitudes for this example posterior in comparison with the true input observations. Overall we find very good agreement, except for a few outlier cases. The red circle indicates the area populated by the 60 per cent of the samples (containing instances from both peaks) with the lowest distance to the nearest neighbour used as a resimulation proxy. This set matches the true observations almost perfectly. All of the outliers exhibit larger nearest neighbour distances (especially the far outliers). Consequently, our resimulation approximation is less precise for these objects, which likely explains the offset from the true observation. Therefore this diagram confirms the validity of the predicted posterior and the identified degeneracy.

The final example in the right-hand column shows another degenerate case that could be explained as either a younger $\sim$3 M$_\odot$ or a much older $\sim$2.4 M$_\odot$ star. Here, the most likely explanation of the observation as given by our MAP estimate is in fact not the true one, which falls into the secondary peak. This result may seem unsatisfying at first glance, but a true posterior distribution describes all *possible* physical parameters that can explain the given observation. That means that the most likely combination does not necessarily have to be the one that generated the observation. In fact these two degenerate examples show the great strength of the cINN approach for this type of degenerate regression problem, as even in the second case, the true solution is part of the posterior distribution as the second most likely result. The interpretation of
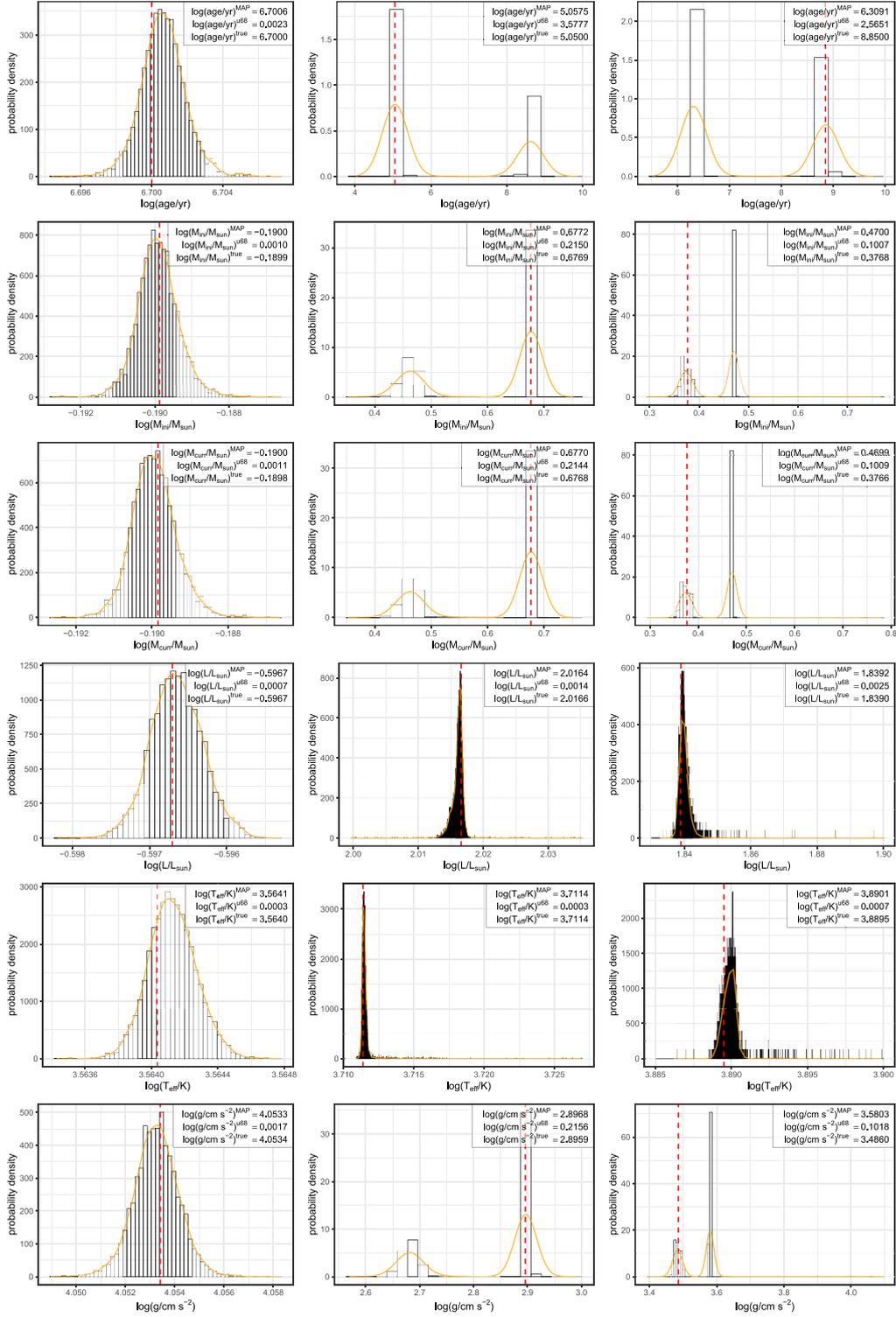
Figure 7. Predicted posterior distributions for three test examples (column wise) as predicted by the cINN trained on 'Wd2_I'. The red dotted line in each histogram indicates the known true value for the given test observation. The orange line represents the kernel density estimate of the predicted distribution used to locate the MAP solution. The left-hand column shows an example case where the cINN is able to constrain the physical parameters of this observation extremely well. The remaining two columns show degenerate examples where the predicted posterior distributions of some parameters (e.g. age and mass) show multimodalities as a consequence. The middle column test observation shows an example case where the MAP of the predicted bimodal distribution coincides with the true value, while in the right-hand column case the true value falls on to the second peak of the distribution. Note the different scaling in each column.
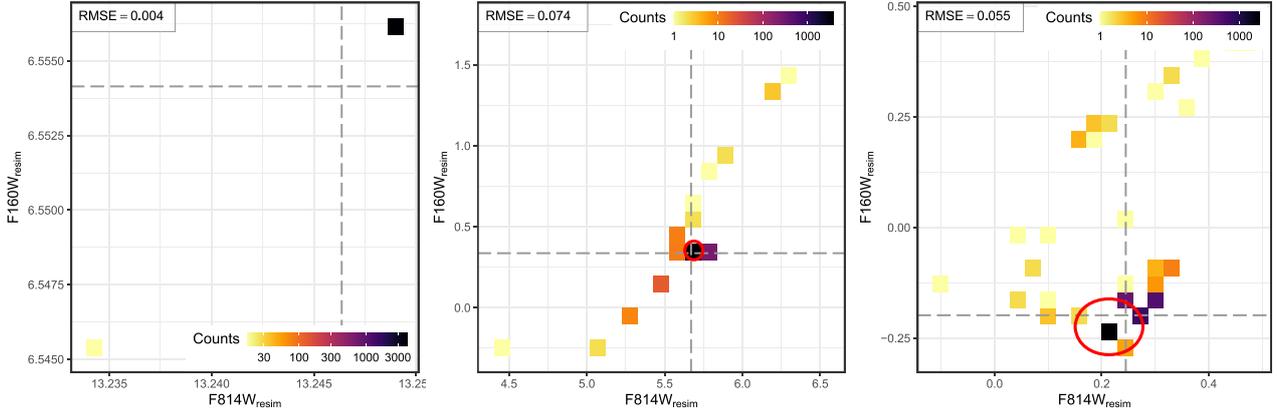
Figure 8. 2D histograms of the 'resimulated' magnitudes for the example posteriors in Fig. 7 (columns match accordingly). The grey dashed lines indicate the observed magnitudes, while the red circles in the middle and right-hand panels indicate the area in which 60 per cent of the samples are located that have the lowest distances to the nearest neighbour chosen as the resimulation proxy. Again, it should be mentioned that the axis scaling is very different in each plot.
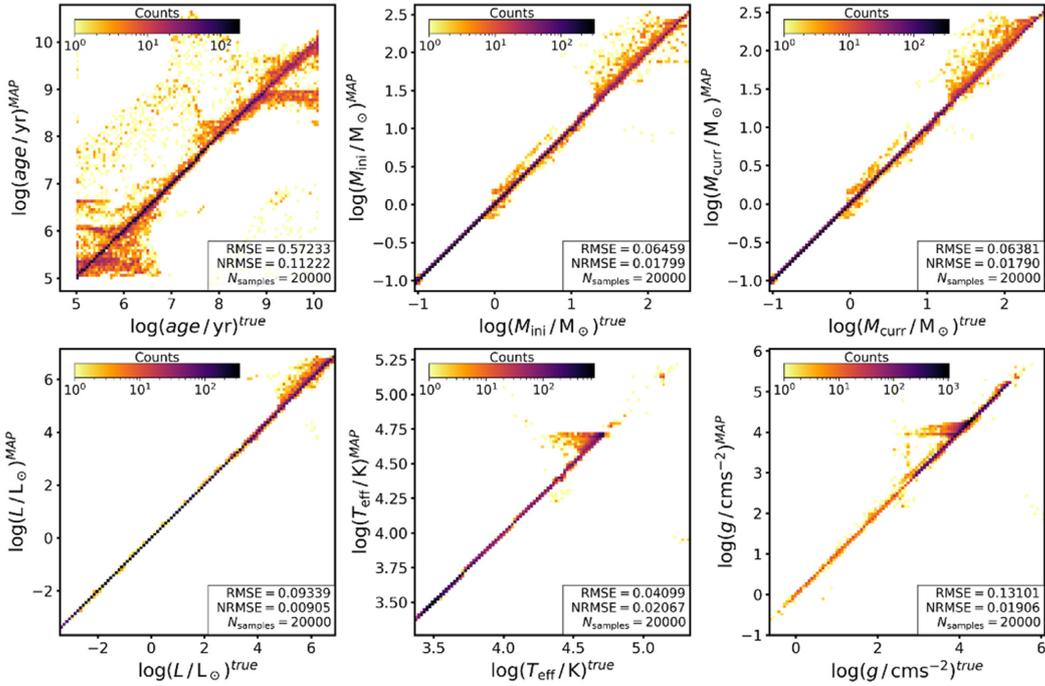
Figure 9. 2D histograms of the MAP estimates plotted against the true values for the six physical parameters we predict with the cINN trained on 'Wd2_I' for 20 000 cases from our test set. From top left to bottom right, we show age, M_{ini}, M_{curr}, L, T_{eff}, and g.

cases like these, as always, benefits from additional astrophysical constraints.

The right-hand panel of Fig. 8 provides the resimulation approximation for this example. Again we find a good match with the observation for the objects for which our approximation is the most precise. Only objects with large distance to the nearest neighbour, so less precise resimulation approximation, deviate more significantly.

To assess the limitations of the method tested on the synthetic data we compare the predicted point estimates with the true values (as previously summarized by the rmses in Table 1) for the 20 000 test observations in Fig. 9. The figure highlights how well the cINN

predicts M_{ini}, M_{curr}, L, T_{eff}, and g as only very few predictions (note the logarithmic colour scale) fall off a perfect 1-to-1 correlation between the predicted and true values. However, for T_{eff} and g we observe some structure (around $\log(T_{eff}/K) \approx 4.75$ and $\log(g/\mathrm{cm\,s^{-2}}) \approx 4$, respectively) that seems systematic in nature. For the effective temperature there is also a deviation from the 1-to-1 correlation for $\log(T_{eff}/K) > 4.75$.

We find the largest scatter in the age prediction, confirming that this parameter is the most difficult to predict for the cINN. It has the most trouble with predicting ages for the very young ($\log(age/yr) < 6.5$) and the oldest ($\log(age/yr) > 8.5$) objects in the test set, as we find the most deviations from the perfect correlation here. Still, even
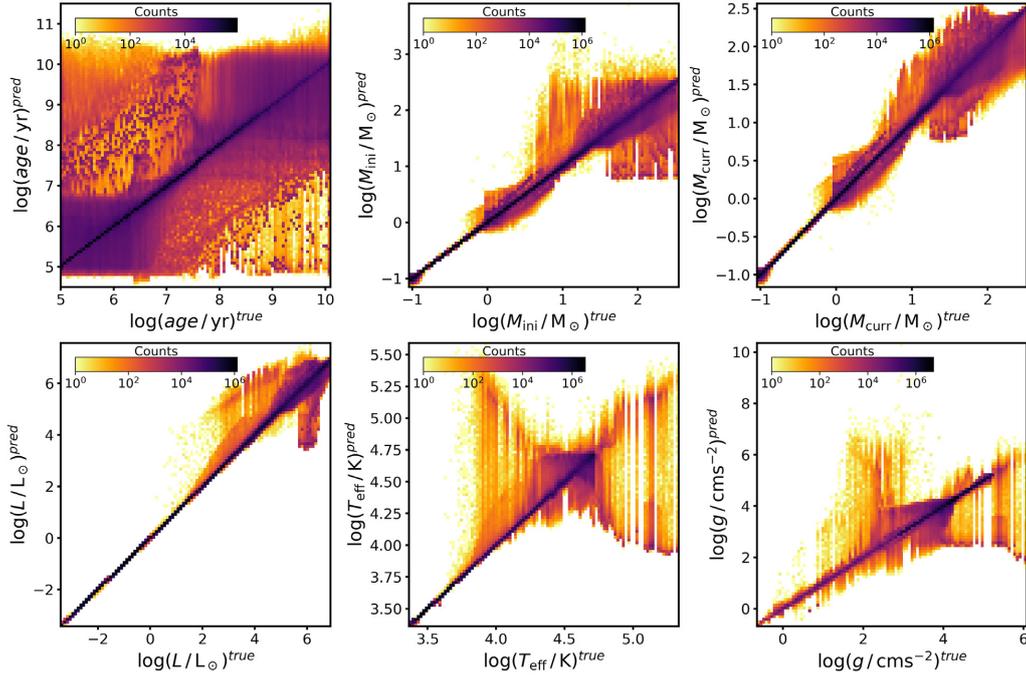
Figure 10. 2D histograms of the entire predicted posteriors plotted against the true value of the six physical parameters of the cINN trained on 'Wd2_I' for 20000 cases from the test set. From top left to bottom right age, M_{ini}, M_{curr}, L, T_{eff}, and g are shown.

in this regime, there is a majority of good predictions (note again the logarithmic colour scale).

The difficulty in predicting the correct age becomes further apparent when visualizing the posterior distributions in relation to the true values, as in Fig. 10. Here, we plot the spread of the posterior distribution of every physical parameter against the true value for all 20 000 test observations. Again we observe that for all physical parameters except age the cINN provides well constrained posterior distributions that are in many cases quite narrow and symmetric around the true value. Similar to the systematic structures in the MAP estimates we find 'arrow'-like 'artefacts' for T_{eff} and g here. For $\log(T_{\text{eff}}/\text{K}) > 4.75$, we also discover two 'branches', indicating a strong bimodal degeneracy in this range that explains the deviation from the 1-to-1 correlation in Fig. 9, as the MAP estimates seemingly tend to fall into this lower branch.

The age posterior distributions appear to be much wider, although the visual effect is amplified in Fig. 10 by the logarithmic colour scaling, chosen to better visualize outliers. Most of the predicted posterior distributions are also well centred on the true value, but nevertheless we find many more wide outliers here, indicating ample degeneracy. Analogous to the MAP estimates, the posterior distributions narrow down within the intermediate-age range and widen for the youngest and the oldest stars, also exhibiting the multimodalities previously highlighted in Fig. 7. Despite the slightly discouraging look of the age posteriors it is important to note that in 99.8 per cent of the cases the true value *is part* of the predicted posterior distribution.

To evaluate whether the 'arrow' artefacts observed in the MAP estimates and posteriors of T_{eff} and g are a cINN model specific issue, we retrained the cINN model on modified versions of training set 'Wd2_I', where we increase the number of observables with additional photometric filters. Within the synthetic data sets these additional filters are readily available. Fig. 11 shows the results of this experiment. It provides the posterior against true value diagrams for age and surface temperature for different numbers of additional photometric filters. This sequence shows that the 'arrow'-like structures in T_{eff} and g, as well as the second branch in T_{eff}, are in fact a result of the limited number of photometric filters in our study, as the effect already decreases when the F555W filter is added and basically disappears when we use nine photometric filters. Not surprisingly, the predictions also improve as more observational information is gained, the posterior distributions narrowing down noticeably. Especially, interesting for the age prediction is that we already observe a considerate improvement with five filters (F275W, F336W, and F555W on top of F814W and F160W). Specifically the spread for very young objects ($\log(\text{age/yr}) < 6.5$) decreases significantly. We observe the same improvement in the point estimates (see Fig. B2 in Appendix B). Still, even with the 'ideal' information of the full complement of 17 photometric filters of the '*HST* WFC3 wide' photometric system used by the PARSEC isochrones, the prediction for old stars is not perfect. The age prediction of old stars thus remains the most challenging task within this regression problem (see also the discussion in Sections 4.3 and 5.2). In any case, based on this performance analysis, we *recommend* to use at least five photometric filters in addition to extinction if they are available.

The model trained on 'Wd2_II' does not show significant difference with respect to the 'Wd2_I' cINN model within their range of overlap. The corresponding diagrams of the point estimate and posteriors against the true values for 'Wd2_II', as well as a more detailed discussion can be found in Appendix C, Figs C1 and C2, respectively.
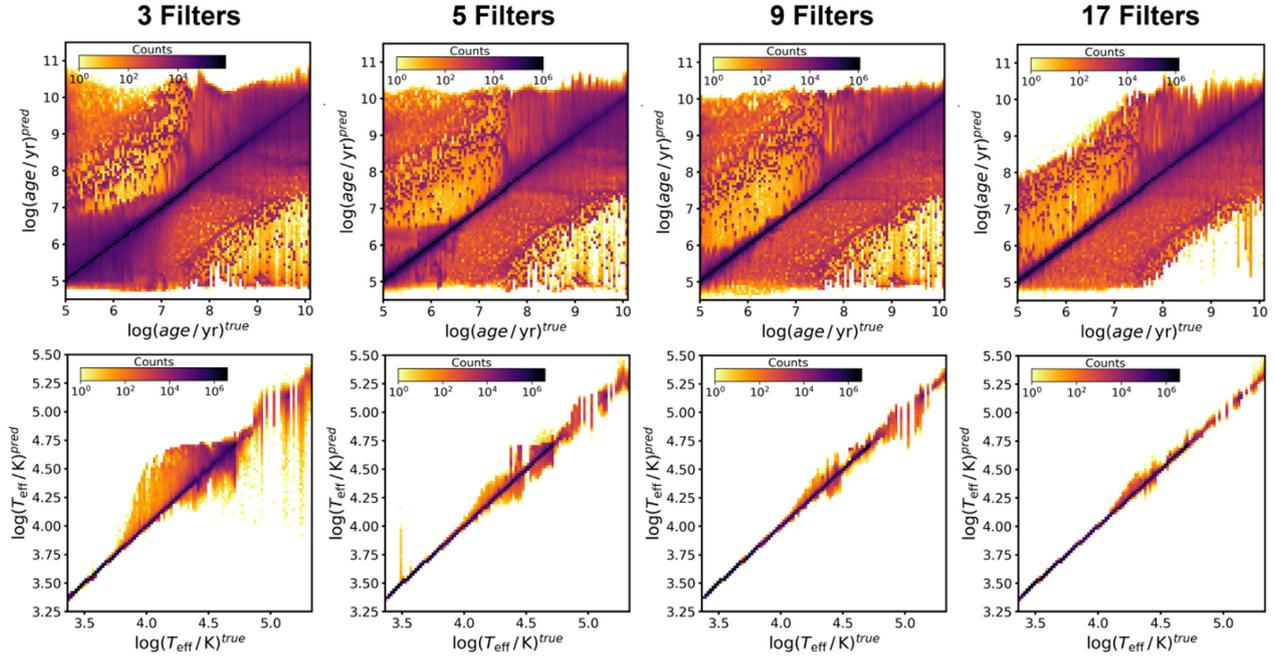
Figure 11. 2D histograms of the predicted posterior distributions plotted against the true value for age (top) and T_{eff} (bottom) for 20 000 test observations as predicted by cINN models trained on 'Wd2_I' with increased numbers of photometric filters compared to our standard 'Wd2_I' setup. The three filter case (first column) entails the *HST* filters F555W, F814W, and F160W, the five filter one (second column) adds F275W and F336W to that, and the nine filter setup (third column) further includes F438W, F606W, F775W, and F110W on top of the previous. The final 17 filter case (fourth column) entails all previous filters in addition to F218W, F225W, F390W, F475W, F625W, F105W, F125W, and F140W. Compared to our standard 'Wd2_I' with only two filters in Fig. 9 this sequence highlights how the increased feature abundance improves the predictive capability of the cINN as the 'arrow'-like artefacts in T_{eff} disappear and the age posterior distributions decrease in width, especially so for the young test observations.

4.3 NGC 6397_I and NGC 6397_II

Overall the training results of model 'NGC 6397_I' match those of 'Wd2_I', except for the previously described slight improvements in accuracy. Judging by our performance experiments in dependence of filter coverage carried out on 'Wd2_I', these improvements are likely caused by the larger number of photometric filters, five instead of the two used for 'Wd2_I' (see Appendices B and D).

In general, of all trained models 'NGC 6397_II' provides the smallest rmses across all predicted physical parameters and lowest median uncertainty for all parameters but age. Given how well the 'Wd2_I' and 'NGC 6397_I' models already constrain the posterior distributions for all parameters (except age), this extra performance gain can be attributed to the more limited physical parameter space. The nrmse of this cINN model confirms again that the age prediction for very old stars (1 Gyr and above) is the most difficult part of this regression problem. We find that the age posterior distributions tend to be quite broad and that the cINN has a tendency to extrapolate with predicted posterior distributions ranging from $\log(age/yr) = 8$ to above $\log(age/yr) = 11$ (see Fig. E2 in Appendix E), outside the boundaries of the training set range of 9–10.13. This extrapolatory behaviour within the 1–10 Gyr range appears in the 'Wd2_I' and 'NGC 6397_I' models as well, but to a lesser degree. From the age MAP estimate against true plot of the 'NGC 6397_II' model, we also find that, while most predictions fall on the ideal 1-to-1 correlation, there is a faint trace of an almost flat 'branch' at $\log(age/yr) \approx 9.6$ (see Fig. E1 in Appendix E). This might suggest that the cINN has a slight tendency to predict something akin to a mean age value (9.6 is exactly the average) over the trained range when it encounters a star with uncertain age.

4.4 On the age prediction of main-sequence stars

One matter we have not discussed in detail so far is the age prediction for main-sequence objects. With traditional isochrone fitting methods recovering the age of a main-sequence star from photometry alone is a notoriously difficult, if not impossible task. Our approach, on the other hand, successfully predicts ages across the entire spectrum of objects, including synthetic main-sequence stars. Given the difficulties traditional approaches have, this could be an indication that our cINN models achieve this task only by overfitting the synthetic training data. To ascertain whether this is the case we perform a test prediction with the 'Wd2_I' and 'NGC 6397_I' models on synthetic data generated from different stellar evolution models, namely the MIST (Paxton et al. 2011, 2013, 2015; Choi et al. 2016; Dotter 2016) and Dartmouth (Dotter et al. 2007, 2008) isochrone tables. These models also provide synthetic photometry, but treat the underlying physics slightly different than PARSEC. Note that the Dartmouth isochrones only cover an age range of 1–15 Gyr, while the MIST tables are available over a similar $\log(age/yr)$ span of 5–10.3 as the PARSEC models. For the test, we choose data sets matching the corresponding metallicities for 'Wd2_I' and 'NGC 6397_I', and for simplicity only treat the zero extinction case. Additionally, for the MIST data, we remove the post-asymptotic giant branch phase as our selection of PARSEC models (version 1.2s) does not include it.

In the solar metallicity 'Wd2_I' case, we retrieve overall excellent results (see Fig. 12). In particular, for the MIST data the cINN recovers $\log(L)$, $\log(T_{eff})$, and $\log(g)$ almost perfectly, except for a few instances of massive post-main-sequence stars. M_{ini} and M_{curr} are also recovered well, but exhibit more scatter than in our PARSEC
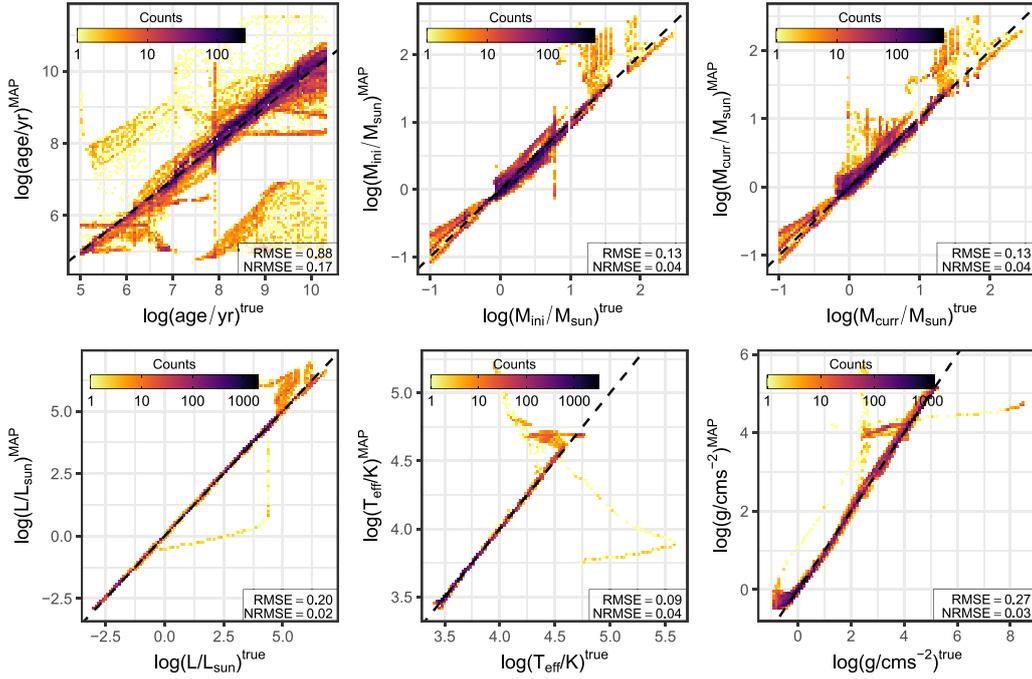
Figure 12. 2D histograms of the MAP predictions for the six physical parameters on 40 000 samples from the MIST isochrone tables as predicted by the 'Wd2_I' cINN model. Note that the nrmses are normalized to the parameter ranges of the MIST ground truth here instead of the ranges in the 'Wd2_I' training set.

test case. Lastly, while the age prediction also exhibits more of a spread around a perfect 1-to-1 correlation, with a median absolution deviation of only 0.2 dex, the cINN correctly retrieves ages for most samples, *including* main-sequence objects. Fig. 13 shows the predicted HRD in comparison to the MIST ground truth, highlighting the excellent performance of the cINN. Note that the predicted ages are represented by the best-fitting peak in the age distribution here in order to account for multimodal distributions found for, that is, post-main-sequence objects. We find a similar success for the Dartmouth models, recovering luminosity, temperature, and gravity near flawlessly (see Fig. B3 in Appendix B). The initial mass predictions are overall also fairly accurate, but exhibit a slight systematic overprediction below 0.5 M$_\odot$, likely an effect of the known model discrepancy between Dartmouth and PARSEC in the subsolar mass regime. The age prediction on the other hand is slightly less successful here. While we can recover ages for most post-main-sequence objects, taking multimodalities in the posteriors into account, and for some main-sequence stars down to about 0.75 M$_\odot$, below this mass limit we find larger errors (see also Fig. B4 in Appendix B). A likely explanation for this behaviour is a combination of the fact that the cINN also struggles in the range above one Gyr on the PARSEC test data and the significant model difference of Dartmouh and PARSEC in the low-mass regime.

With the low-metallicity 'NGC 6397_I' model, we are also fairly successful on the MIST synthetic test data (Figs D3 and D4 in Appendix D). Interestingly, despite using photometry in three more filters we get overall larger errors compared to the 'Wd2_I' test. It appears that the differences between the stellar evolution models, for example, in the model stellar atmospheres, become more significant outside of the solar metallicity case. The 'NGC 6397_I' model also

recovers luminosity and temperature well, but has more difficulties with the age prediction. Still for a large fraction of test objects, both main- and non-main sequence, a correct age is inferred (median absolute deviation of 0.3 dex). For the Dartmouth test data, we find overall the worst results with the 'NGC 6397_I' model in this experiment (see Figs D5 and D6 in Appendix D). While the cINN recovers luminosity, temperature, and gravity decently for most test samples, we find larger systematic deviations in the low brightness regime. Likewise we find a significant discrepancy for the predicted initial masses within a range from 0.25 to 0.6 M$_\odot$ and for some objects above 0.8 M$_\odot$. Lastly, the age prediction fails completely for this synthetic test set with the 'NGC 6397_I' model systematically underestimating the age. Given that the prediction performance on the MIST data is acceptable, we conclude that the significant model discrepancy between Dartmouth and PARSEC at this metallicity, especially in the synthetic photometry, is the primary reason for the cINNs difficulties.

In summary, these experiments provide good evidence that our cINN models have not simply overfit the synthetic PARSEC training data as they are able to recover correct ages in most cases for test data from different stellar evolution codes, including ages of main-sequence objects. Furthermore, this test shows that the cINN generalizes well to slightly different populations and especially excels in recovering luminosity, temperature, and surface gravity. Concerning the predictions for main-sequence stars, we believe that a combination of the latent variable approach, encoding enough of the lost information, and the fact that we are using *perfect* photometry allows the cINN to correctly recover ages for these objects. Consequently, as real photometry is never perfect, we acknowledge that the cINN age prediction for any real main-sequence
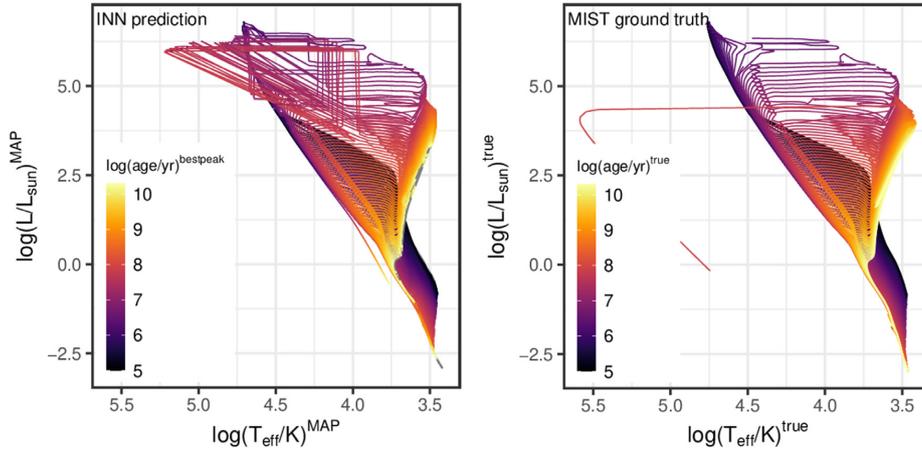
Figure 13. Left: cINN prediction for the HRD of the MIST isochrones. Note that MAP estimates are used for luminosity and effective temperature here, but the colour code that indicates the predicted age does not correspond to the MAP age prediction but rather the best-fitting peak of the predicted age posterior. The latter is done to take multimodal age posteriors into account. Right: ground truth HRD of the MIST isochrones, colour coded according to their age.

star needs to be treated with caution. We will further discuss this matter in our application to the real NGC 6397 data, as this cluster consists primarily of main-sequence sources, contrary to the young Wd2.

5 PREDICTION

With the excellent performance of the cINN on the synthetic training data for Wd2 and NGC 6397, we can now benchmark the method on real observational data. As with the synthetic test set, to retrieve the posterior distributions we sample the latent variables 4096 times for each star and determine point estimates for all physical parameters as described in Section 3.5. Since we have seen no significant differences between the full models and those that entail prior knowledge about the age on the synthetic data, in the following we take the full model predictions as our primary reference and perform a short comparison with the other models at the end of each section (providing further details in the Appendix).

5.1 Westerlund 2

Fig. 14 presents our cINN prediction results for all six physical parameters, showing their MAP estimates colour coded on the optical CMD of Wd2 (cluster members only). Overall the results are very reasonable for Wd2, from subsolar masses for low-mass pre-main-sequence (PMS) stars to above solar masses for UMS stars, with the correct gradients of L, T_{eff}, and g versus magnitude and colour. On top of that the median 1.27 Myr cluster age from MAP estimates is well within the previously determined age range of 1.04 ± 0.72 Myr. The resulting HRD, shown in Fig. 15 (top left panel as per MAP estimates, top right panel as per entire posteriors), also matches fairly well the 1 Myr isochrone traced in red for comparison. There is a noticeable spread around the isochrone, but most of the stars are correctly placed within the PMS regions of the diagram. Notable is only a small vertical feature at the extreme right of the predicted HRD, highlighted by the orange points in the top left panel of Fig. 15, which appears to be deviating more systematically from the 1 Myr isochrone. These 502 stars, all located at the very red edge of the CMD (bottom left panel of Fig. 15) have a median photometric error of 0.15 mag. It is quite possible that the cINN prediction entails

this vertical artefact due to photometric uncertainties, which are not accounted for in our setup.

We also find other mispredictions from the cINN. For 584 stars (179 among the HRD outliers), the initial mass MAP estimate falls below the $0.09 M_{\odot}$ minimum of the training set and in 292 cases even below the H-burning threshold of $0.072 M_{\odot}$ (Solar metallicity, Chabrier 2002). With a minimum of $0.05 M_{\odot}$, the mass estimates for these stars (red points, Fig. 15 bottom centre panel) are still physically plausible for, that is, young brown dwarfs, but this extrapolation might indicate a systematic error. Like the HRD outliers these objects are subject to a notable amount of photometric uncertainty (median of 0.2 mag in F814W), being a likely culprit for these mispredictions.

For another 818 stars (343 also in HRD outliers), the MAP age estimate is below the 0.1 Myr training set minimum, going down to 0.02 Myr. Given their location at the red edge of the CMD (blue points, Fig. 15 bottom left panel,) these results are somewhat plausible but not convincing. Aside from the photometric uncertainties, limitations of the Zeidler et al. (2015) prescription to estimate stellar extinction from gas colour excess could provide an explanation for these results, if, for example, the stellar extinction has been underestimated for these objects.

Lastly, a number of stars are predicted to be unreasonably old for Wd2. These are located primarily at the very blue and red edges of the PMS population in the CMD, but we also find 86 among them on the turn-on (highlighted in Fig. 16). The former could potentially be field contaminants that survived our initial rejection using Besançon models in the direction of Wd2 (Zeidler et al. 2015) and are correctly identified as old. Evidence for this hypothesis is that we identify these outliers in our age prediction primarily in the CMD region where Zeidler et al. (2015) find an overlap of the Besançon models and the cluster constituents.

For the 86 turn-on stars only the MAP age estimate is incorrect, as almost all of them show degenerate age posteriors with a prominent second peak close to the supposed cluster age. Fig. 16 presents three example age posteriors for these turn-on objects and one from the majority of well-constrained solutions (bottom left panel) for comparison. In the top left posterior example, a common case, an old age appears as the most likely solution, but we find a secondary maximum at the cluster age. The top right panel represents another frequent outcome among these 86 stars, where the young and old
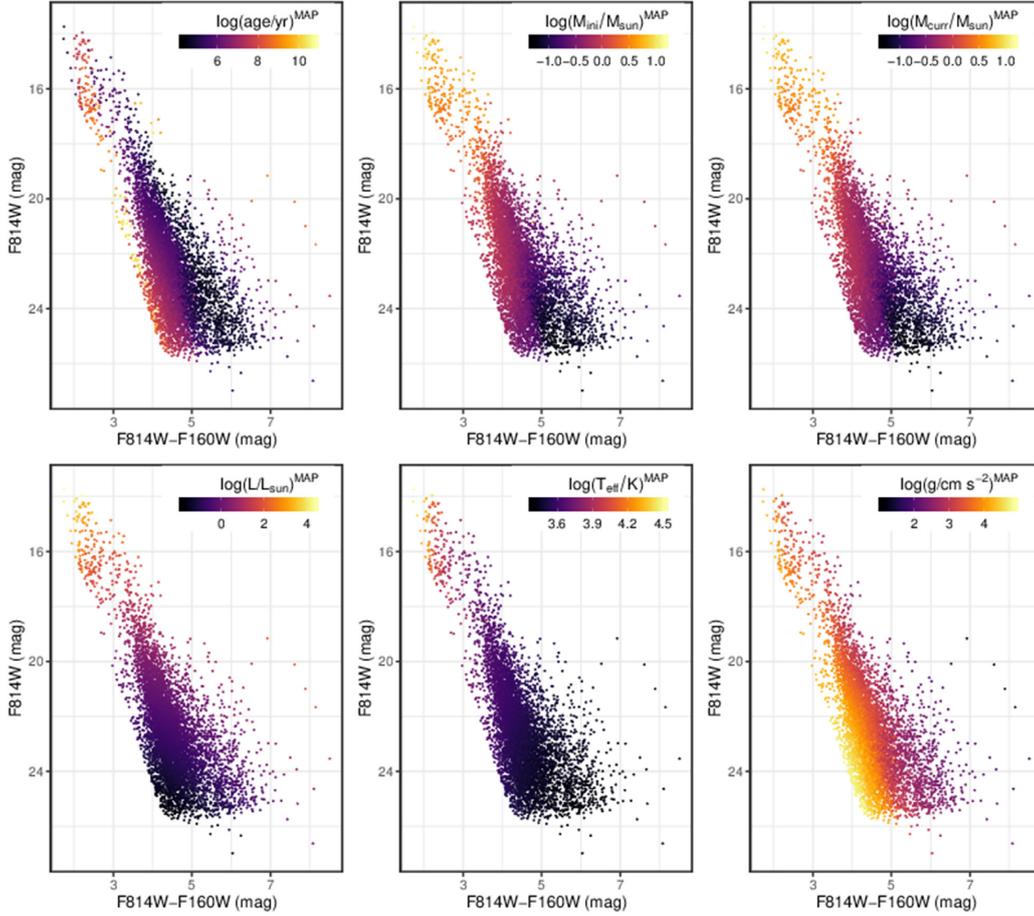
Figure 14. Optical CMDs of the Wd2 *HST* data, colour coded according to the MAP estimates for the six physical parameters predicted with the cINN trained on 'Wd2_I'.

solution are almost equally likely. The (rarely occurring) final case in the bottom right shows a 'complete failure', where no prominent secondary maximum exists at the cluster age. Given that (field) RGB stars can very well overlap with PMS stars within the main-sequence turn-on region, these results demonstrate again the great strength of the cINN approach as it recognizes and shows this possibility in the predicted age posterior distributions. At the same time, these examples serve as a reminder that careful post-processing (e.g. identification of all major peaks) of the predicted posterior distributions is necessary to avoid possible false conclusions by, for example, relying only on MAP point estimates.

Comparing the predictions on the Wd2 *HST* data between the models 'Wd2_I' and 'Wd2_II', we find that they agree well with each other. See Appendix C and Fig. C5 for more details. We conclude that inclusion of prior knowledge in the form of a simple range cut of the training set does not benefit the cINN approach in the Wd2 case.

5.1.1 Cluster age

Having assessed the overall satisfying prediction results of the 'Wd2_I' cINN model we now derive some physical properties of the cluster and compare them to previous studies.

To begin, we derive a cluster age from our individual stellar age predictions. As previously mentioned from the MAP stellar age estimates, we find a median age of the cluster stars of $1.27^{+3.62}_{-0.94}$ Myr. Determining the cluster age as the most likely value from the sum of all the individual age posterior distributions (Fig. 17) using a kernel density estimate, we find a value of $1.04^{+8.48}_{-0.90}$ Myr (MAP and edges of 68 per cent confidence interval). We find an almost identical result for the same derivation with the 'Wd2_II' model (see Appendix C and Fig. C6). While we cannot constrain the cluster age more precisely than the previous study by Zeidler et al. (2016), both of our values match the previously derived age within their errors. This is a very satisfactory result given that our method derives the cluster age without any prior knowledge, just on the basis of the stellar magnitudes in two photometric broad-band filters and an extinction estimate.

5.1.2 The stellar initial mass function

As our method predicts the initial mass of each star of Wd2, we can also analyse the IMF of the cluster, shown in Fig. 18. We suffer from incompleteness at the low-mass end and from saturation at the high-mass end but nevertheless, using the range from 0.5 to $\sim 5.6\,M_\odot$ as a proxy to derive the slope of the high-mass IMF, we find a value of $\alpha = 2.39 \pm 0.20$, which matches the Salpeter IMF slope of $\alpha =$
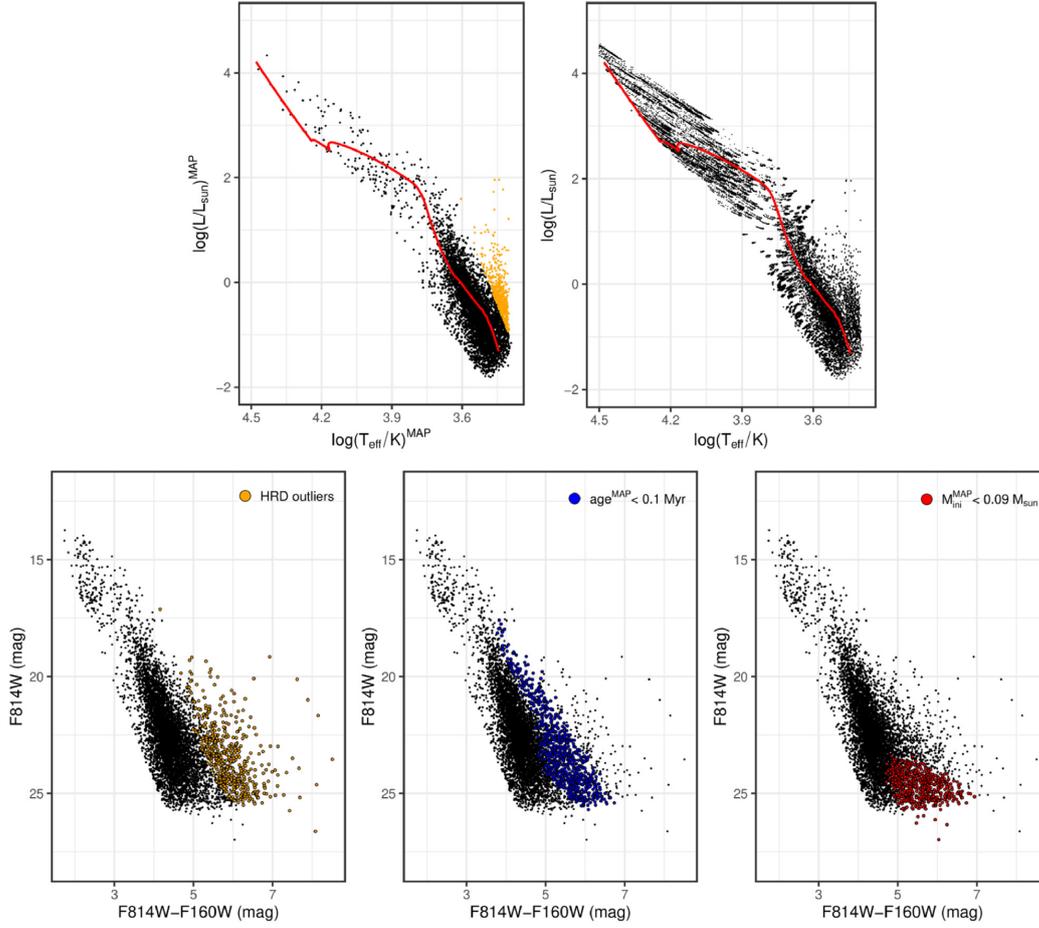
Figure 15. The top panels depict predicted HRDs for the Wd2 cluster constituents provided by the cINN model trained on 'Wd2_I'. The left-hand panel shows the HRD based on the MAP predictions for $\log(L)$ and $\log(T_{\rm eff})$, while in the right-hand panel the entire posterior distributions of these two parameters are plotted for every star. The red line in both diagrams indicates a 1 Myr isochrone for comparison. The orange points in the first panel indicate a possible vertical artefact in the cINN prediction. The bottom left panel indicates the observed CMD position of these HRD outliers. In the bottom centre and right-hand panels, we indicate the observed CMD positions of a few stars for which the predicted age (blue) or initial mass (red), respectively, are below the lower limits of the training data.

2.35 within 1σ. Zeidler et al. (2017) determine a present-day mass function (PDMF) with a slope of $\alpha = 2.53 \pm 0.05$ for the survey area of our Wd2 data. Presuming that the PDMF should not deviate too much from the IMF given the young age of Wd2, our slope is in good accordance with the result from Zeidler et al. (2017).

5.1.3 Mass segregation

Zeidler et al. (2017) also find evidence for mass segregation in Wd2 through the analysis of the PDMF within different annuli around the mid-point between the main and northern subcluster of Wd2. Using our individual stellar mass predictions, we try to confirm this finding by computing the mass segregation ratios (MSR) $\Lambda_{\rm MSR}$ (Allison et al. 2009) and $\Gamma_{\rm MSR}$ (Olczak et al. 2011). These two quantities are derived by constructing a minimum spanning tree (MST) for the N most massive stars within the population and comparing it with k MSTs of N random stars from the stellar sample. For $\Lambda_{\rm MSR}$, we then compute the tree length $l_{\rm mass}$ of the tree with the massive stars and the average tree length $\langle l_{\rm rand} \rangle$ of the k trees of random stars, so that

we find the MSR as

$$\Lambda_{\rm MSR} = \frac{\langle l_{\rm MST}^{\rm rand} \rangle}{l_{\rm MST}^{\rm mass}} \pm \frac{\sigma_{\rm rand}}{l_{\rm mass}}, \tag{19}$$

where $\sigma_{\rm rand}$ is the standard deviation of $\langle l_{\rm rand} \rangle$ (Allison et al. 2009).

$\Gamma_{\rm MSR}$ is given by the ratio between the mean edge lengths $e_{\rm mass}$ and $\langle e_{\rm rand} \rangle$:

$$\Gamma_{\rm MSR} = \frac{\langle e_{\rm rand} \rangle}{e_{\rm mass}}. \tag{20}$$

Here, we proceed in a fashion similar to $\Lambda_{\rm MSR}$, except that we now calculate the geometric instead of the arithmetic mean. For each of the k random MSTs, we determine the geometric standard deviation according to

$$\Delta e_{\rm rand}^k = \exp\left(\sqrt{\frac{\sum_{i=1}^{N} \left(\ln e_i^k - \ln e_{\rm rand}^k \right)^2}{N}} \right), \tag{21}$$

where e_i^k are the N edges of the kth tree (Olczak et al. 2011), and then derive the upper and lower 1σ intervals as the means of the k lower and upper 1σ intervals (note that $\Delta e_{\rm rand}^k$ is a multiplicative standard
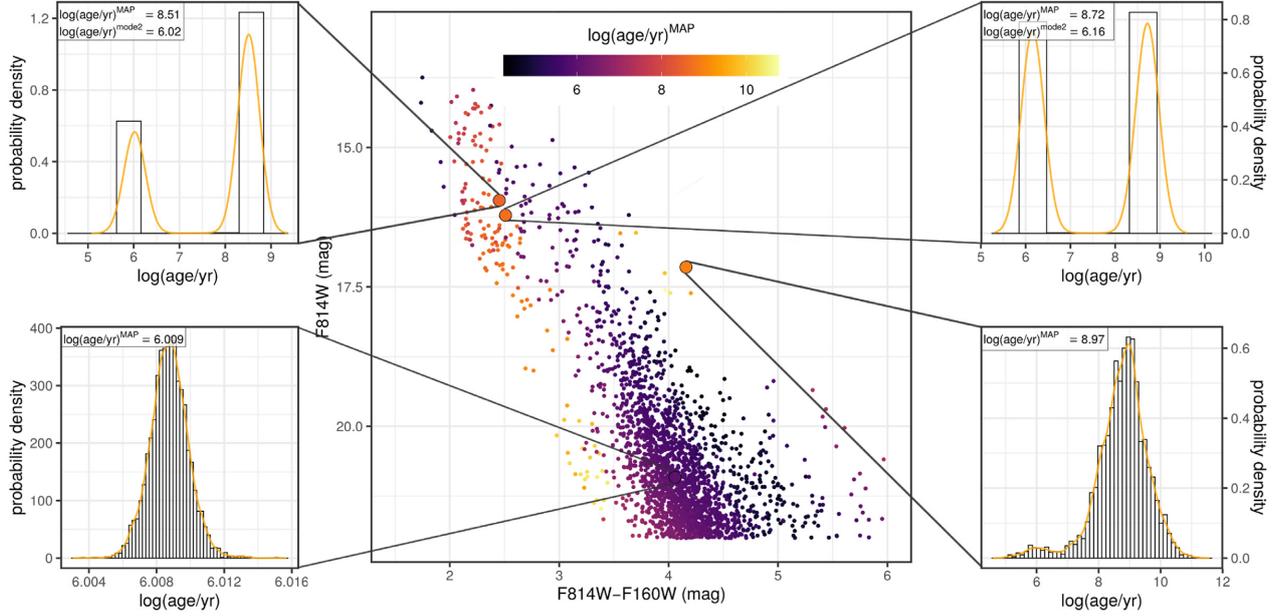
Figure 16. The middle panel shows a zoom-in of the optical CMD of the Wd2 cluster constituents colour coded according to the MAP prediction of log (age). The four panels in the left and right show the predicted age posterior distributions of the highlighted stars in the CMD. The bottom left panel is an example PMS star for which our approach provides excellent results, returning a very narrow age distribution at the proposed cluster age. The remaining three cases are taken from the 86 stars likely on the turn-on for which the MAP age estimate is significantly above the suggested age of Wd2. The two posterior distributions on the top show commonly observed behaviour among these 86 stars, where we find a second peak in the age posterior distribution, either less (right-hand panel) or almost equally likely (left-hand panel), which is more consistent with Westerlund's suggested age. The bottom right panel is a rare example where the age posterior distribution shows no significant second peak and the age of the star is predicted to be too old for Wd2.
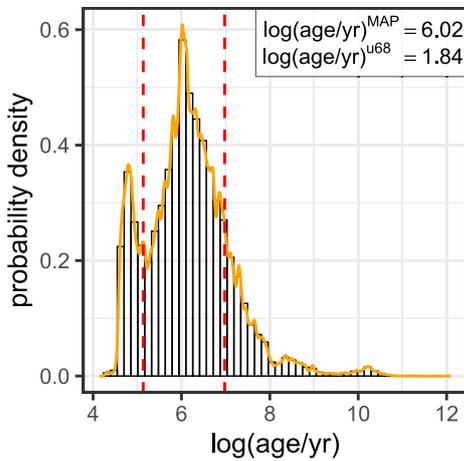
Figure 17. Histogram of the sum of the age posterior distributions of all Wd2 cluster stars as predicted by the 'Wd2_I' cINN model. The orange line indicates a kernel density fit to this 'cumulative' posterior distribution to determine the most likely cluster age. The red dashed lines mark the width of the 68 per cent confidence interval.

deviation):

$$\Gamma_{\mathrm{MST}}^{\mathrm{upper}} = \frac{1}{k} \sum_{i=1}^{k} \frac{e_{\mathrm{rand}}^i}{e_{\mathrm{mass}}} \Delta e_{\mathrm{rand}}^i,$$

$$\Gamma_{\mathrm{MST}}^{\mathrm{lower}} = \frac{1}{k} \sum_{i=1}^{k} \frac{e_{\mathrm{rand}}^i}{e_{\mathrm{mass}}} \frac{1}{\Delta e_{\mathrm{rand}}^i}. \tag{22}$$
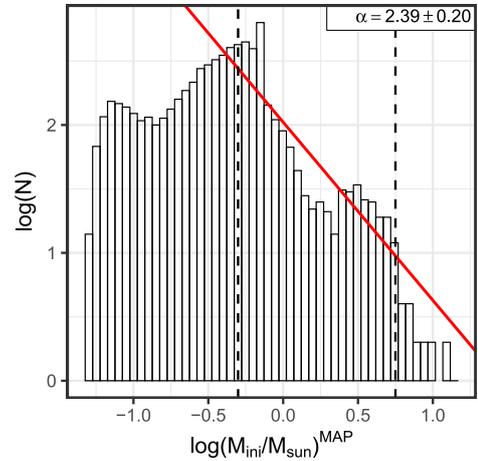
Figure 18. The predicted IMF of Wd2 based on the MAP estimates of M_{ini} of the individual stars provided by the 'Wd2_I' cINN model. The black dashed lines indicate the range (0.5 to $\sim 5.6\,\mathrm{M_\odot}$) used to fit the high-mass slope of the IMF. The fit with a slope of $\alpha = 2.39 \pm 0.2$ is given by the red line.

Values of $\Lambda_{\mathrm{MST}} \approx 1$ and $\Gamma_{\mathrm{MST}} \approx 1$ indicate that the N massive and the N randomly selected stars are similarly distributed, while $\Lambda_{\mathrm{MST}} \gg 1$ ($\Gamma_{\mathrm{MST}} \gg 1$) signifies mass segregation and $\Lambda_{\mathrm{MST}} \ll 1$ ($\Gamma_{\mathrm{MST}} \ll 1$) suggests inverse mass segregation where the most massive stars are more spread outwards (Dib, Schmeja & Parker 2018). Following the suggestion in Olczak et al. (2011), we calculate the number k of random population MSTs based on the number N of massive stars, such that a fraction $p = 0.99$ of the total population of M stars is
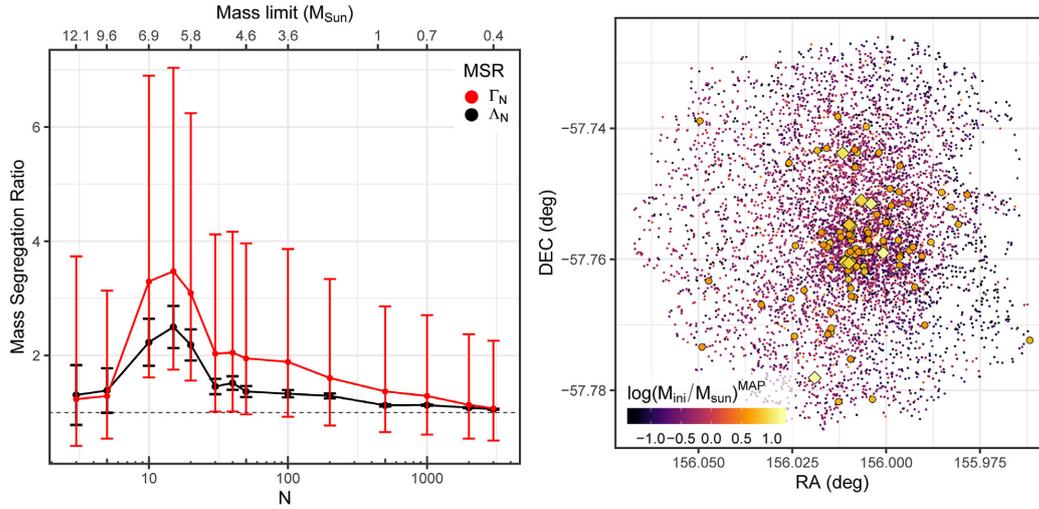
Figure 19. The left-hand panel shows the MSRs Λ_{MSR} (Allison et al. 2009) and Γ_{MSR} (Olczak, Spurzem & Henning 2011) for different numbers of the most massive stars of Wd2 based on the MAP estimates of the initial mass of the stars. Note that the second *x*-axis in this diagram denotes the corresponding lower mass limits, that is, the mass of the lowest mass star within the set of the *N* most massive stars. On the right, the spatial distribution of the Wd2 stars is shown colour coded according to the MAP estimate of the initial mass. The stars highlighted by the large diamond symbols are the 10 most massive stars in our prediction, while the large circles (plus the diamonds) indicate the 100 most massive stars.

covered according to

$$k = \mathrm{ceil}\left(\frac{\ln(1-p)}{\ln\left(1-\frac{N}{M}\right)}\right), \qquad (23)$$

where ceil(x) denotes the ceiling function, that is, the function rounding up to the next larger integer.

In the left-hand panel of Fig. 19, we present our resulting MSRs for different numbers *N* of the most massive stars drawn from the total population. We find some evidence for mass segregation as $\Lambda_{MST} > 1$ and $\Gamma_{MST} > 1$ for the 10–100 most massive stars. With a maximum MSR of ~ 3.4 within this range, however, our analysis suggests that the mass segregation is not strongly pronounced. The spatial distribution of the 10 (diamond markers) and the 100 (large circles + diamond markers) most massive stars shown for comparison in the right-hand panel of Fig. 19 confirms this finding, as the most massive stars appear slightly more clustered towards the centre but not to an excessive degree. The decrease in MSR for the five and three most massive stars is likely due to the fact that the single most massive star ($M_{\mathrm{ini}}^{\mathrm{MAP}} \approx 14.7\,\mathrm{M}_\odot$) in our sample is actually located away from the centre of Wd2 (the southernmost diamond in the diagram), which induces large tree and edge lengths in the MST.

In conclusion, our results for cluster age, slope of the IMF, and observed mass segregation, derived from the cINN predictions of Wd2, are in good accordance with previous studies. Therefore, the cINN method performs to a very satisfactory degree on the actual observational data of Wd2.

5.2 NGC 6397

For NGC 6397, our cINN predictions do not achieve the same success on the real *HST* data as for Wd2. Fig. 20 summarizes our results showing the MAP estimates for the physical parameters colour coded for every star in the UV-I CMD. Overall we find fairly plausible values for all parameters, except for age. For instance, most predicted masses are below one solar mass, which is expected

for a 13 Gyr old cluster given that more massive stars should already have disappeared. With the age prediction, however, we find worse results. A large fraction of stars is predicted to be much younger than what would be reasonable for NGC 6397, considering that some of them are located on the RGB and the main-sequence turn-off, the features traditionally used to date globular clusters. The top left and top centre panels of Fig. 21 show the age prediction more in detail, separating those stars in the CMD for which the MAP estimate is plausible (above 1 Gyr) from those where it is definitely incorrect (below 1 Gyr). Only 1/5 of the stars (999 out of 4831) have plausible MAP age estimates. Of the remaining 3832 stars, only 359 have a second or third mode in their predicted age posterior distributions that falls above 1 Gyr. The top right panel in Fig. 21 shows that most of these are located at the turn-off and bottom of the RGB, an indication that the cINN has learned, at least to some degree, that stars located on the turn-off may be old. But even including these 359 additional stars, where a plausible solution is part of the posterior distribution, we still find that for more than two-thirds of the observational data our age prediction fails entirely. Failure may be expected for some of the stars within the NGC 6397 sample as our training set does not include, for example, white dwarfs, so that a misprediction in these cases is easily explained. If we subtract the latter cases and the 359 turn-on stars with a plausible second mode, we find that the age prediction fails primarily for low-mass main (LMS) sequence stars.

As we have previously discussed, predicting the age of LMS stars is arguably an extremely difficult task as stars with a wide range of ages share very similar observational features. Even though the cINN estimates an age within a plausible range for a number of LMS stars, at least down to about 23 mag in F275W, some of these predictions are still flawed, as can be seen in the histograms in the bottom row of Fig. 21. There are, in particular, cases here where the MAP estimates are too large, sometimes even way above the age of the universe. With only a minority of stars with plausible MAP age estimates, deriving a cluster age as the most likely age provided by the sum of the age posteriors (Fig. 22) is not applicable. The most likely
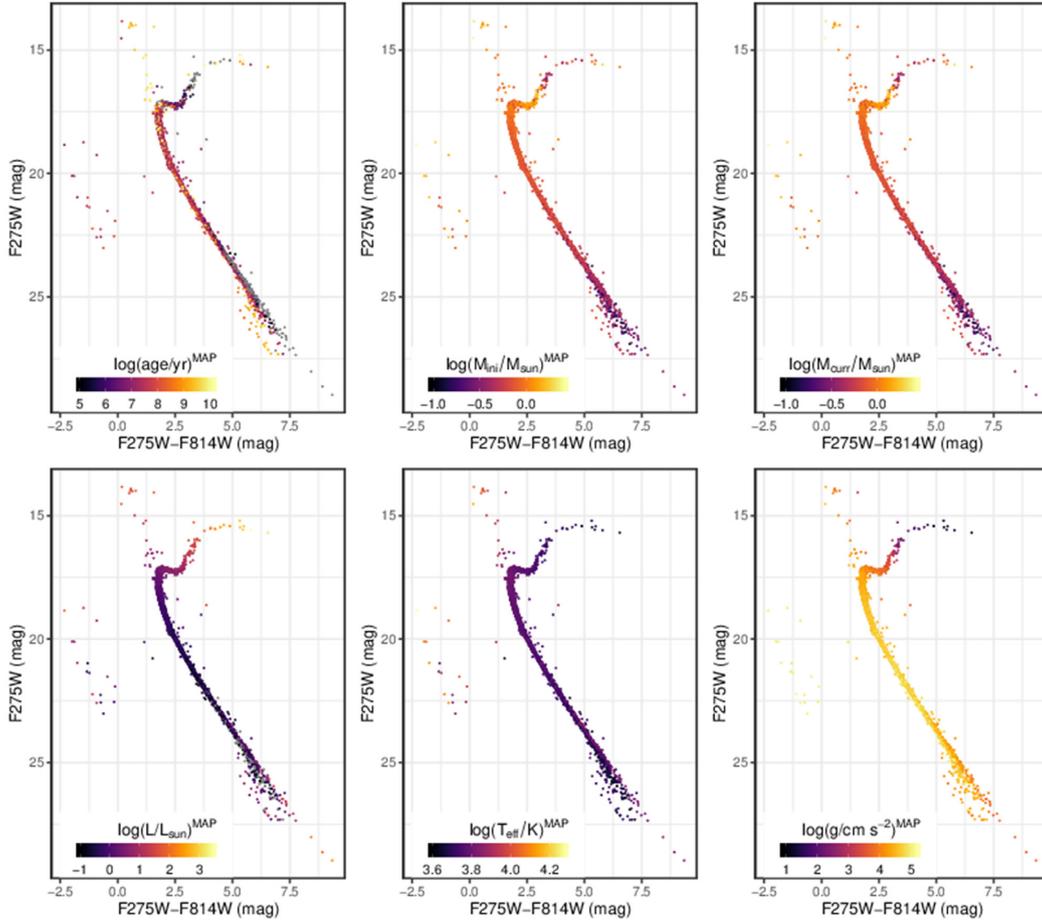
Figure 20. UV-I CMDs of NGC 6397 colour coded according to the MAP estimates of the six physical parameters log (age), log ($M_{\rm ini}$), log ($M_{\rm curr}$), log (L), log ($T_{\rm eff}$), and log (g) as predicted by the cINN trained on 'NGC 6397_I'.

age value would be 23.4 Myr, way too low, and the barely prominent second peak, while in the vicinity of the relatively well known cluster age, still underestimates the age with a value of 7.9 Gyr.

In contrast to the problematic age estimates, luminosity and effective temperature appear to be predicted quite well by the cINN model. The left-hand panel in Fig. 23 shows the predicted HRD for NGC 6397 based on the MAP estimates of log (L) and log ($T_{\rm eff}$). The cINN prediction traces the 13.4 Gyr isochrone (red line) overall very closely, but there are a few outliers, indicated by orange points in the diagram. Among them we find the white dwarfs and blue stragglers as revealed by the CMDs in the remaining panels of Fig. 23. Apart from these cases, there are a couple of LMS stars for which the prediction deviates noticeably from what one would expect, similar to what we found with the predicted ages. This leads us to a possible hypothesis to explain the problems our cINN approach encounters with the low-mass NGC 6397 stars. The main-sequence outliers fall primarily into a region in the observable space where the PARSEC isochrone models cannot properly fit the observed data (note the deviation between the red isochrone and the data points, most obvious in the UV-I CMD in Fig. 23c). Given that the deviation between model and data is most severe starting, for example, at ∼23 mag in F275W in the UV-I CMD, this deficiency of the models could also explain why we find very few plausible age predictions below this magnitude.

The deviation between the model and the data is not only present for the LMS, but also in the RGB, as shown by the CMDs. This could further explain the many age miss-predictions for even the RGB constituents.

If a problem with the underlying stellar evolutionary models is indeed the root of the cINN prediction shortcomings, restricting the training set to a narrower range as in our 'NGC 6397_II' model cannot be a remedy (see Figs E3–E5 in Appendix E). While narrowing the range provides age predictions much closer to the actual age of the cluster, a large number of cases show overprediction (1013 stars have an age MAP estimate above 13.5 Gyr) as well as several instances of extrapolation far below the minimum age of the 'NGC 6397_II' training set (1751 stars with age$^{\rm MAP}$ < 1 Gyr, going down to 0.1 Myr and below). The CMD positions of the latter outliers (Fig. 24) provides further support for the hypothesis that the discrepancy between model and observations causes the prediction issues: the cINN underestimates the age predominantly for stars located where the observed LMS population deviates the most from the theoretical data. For further evidence on this matter and additional details on the 'NGC 6397_II' results we refer to Appendix E.

Returning to our discussion in Section 4.4 concerning the age estimation for main-sequence sources from photometry alone, it is necessary to mention here that the complexity of the task itself
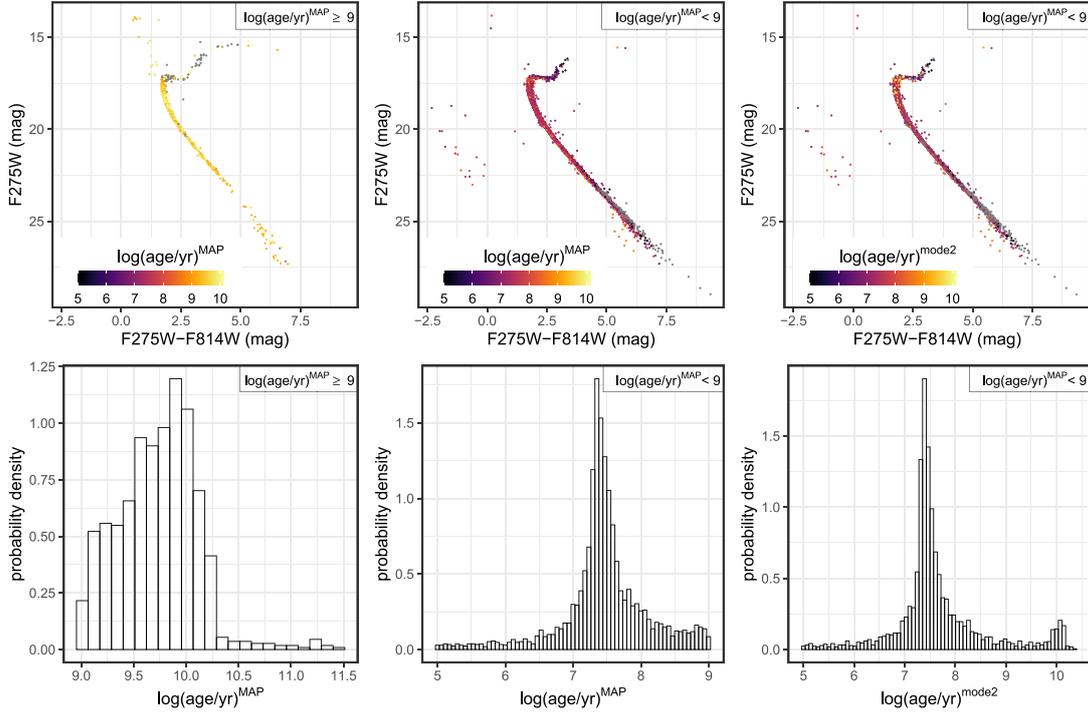
Figure 21. The top left panel shows the CMD of the 999 NGC 6397 stars for which the MAP estimate of the age, derived from the 'NGC 6397_I' cINN prediction, is above 1 Gyr as indicated by the colour coding. The top middle panel shows the CMD position of the remaining 3832 stars, also colour coded according to log (age/yr)^{MAP}, for which the MAP estimate is below 1 Gyr for comparison. The top right panel shows the CMD of the same stars as the middle panel, now colour coded according to the second most likely mode (the second highest local maximum) of their posterior distribution. Stars that show no second mode in the predicted age posterior are indicated in grey. The bottom row shows the histograms of the age MAP estimates (or second mode in the age posterior in the last case) of the stars in the corresponding panels in the top row.
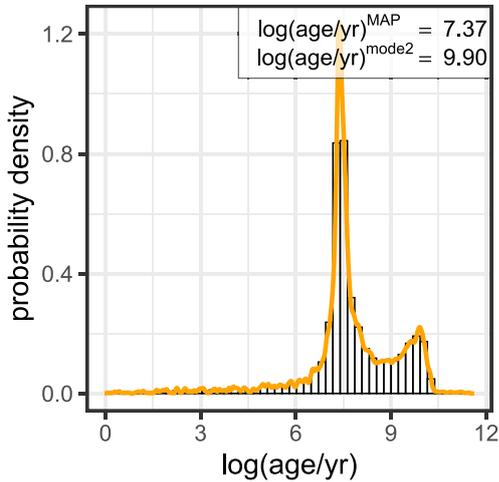
Figure 22. Histogram of the sum of the age posterior distributions of all NGC 6397 stars, predicted by the cINN trained on 'NGC 6397_I'. The orange line indicates a kernel density fit to this cumulative posterior distribution to determine the most likely cluster age.

likely also plays a role in the NGC 6397 prediction outcome. While the evidence for the main culprit being discrepancy between model and observations appears conclusive to us, given also its similarity to the model–model difference issues we have discovered with the

prediction on the low-metallicity Dartmouth isochrone data, we have to acknowledge an additional caveat here. Even where the PARSEC models do match the observations, we find that the age estimates for the LMS constituents are *just* plausible but do not exactly recover the known age of NGC 6397. Based on this consideration and with regard to the difficulty that traditional methods have with dating main-sequence objects, we have to surmise that the cINN might not necessarily outperform known approaches on this specific aspect. Consequently, our age estimates for real main-sequence stars should be treated with caution.

In conclusion our prediction of physical parameters for the globular cluster NGC 6397 does not achieve the same satisfying results as the cINN model does for Wd2. While the predicted masses fall within reasonable ranges and the HRD constructed from the MAP estimates of L and T_{eff} traces the theoretical position of the cluster reasonably well, we find a significant number of outliers and major problems with the prediction of the age. Here, the cINN tends to overestimate the age for the stars it recognizes as old, while severely underestimating the age for a majority of stars of the globular cluster. Comparing the location of these outliers in the CMD with the underlying PARSEC models, we believe that these issues are primarily rooted in a mismatch, contrary to the Wd2 case, bewteen the isochrones and the observations of NGC 6397. We find large discrepancies especially for the LMS and RGB stars (see e.g. Fig. 23). This ultimately demonstrates again that even an ML approach as powerful as the cINN is always only as good as the underlying physical model. Therefore, it is crucial to choose models that provide the best agreement with the data.
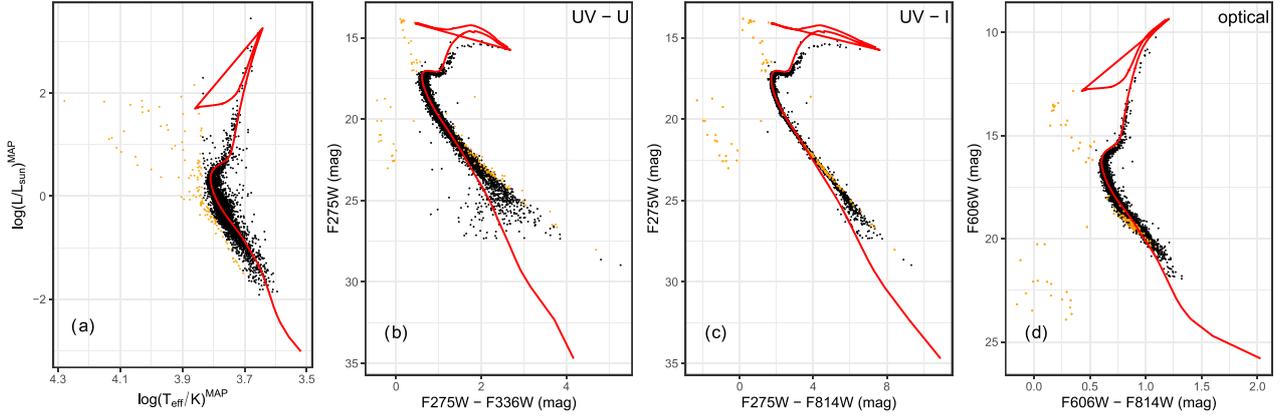
Figure 23. Panel (a) shows the predicted HRD of NGC 6397 based on the MAP estimates of $\log(L)$ and $\log(T_{\mathrm{eff}})$ by the cINN trained on 'NGC 6397_I'. The remaining panels (b)–(d) show in order the UV-V, UV-I, and the optical CMDs of the NGC 6397 data, respectively. The orange marked stars are those for which the cINN prediction of the HRD position deviates strongly from the supposed age of the cluster of ~ 13 Gyr, as indicated by the red isochrone in all four diagrams. This series shows that, aside from the white dwarfs and blue stragglers, the cINN prediction fails where the PARSEC isochrone models fail to fit the data.
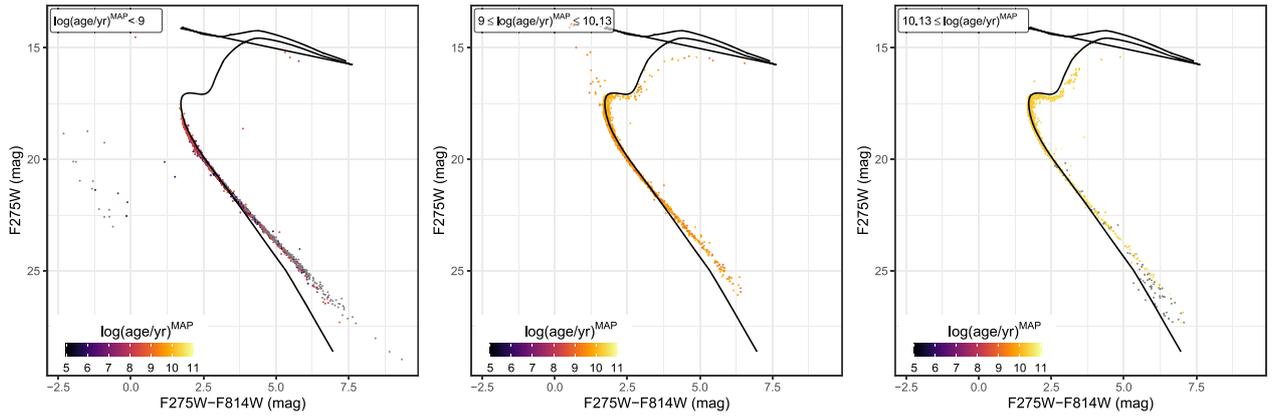
Figure 24. UV-I CMD of the NGC 6397 data colour coded according to the MAP age estimates predicted by the cINN trained on 'NGC 6397_II'. In order the three panels show the stars for which we find $\log(\mathrm{age/yr}) < 9$, $9 \leq \log(\mathrm{age/yr}) \leq 10.13$, and $10.13 < \log(\mathrm{age/yr})$, respectively. The red line in all diagrams indicates a 13 Gyr isochrone for comparison. This sequence demonstrates that the age prediction of the 'NGC 6397_II' fails noticeably for those stars, where the observations deviate the most from the theoretical model.

6 POSSIBLE EXTENSIONS

6.1 Extinction as a physical parameter

As mentioned in the Introduction, we keep the regression problem as simple as possible for this study. Therefore, we adopt a single value for the metallicity of each cluster and assume that individual stellar extinctions are known. Nevertheless, given the way our training sets are constructed we can easily move the extinction from the observable to the physical parameter space. Because of that, and also in view of future development of this method, we perform one 'feasibility' test for both 'Wd2_I' and 'NGC 6397_I' where the cINN trains to predict extinction instead of taking it as an input.

Without further modifications to our approach, the prediction of extinction does not work very well for 'Wd2_I'. Not only does the predicted extinction on the synthetic test set exhibit a large rmse (~ 1.7 mag) for the point estimates, the prediction of the remaining parameters also suffers greatly (e.g. age rmse increases to 1.3 dex).

In the 'NGC 6397_I' case, however, we find that the cINN can easily predict the extinction. Here, the rmse of the point estimates is only 0.008 mag (first panel in Fig. 25), and there is no significant degradation of the predictive capabilities for the other parameters. Part of the failure of the Wd2 model can likely be attributed to the much larger extinction range, 12 mag, adopted for 'Wd2_I' training set in comparison to the only 3 mag range for the 'NGC 6397_I' case. Another possibility, however, could be that the 'Wd2_I' training set only uses two photometric filters, too few to properly constrain the extinction. A more optimized architecture of the cINN may be required in these cases.

Given this promising outcome for the extinction prediction of the 'NGC 6397_I' cINN model on the synthetic data, we further evaluate its performance by predicting the *known* extinction of the real NGC 6397 data. The results are shown in panels 2–4 of Fig. 25. The histogram of the MAP estimates in panel 2 shows that the prediction for A_V is fairly accurate, the mean being even within the narrow 3σ range determined by Brown et al. (2018). There are a
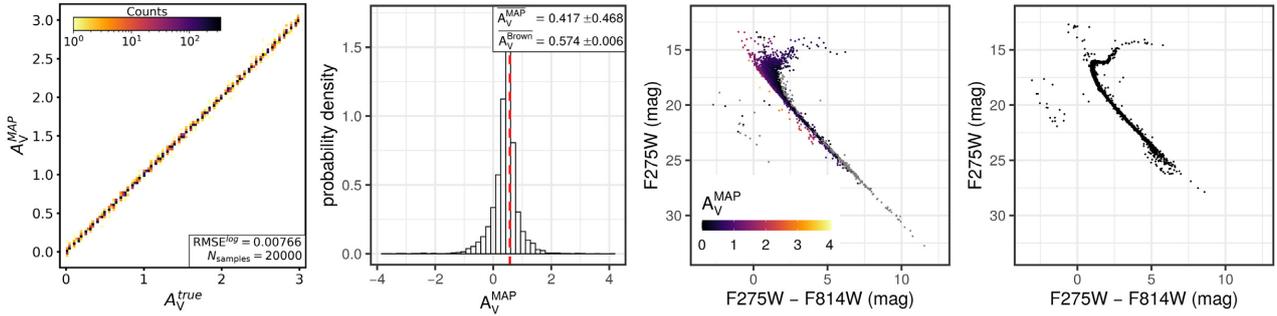
Figure 25. The first diagram shows the MAP estimates of extinction against the true values for 20 000 test observations as predicted by a cINN trained on 'NGC 6397_I'. The second panel shows a histogram of the predicted A_V MAP estimates of the NGC 6397 *HST* data in comparison to the known mean extinction (red dashed line, Brown et al. 2018). The third and fourth panels show extinction corrected UV-I CMDs of the NGC 6397 *HST* data, once corrected via the MAP extinction estimates and once by sampling from $A_V = 0.5735 \pm 0.0062$ (as suggested by Brown et al. 2018). In the third panel, the stars are additionally colour coded according to A_V^{MAP} and the grey points indicate stars for which the MAP estimate is an unphysical value below 0.

few cases where the cINN predicts an unphysical negative extinction value. As we have not enforced the extinction value to be positive during training, by, for example, taking its logarithm, this could easily be remedied in further optimization. The third and fourth panels show the predicted extinction corrected CMD of NGC 6397 (colour coded according to A_V^{MAP}) and the extinction corrected CMD retrieved by randomly sampling the extinction values of Brown et al. (2018). The overall shape matches fairly well, especially the turn-off points are in good agreement. The cases where the cINN prediction fails (grey points in the third panel) are mostly related to the issue with the PARSEC isochrones, that we have discussed in Section 5.2.

Nevertheless, this test shows that extinction prediction is very well within the capability of the cINN method, at least if enough photometric filters are available as features, the considered extinction range is small enough, or a combination of both.

6.2 Other

Beside the prediction of extinction there are several physical effects, such as variability or photometric uncertainties, that provide room for extensions to our approach. While the latter can be taken into account using a weighted sampling strategy within the uncertainties of the observational data, developing an intrinsic uncertainty propagation mechanism would be a powerful extension to the approach.

In this paper, we have also presented one approach to incorporating prior knowledge (specifically metallicity and age) into our method by curating training sets accordingly. Taking the young cluster Wd2 as an example, that is, comparing the results of 'Wd2_I' and 'Wd2_II', this procedure does not necessarily change the cINN outcome. The alternative of treating prior knowledge intrinsically rather than through training set modification is another possible extension that could benefit the approach. A possible way to do so could be through modification of the target distribution of one or more latent variables. Within the approach we have presented, using Gaussian distributions as targets is the simplest choice, but an arbitrary one. In principle any distribution (for which a log-likelihood can be defined) can serve as the target during training of the cINN. Therefore, we plan to investigate if setting prior knowledge, for example, a distribution of plausible ages for a given cluster, as the target distribution of one of the latent variables will have the desired effect of incorporating additional prior information more effectively.

7 SUMMARY AND CONCLUSIONS

In this introductory paper, we present the first application of a novel INN approach to the task of predicting physical parameters for individual stars based on photometric observations. In many such inverse regression problems, the mapping from the physical parameters of interest x to the associated observables y is subject to an inherent information loss that induces degeneracies as y no longer captures all variance of x. To retain this information otherwise lost, the cINN encodes all variance of x that is not covered by y in latent (not observable) variables z by learning a mapping from x to z conditioned on y. Due to the invertible architecture of this network, after learning this forward mapping it automatically provides a solution for the inverse mapping $x = g(z; y)$, and by sampling the latent variables z one obtains estimates for the full posterior distributions $p(x|y)$ of interest.

We introduce cINNs to the analysis of photometric data in this pilot study by training and testing on synthetic data from the PARSEC stellar evolutionary models (Bressan et al. 2012) and performing a benchmark analysis on real observational data obtained by the *HST* for the young cluster Wd2 and the old globular cluster NGC 6397. These clusters are chosen to cover the extremes of the cluster range, that is, very young and very old, in order to gain first insights into the systematics of our approach, but not to conduct an exhaustive analysis of the whole spectrum of possible cluster parameters. We construct the synthetic training sets by adopting isochrone model tables of the correct metallicity for Wd2 and NGC 6397, respectively, with the aim to predict age, initial and current mass, luminosity, effective temperature and surface gravity of each cluster star. To overcome sampling issues in the M_{ini} and age spaces of the isochrone tables, we first oversample each individual isochrone using a spline interpolation and then extract sample points to fill up underpopulated areas within the parameter space. To simplify this regression problem we use extinction as an observable parameter, in addition to the available photometry. To account for extinction within the synthetic training set, assuming every isochrone model point to be a synthetic star, we add multiple examples of the same star with different amounts of extinction to the training sets.

In order to evaluate how the cINN prediction on the real data behaves when we include prior knowledge about the age of the respective clusters, we construct two training sets for each cluster, one encompassing the entire age range of our theoretical models, from $\log(\text{age/yr}) = 5$ to 10.13, the other with a reduced age range

close to the actual cluster age ($\log\,(\text{age/yr}) = 5$ to 8 for Wd2 and 9 to 10.13 for NGC 6397). To derive point estimates from the predicted posterior distributions, we use kernel density estimation with bandwidths determined according to Silverman's rule of thumb. In this way we find the most likely values of physical parameters (MAP estimates) in the marginalized distributions. We ascertain the training performance of our four models on a test set of 20.000 random synthetic observations that are excluded for the training process. Using this sample, we determine for each parameter the median calibration error, the uncertainty at 68 per cent confidence as well as the rmse/nrmse between the MAP parameter estimates and the known true parameter values. Using a simple nearest neighbour approach on the training data, we also approximate a resimulation error for the predicted posteriors on the synthetic test set. Furthermore, to ascertain how well our models generalize to new populations, we test our models on synthetic data from two different stellar evolution models, namely isochrone tables from MIST (Choi et al. 2016; Dotter 2016) and Dartmouth (Dotter et al. 2008).

Our main results from the tests on *synthetic data* are the following:

(i) Once trained ($\sim 2\,\text{h}$) the cINN can rapidly predict a posterior distribution for a single star. Using GPU acceleration on a Nvidia GTX 1080, the cINN can predict about 35 posterior distributions with 4096 samples per second.

(ii) On the synthetic test data, the prediction of initial/current mass, luminosity, effective temperature, and surface gravity works extraordinarily well with posterior distributions that are narrowly constrained around the true values and low rmses of the derived MAP parameter estimates.

(iii) Predicting the stellar age is a more difficult task. The predicted posteriors tend to be broader and often exhibit multimodalities, revealing ample degeneracies in the age prediction. While we can confirm that the true value is part of the predicted distribution in more than 99 per cent of the cases, there are several instances where the true solution does not coincide with the most likely outcome of the posterior, falling into a second peak instead. In itself, this is not problematic as a true posterior describes all possible parameters that could explain a given observation, such that the most likely prediction does not have to be the one that generated the given observation. However, as a consequence we find significantly more cases in the age prediction where our point estimates deviate from the true value. The intermediate age range from $\log\,(\text{age/yr}) = 6.5$ to 8.5 is the least affected, whereas the predictions for very old stars ($> 1\,\text{Gyr}$) show a notable amount of instances where the MAP estimate is off by ~ 0.4 dex on average. When fewer photometric filters (2 instead of 5) are available, as in the case of our Wd2 training set, we also observe more deviations for the very young stars ($< 10\,\text{Myr}$). Nevertheless, overall these cases where the MAP estimate deviates strongly from the true value are still a *minority*.

(iv) Our nearest neighbour resimulation approximation returns small errors, confirming the validity of the predicted posterior distributions and identified degeneracies.

(v) The predictive performance of the cINN improves (especially for the age) when more photometric filters are included in the observables. However, even with perfect information (17 photometric filters here) the prediction of age for old stars (1 Gyr and older) still remains highly challenging.

(vi) Our models generalize overall very well to the synthetic data of other stellar evolution models and perform particularly well in recovering luminosity, surface temperature, and gravity. Ages and masses are also predicted fairly accurately for most samples, but tend to exhibit larger errors. Specifically, our models manage to also recover ages of main-sequence objects on these different synthetic test sets, providing confirmation that our models do not simply overfit their training data for these hard to predict cases. However, if there are significant discrepancies between the investigated models and the PARSEC isochrones, such as in our low-etallicity Dartmouth test, the age prediction can fail severely. We conclude that our cINN model manages to recover ages for synthetic main-sequence stars through a combination of its latent variable approach and the perfect synthetic photometry. For real main-sequence observations, however, we suggest to treat the predicted ages with caution.

Applied to *observed data* of Wd2 and NGC 6397, we find:

(vii) The cINN predictions based on the *HST* data of Wd2 return excellent results. With a median of $1.27^{+3.62}_{-0.94}$ Myr of the age MAP estimates and a most likely value of $1.04^{+8.48}_{-0.90}$ Myr from the sum of all age posteriors, the cINN results are in good accordance with the previously determined age of Wd2 of 1.04 ± 0.72 Myr (Zeidler et al. 2016). Furthermore, the cINN correctly recognizes that stars located on the turn-on could potentially also be RGB stars and thus returns multimodal age posterior distributions, highlighting this degeneracy.

(viii) Based on the cINN mass estimates, we are able to construct the IMF of Wd2 and fit its high-mass slope. We find a value of $\alpha = 2.39 \pm 0.2$, which corresponds to the Salpeter slope within 1σ and is in accordance with the previously determined slope of the PDMF of $\alpha = 2.53 \pm 0.05$ (Zeidler et al. 2017). We also find evidence for mass segregation based on the MSRs Λ_{MSR} and Γ_{MSR} of individual stars, again confirming previous results by Zeidler et al. (2017).

(ix) For NGC 6397, the cINN predictions are not as good as for Wd2. While certain properties are recovered, such as the predicted HRD which traces the isochrone corresponding to the known clusters age relatively well, there are glaring issues with the prediction of the cluster age. The majority of stars is predicted to be much younger than the actual cluster age, while stars that are correctly identified as old tend to have an overestimated age. We identify the culprit for these unsatisfying results in the PARSEC evolutionary models, as they do not fit the observations of this globular cluster well enough. This example highlights that the careful selection of the underlying physical model is of utmost importance for our approach.

(x) When enough photometric filters are available (e.g. 5 in the case of NGC 6397) the cINN can also predict extinction very well, instead of using it as an input observable, without losing accuracy in the prediction of the other physical parameters.

Overall the results presented in this paper demonstrate that the cINN is a very powerful approach that can solve the problem of predicting physical parameters from photometry data if the underlying physical models are selected carefully to match the observations. In other words, the possibility of solving the inverse problems (from observations to the physical parameters of each star) relies heavily on the quality of the forward modelling (from physical parameters to synthetic observations). In the case of Wd2, we correctly recover the main cluster properties using *only 2* photometric filters and an estimate of stellar extinction as an input. The cINN method can successfully learn and highlight degeneracies that appear within the given problem, making it an excellent tool for tasks that are subject to degenerate mappings from physical to observable parameter space.

Given its excellent prediction efficiency, we believe that the cINN approach could become a very valuable tool in the big data epoch of astronomy. In particular current and future all-sky/very-wide-field surveys like Pan-Starrs as well as upcoming observational facilities such as the Vera Rubin Observatory (Large Synoptic Survey

Telescope, LSST) or the *Roman Space Telescope* (formerly *WFIRST*) will provide enormous amounts of data, for which efficient and robust deep learning approaches, such as the cINN, will truly be able to show their strength. With this in mind we plan to employ the cINN approach on data from large *HST* surveys, such as the Hubble Tarantula Treasury Project or the Measuring Young Stars in Space and Time survey, to characterize more complex stellar populations in a subsequent study.

In this paper, we purposefully keep the regression problem as simple as possible, in particular we only consider the single metallicity case, using extinction as an input parameter and ignore photometric errors. We plan to address these effects, together with variability and binarity, in future studies. As demonstrated for the example of the old globular cluster NGC 6397, predicting extinction is already well within the capacity of the cINN. However, it might require some architecture optimization to support observations with a low number of filters or regions with a large range of differential extinction. Photometric errors can be taken into account to some degree at this stage already, at least at the prediction stage, by simply resampling the observations according to their errors and performing a weighted addition of the resulting posterior distributions. In the future, we also plan to investigate if some intrinsic treatment is possible at the training stage of the network, for example, by incorporating uncertainties in the training set. Another avenue that we aim to pursue is the possibility of considering prior knowledge as part of the training strategy rather than incorporating it in the training set. Ultimately our goal is to provide observers with a robust, efficient and general tool to analyse observations and retrieve the key physical parameters of their targets.

DATA AVAILABILITY

The HUGS data of N6397 used in this article are available in the Mikulski Archive for Space Telescopes, at http://dx.doi.org/10.17909/T9810F.

The *HST* data of Wd2 underlying this article were provided by Elena Sabbi by permission. Data will be shared on request to the corresponding author with permission of Elena Sabbi.

The data outcomes underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Allison R. J., Goodwin S. P., Parker R. J., Portegies Zwart S. F., de Grijs R., Kouwenhoven M. B. N., 2009, MNRAS, 395, 1449
Anthony-Twarog B. J., Twarog B. A., Suntzeff N. B., 1992, AJ, 103, 1264
Ardizzone L., Kruse J., Rother C., K"othe U., 2019a, in International Conference on Learning Representations. Available at https://openreview.net/forum?id=rJed6j0cKX
Ardizzone L., Lüth C., Kruse J., Rother C., Köthe U., 2019b, preprint (arXiv:abs/1907.02392)
Ascenso J., Alves J., Beletsky Y., Lago M. T. V. T., 2007, A&A, 466, 137
Bellinger E. P., Angelou G. C., Hekker S., Basu S., Ball W. H., Guggenberger E., 2016, ApJ, 830, 31
Bressan A., Marigo P., Girardi L., Salasnich B., Dal Cero C., Rubele S., Nanni A., 2012, MNRAS, 427, 127
Brown T. M., Casertano S., Strader J., Riess A., VandenBerg D. A., Soderblom D. R., Kalirai J., Salinas R., 2018, ApJ, 856, L6
Cantat-Gaudin T. et al., 2020, A&A, 640, A1
Cardelli J. A., Clayton G. C., Mathis J. S., 1989, ApJ, 345, 245
Carraro G., Turner D., Majaess D., Baume G., 2013, A&A, 555, A50
Chabrier G., 2002, ApJ, 567, 304
Chen Y., Girardi L., Bressan A., Marigo P., Barbieri M., Kong X., 2014, MNRAS, 444, 2525
Chen Y., Bressan A., Girardi L., Marigo P., Kong X., Lanza A., 2015, MNRAS, 452, 1068
Choi J., Dotter A., Conroy C., Cantiello M., Paxton B., Johnson B. D., 2016, ApJ, 823, 102
Da Rio N., Gouliermis D. A., Gennaro M., 2010, ApJ, 723, 166
Dib S., Schmeja S., Parker R. J., 2018, MNRAS, 473, 849
Dinh L., Sohl-Dickstein J., Bengio S., 2016, preprint (arXiv:1605.08803)
Dotter A., 2016, ApJS, 222, 8
Dotter A., Chaboyer B., Jevremović D., Baron E., Ferguson J. W., Sarajedini A., Anderson J., 2007, AJ, 134, 376
Dotter A., Chaboyer B., Jevremović D., Kostov V., Baron E., Ferguson J. W., 2008, ApJS, 178, 89
Feigelson E. D., Babu G. J., 2012, Modern Statistical Methods for Astronomy: With R Applications. Cambridge University Press, Cambridge
Goodfellow I. J., Bengio Y., Courville A., 2016, Deep Learning. MIT Press, Cambridge, MA, USA
Gratton R. G., Bragaglia A., Carretta E., Clementini G., Desidera S., Grundahl F., Lucatello S., 2003, A&A, 408, 529
Hastie T., Tibshirani R., Friedman J., 2009, The Elements of Statistical Learning: Data Mining, Inference and Prediction, 2 edn. Springer, Berlin, http://www-stat.stanford.edu/ tibs/ElemStatLearn/
Hyvärinen A., Oja E., 2000, Neural Netw., 13, 411
Ivezic Z., Connolly A. J., VanderPlas J. T., Gray A., 2014, Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data. Princeton Univ. Press, Princeton, NJ
Jackson R. J., Deliyannis C. P., Jeffries R. D., 2018, MNRAS, 476, 3245
Jørgensen B. R., Lindegren L., 2005, A&A, 436, 127
Kingma D. P., Dhariwal P., 2018, preprint (arXiv:1807.03039)
Kounkel M., Covey K., Stassun K. G., 2020, preprint (arXiv:2004.07261)
Kraft R. P., Ivans I. I., 2003, PASP, 115, 143
Miller A. A. et al., 2015, ApJ, 798, 122
Nardiello D. et al., 2018, MNRAS, 481, 3382
Olczak C., Spurzem R., Henning T., 2011, A&A, 532, A119
Olney R. et al., 2020, AJ, 159, 182
Paszke A. et al., 2019, in Wallach H., Larochelle H., Beygelzimer A., d'Alché-Buc F., Garnett R., eds, Advances in Neural Information Processing Systems 32 (NIPS 2019). Curran Associates, Inc., Red Hook, NY, p. 8024
Paxton B., Bildsten L., Dotter A., Herwig F., Lesaffre P., Timmes F., 2011, ApJS, 192, 3
Paxton B. et al., 2013, ApJS, 208, 4

Paxton B. et al., 2015, ApJS, 220, 15

Piotto G. et al., 2015, AJ, 149, 91

Sabbi E. et al., 2020, ApJ, 891, 182

Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525

Sharma K., Kembhavi A., Kembhavi A., Sivarani T., Abraham S., Vaghmare K., 2020, MNRAS, 491, 2280

Silverman B. W., 1986, Density Estimation for Statistics and Data Analysis. Chapman and Hall, London

Tang J., Bressan A., Rosenfield P., Slemer A., Marigo P., Girardi L., Bianchi L., 2014, MNRAS, 445, 4287

Valls-Gabaud D., 2014, in Lebreton Y., Valls-Gabaud D., Charbonnel C., eds, EAS Publ. Ser., The Ages of Stars. EDP Sciences, Les Ulis Cedex A, France, p. 225

Vargas Álvarez C. A., Kobulnicky H. A., Bradley D. R., Kannappan S. J., Norris M. A., Cool R. J., Miller B. P., 2013, AJ, 145, 125

Vulic N., Barmby P., Gallagher S. C., 2018, MNRAS, 473, 4900

Zeidler P. et al., 2015, AJ, 150, 78

Zeidler P. et al., 2017, in Charbonnel C., Nota A., eds, IAU Symp. Vol. 316, Formation, Evolution, and Survival of Massive Star Clusters. Cambridge Univ. Press, Cambridge, p. 55

Zeidler P., Grebel E. K., Nota A., Sabbi E., Pasquali A., Tosi M., Bonanos A. Z., Christian C., 2016, AJ, 152, 84

APPENDIX A: GENERAL

Table A1 summarizes the A_λ/A_V values we have derived according to equation (3) for all of the *HST* filters following the Cardelli et al. (1989) Milky Way extinction curve. The wavelength dependent coefficients $a_\lambda(x)$ and $b_\lambda(x)$, where $x = \lambda^{-1}(\mu m^{-1})$, are given by equations (2)–(4) in Cardelli et al. (1989) and are defined as follows. For the infrared regime $0.3\,\mu m^{-1} \leq x \leq 1.1\mu m^{-1}$, they are given by

$$a(x) = 0.574x^{1.61},$$
$$b(x) = -0.527x^{1.61}. \quad (A1)$$

In the optical and NIR regime, $1.1\,\mu m^{-1} \leq x \leq 3.3\mu m^{-1}$, follows

$$a(x) = 1 + 0.17699y - 0.50447y^2 - 0.02427y^3 + 0.72085y^4$$
$$+ 0.01979y^5 - 0.77530y^6 + 0.32999y^7,$$
$$b(x) = 1.41338y + 2.28305y^2 + 1.07233y^3 - 5.38434y^4$$
$$- 0.62251y^5 + 5.30260y^6 - 2.09002y^7, \quad (A2)$$

where $y = x - 1.82$.

Table A1. Overview of the A_λ/A_V values derived from Cardelli et al. (1989) extinction curve according to equation (3) for the *HST* filters used in the Wd2 and NGC 6397 observations.

	Filter	$\frac{A_\lambda}{A_V}$
$R_V = 3.1$	F275W$_{WFC3}$	1.94436
	F336W$_{WFC3}$	1.65798
	F438W$_{WFC3}$	1.33088
	F606W$_{ACS}$	0.92246
	F814W$_{ACS}$	0.60593
$R_V = 3.8$	F218W$_{WFC3}$	2.53769
	F225W$_{WFC3}$	2.21539
	F275W$_{WFC3}$	1.75064
	F336W$_{WFC3}$	1.49531
	F390W$_{WFC3}$	1.39453
	F438W$_{WFC3}$	1.28651
	F475W$_{WFC3}$	1.15971
	F555W$_{WFC3}$	1.03555
	F606W$_{WFC3}$	0.93420
	F625W$_{WFC3}$	0.88084
	F775W$_{WFC3}$	0.67977
	F814W$_{WFC3}$	0.62821
	F105W$_{WFC3}$	0.39924
	F110W$_{WFC3}$	0.34595
	F125W$_{WFC3}$	0.30448
	F140W$_{WFC3}$	0.25550
	F160W$_{WFC3}$	0.21792

Finally, for the UV regime, $3.3\,\mu m^{-1} \leq x \leq 8\mu m^{-1}$, they are defined as

$$a(x) = 1.752 - 0.316x - \frac{0.104}{(x - 4.67)^2 + 0.341} + F_a(x),$$
$$b(x) = -3.090 + 1.825x + \frac{1.206}{(x - 4.62)^2 + 0.263} + F_b(x), \quad (A3)$$

where $F_a(x) = F_b(x) = 0$ for $x < 5.9$ and

$$F_a(x) = -0.04473(x - 5.9)^2 - 0.009779(x - 5.9)^3,$$
$$F_b(x) = 0.2130(x - 5.9)^2 + 0.1207(x - 5.9)^3 \quad (A4)$$

for $5.9 \leq x \leq 8$.

Fig. A1 shows the prior distributions of all six phyiscal parameters, that is, age, initial M_{ini} and current mass M_{curr}, luminosity L, surface temperature T_{eff}, surface gravity g, for the four training sets 'Wd2_I', 'Wd2_II', 'NGC 6397_I', and 'NGC 6397_II' that we employ in this study.
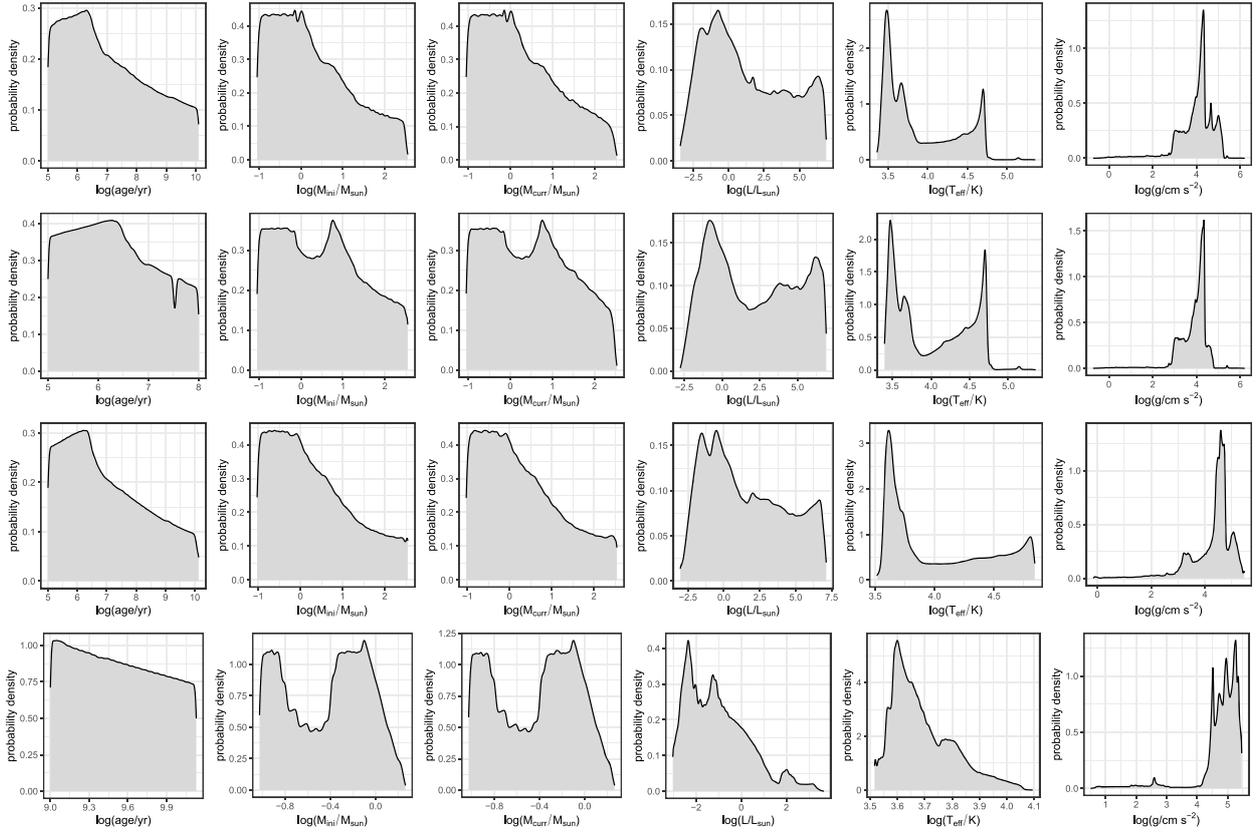
Figure A1. Prior distributions of all physical parameters in our training sets. From top to bottom, the training sets are 'Wd2_I', 'Wd2_II', 'NGC 6397_I', and 'NGC 6397_II', respectively.

APPENDIX B: 'WD2_I'

This appendix provides additional result diagrams for cINN models trained on 'Wd2_I' and variations thereof.

Fig. B1 shows the covariance matrix of the latent variables in the left-hand panel and their corresponding histograms in comparison to the target normal distribution in the right-hand panel as evaluated on the 20 000 test observations using the trained 'Wd2_I' model. These two diagrams serve as an example for convergence of the cINN model.

Corresponding to the series of posterior against true value diagrams presented in Fig. 11 the respective MAP versus true diagrams are shown in Fig. B2. Both these figures demonstrate how the cINN predictive performance improves with an increase of the number of photometric filters used as input.

Fig. B3 shows the 'Wd2_I' cINN prediction results on 10 000 samples from synthetic data of the Dartmouth isochrone tables. This figure highlights that the cINN manages to recover most physical parameters quite accurately on this synthetic data set derived from a stellar evolution model that treats the underlying physics different than the PARSEC models, on which our cINN training sets are based. There are, however, some larger discrepancies in the low-mass regime, where the PARSEC and Dartmouth models deviate most strongly from each other. While the age predictions appear significantly worse than on the MIST synthetic data at first glance, the median absolute error between prediction and ground truth is only 0.2 dex. Additionally, taking multimodalities in the predicted age posteriors into account, most ages can actually be recovered quite accurately. The latter is highlighted in Fig. B4 comparing the predicted to the ground truth HRD for the MIST isochrones.
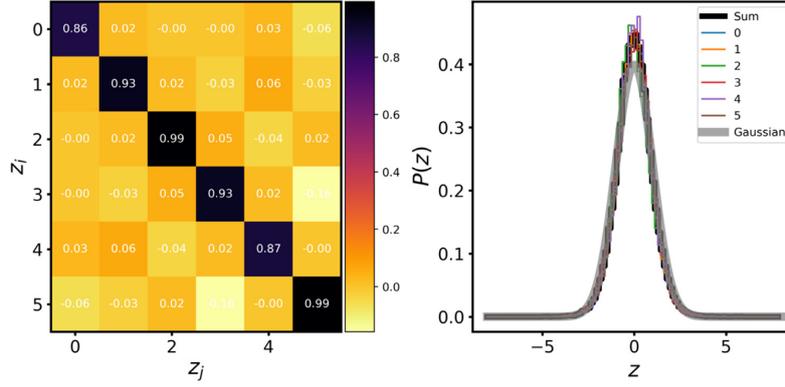
Figure B1. Left: the covariance matrix of the latent variables evaluated on 20 000 test observations provided by the cINN model trained on 'Wd2_I'. Right: histograms of the individual latent variable distributions. The black line indicates the distribution of the sum of all latent variables, while the grey line shows the target normal distribution for reference.
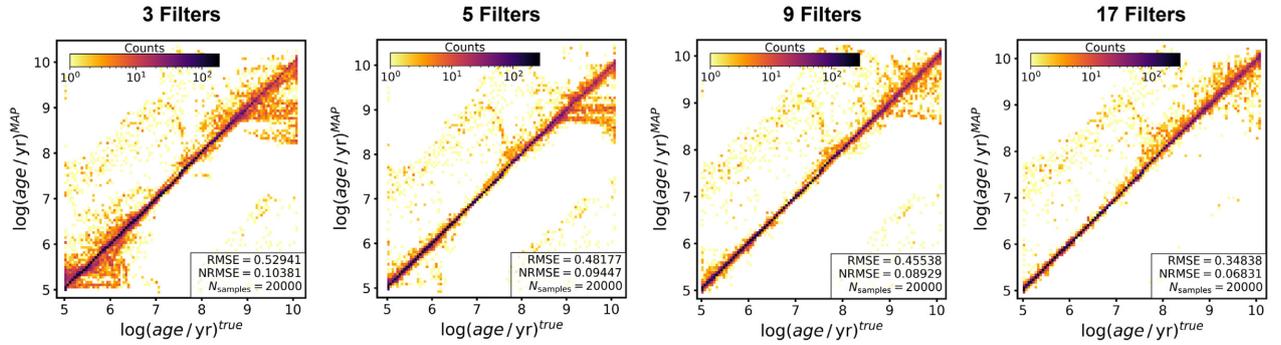
Figure B2. 2D histograms of the MAP estimates of the age against the true value for 20 000 observations from the test set as predicted by the 'Wd2_I' cINN model with increased amounts of features compared to our standard 'Wd2_I' setup. On top of the extinction, the columns indicate increasing numbers of photometric filters used as features. The three filter case entails the *HST* filters F555W, F814W, F160W, the five filter one adds F275W and F336W to that, and the nine filter further includes F438W, F606W, F775W, and F110W on top of the previous. The final 17 filter case entails all previous filters in addition to F218W, F225W, F390W, F475W, F625W, F105W, F125W, and F140W. This sequence shows that the point estimate accuracy of the cINN improves with increasing number of available features as the rmse decreases as well as the number of predictions that fall off the perfect 1-to-1 correlation.
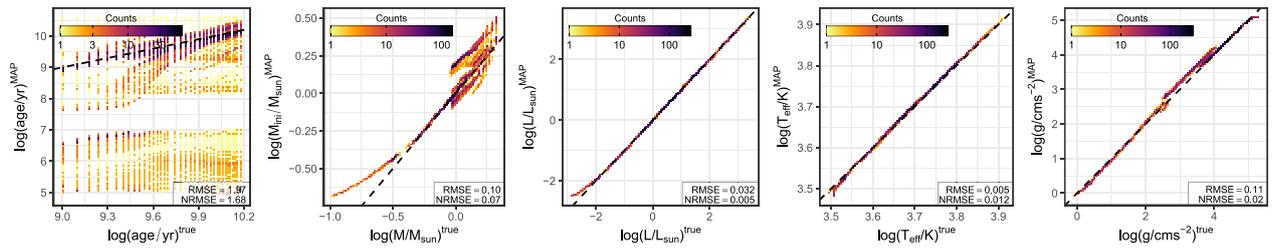
Figure B3. 2D histograms of the MAP predictions for the physical parameters on ∼10 000 samples from the Dartmouth isochrone tables as predicted by the 'Wd2_I' cINN model. Note that the nrmses are normalized to the parameter ranges of the Dartmouh ground truth here instead of the ranges of the 'Wd2_I' PARSEC training data.
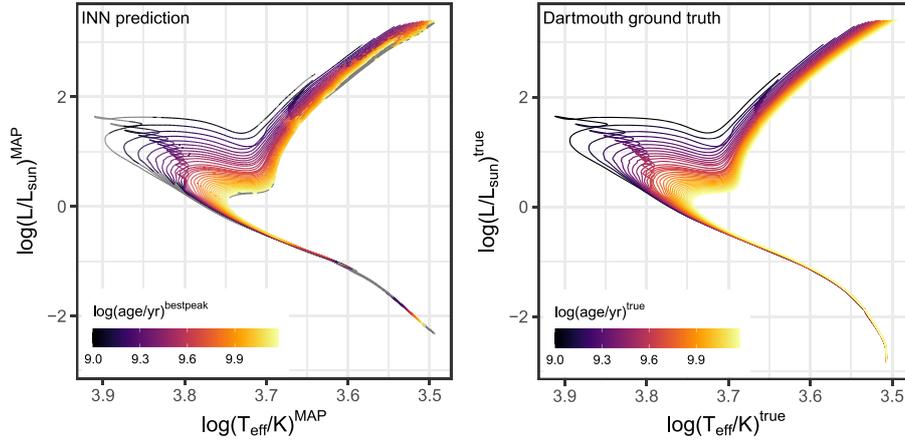
Figure B4. Left: cINN prediction for the HRD of the Dartmouth isochrones using model 'Wd2_I'. Note that MAP estimates are used for luminosity and effective temperature here, but the colour code that indicates the predicted age does not correspond to the MAP age prediction but rather the best-fitting peak of the predicted age posterior. The latter is done to take multimodal age posteriors into account. Right: ground truth HRD of the Dartmouth isochrones, colour coded according to their age.

APPENDIX C: 'WD2_II'

In this appendix, we provide additional plots and further discussion for the cINN trained on 'Wd2_II'.

Figs C1 and C2 present the MAP and posterior against true diagrams for 'Wd2_II' corresponding to Figs 9 and 10 presented in the main paper for 'Wd2_I'. These two diagrams show that the final cINN model on 'Wd2_II' does not differ significantly from the 'Wd2_I' solution. Aside from a few more outlier cases that likely cause the increased rmses of the physical parameters beside age,

the point estimate performance shows similar successes and flaws, especially the same increased number of outliers in the age prediction for the young stars. As already indicated by comparable median uncertainties (except age), the posterior distributions are equally unaffected by the age cut in the 'Wd2_II' training set. The predicted age posteriors also do not show significant changes except for the obvious limitation due to the smaller age range of 'Wd2_II' and a rare tendency to extrapolate down to $\log(\text{age/yr}) = 4$, which we do not observe for 'Wd2_I'. The more limited age range of course eliminates some of the degeneracies that may have caused some of
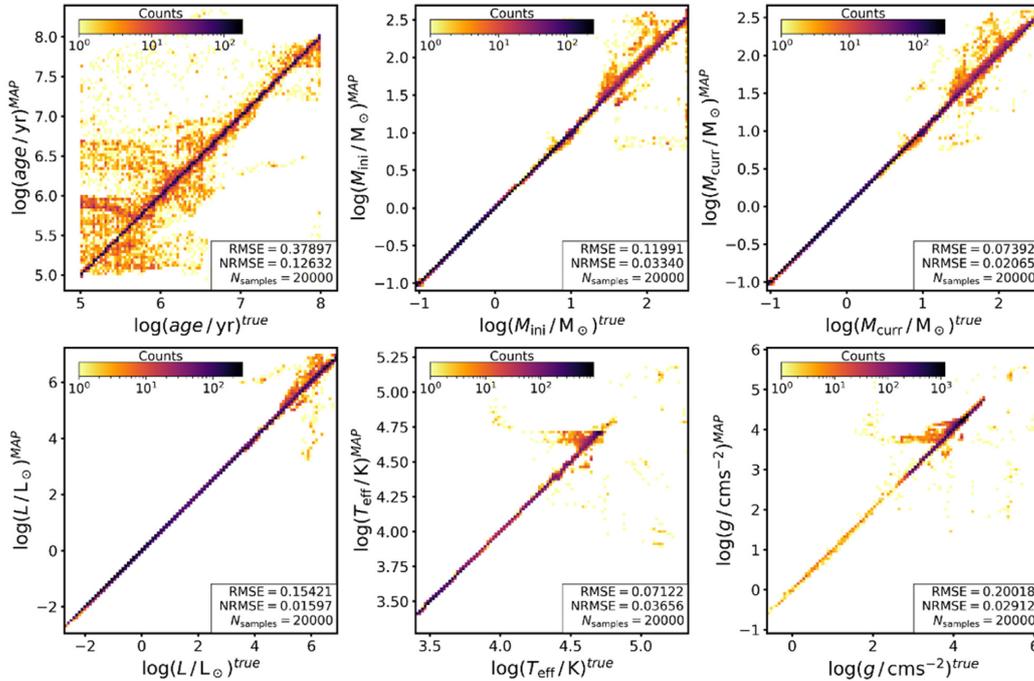
Figure C1. 2D histograms of the MAP estimates plotted against the true values for the six physical parameters we predict with the cINN trained on 'Wd2_II' for 20 000 cases from our test set. From top left to bottom right, we show age, M_{ini}, M_{curr}, L, T_{eff}, and g.
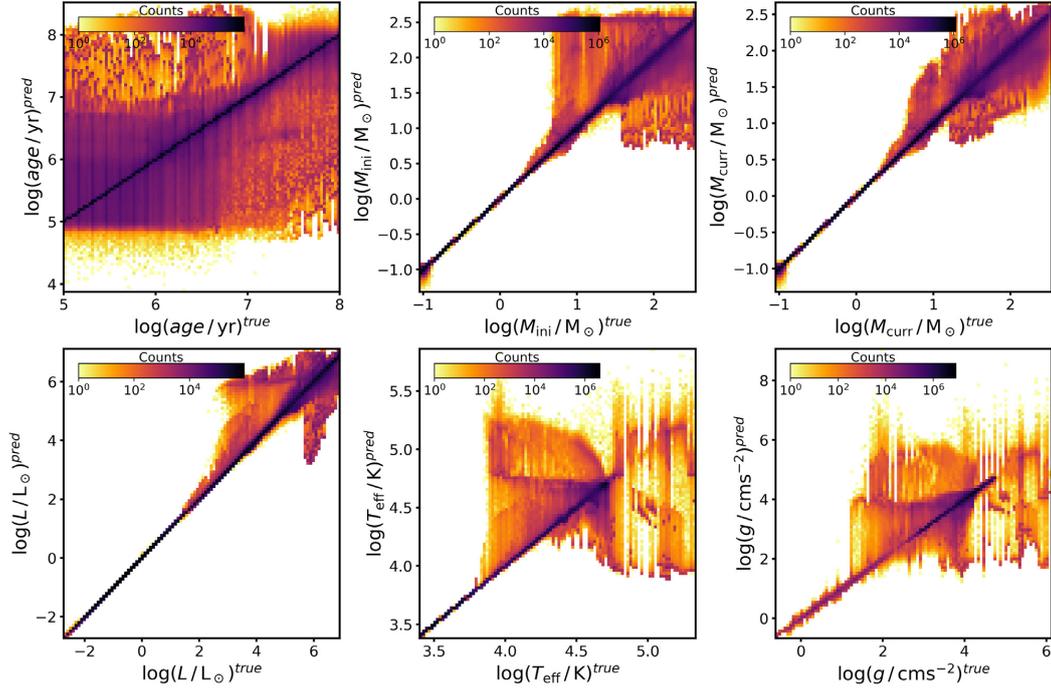
Figure C2. 2D histograms of the entire predicted posteriors plotted against the true value of the six physical parameters provided by the cINN trained on 'Wd2_II' for 20 000 cases from the respective test set. From top left to bottom right age, M_{ini}, M_{curr}, L, T_{eff}, and g are shown.
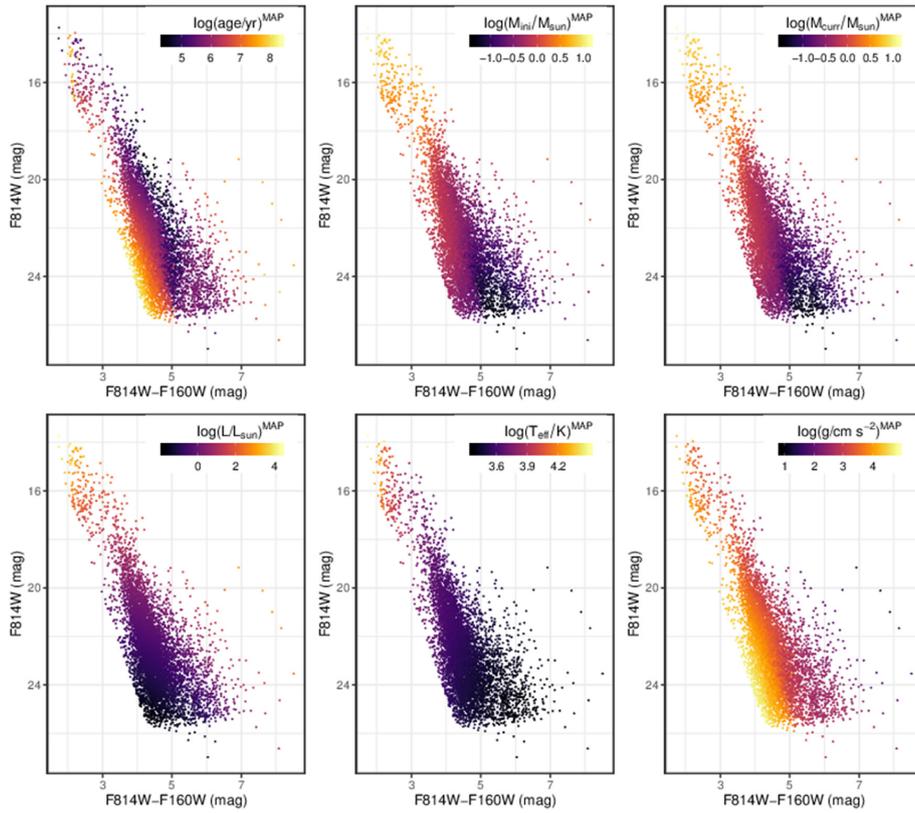
Figure C3. Optical CMDs of the Wd2 *HST* data, colour coded according to the MAP estimates for the six physical parameters predicted with the cINN trained on 'Wd2_II'.
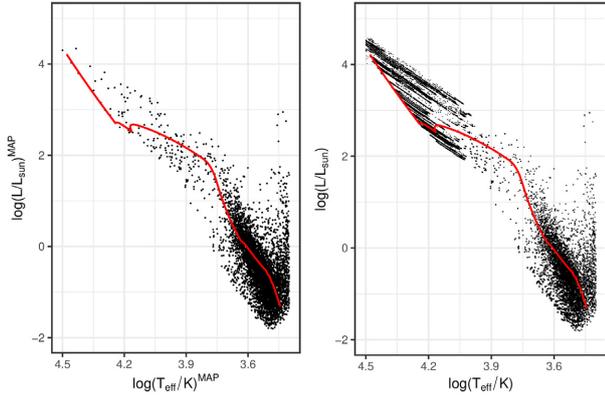
Figure C4. Predicted HRD for the Wd2 cluster constituents provided by the cINN model trained on 'Wd2_II'. The left-hand panel shows the HRD based on the MAP predictions for $\log(L)$ and $\log(T_{\rm eff})$, while in the right-hand panel, the entire posterior distributions of these two parameters are plotted for every star. The red line in both diagrams indicates a 1 Myr isochrone for comparison.

the false MAP estimates in the 'Wd2_I' prediction, like, for example, the few 10–100 Myr old stars in Fig. 9 that have an MAP prediction somewhere between 100 Myr to 1 Gyr. As these are rare cases in the previous test, however, it seems safe to say that the decrease in rmse and median uncertainty is primarily caused by the fact that the posterior distributions should span a smaller range of maximum 3 dex instead of the 5 dex in 'Wd2_I'. The almost equal nrmses confirm this.

Corresponding to Fig. 14 we show the 'Wd2_II' prediction results for all physical parameters in terms of the MAP estimates in the CMDs of Fig. C3.

Analogous to Fig. 15, we show the predicted HRD for the observational Wd2 data as given by the 'Wd2_II' model in Fig. C4.

Fig. C5 shows the comparison of the MAP estimates between the two models. In general the deviations in the MAP for all physical parameters are fairly insignificant except for age, with rms deviations on the order of a few 0.01 dex. In the case of the age prediction the deviations appear more severe, on average about 0.499 dex, likely caused by the cut-off around $\log(\text{age/yr}) = 8$ for 'Wd2_II'. Overall most predictions (note the logarithmic colour coding) fall on to the identity mapping, except for a set of about 500 stars which are placed at $\log(\text{age/yr}) \sim 5.7$ by the 'Wd2_II' model, while the 'Wd2_I' extrapolated an age below 0.1 Myr here. However, as the median relative deviation between the two MAP estimates is only 0.3 per cent it is safe to say that the two models do not differ significantly.

We conclude that the cINN does not get confused in any significant way if there are more potentially degenerate mappings in the training set, as the full model predicts the same physical parameters as the model that incorporates prior knowledge to decrease the amount of degeneracies.

Fig. C6 presents the derivation of a cluster age of Wd2 based on the sums of the age posteriors for the predictions by the 'Wd2_II' model. Here, we find a similar result as for the 'Wd2_I' model shown in Fig. 17, except for a split peak with maxima at 0.5 and 1.04 Myr in the sum of posterior distributions. As these two solutions are part of one major peak they likely belong to the same mode of the distribution, located at the Wd2 cluster age. This demonstrates again that 'Wd2_II' agrees well with 'Wd2_I'.
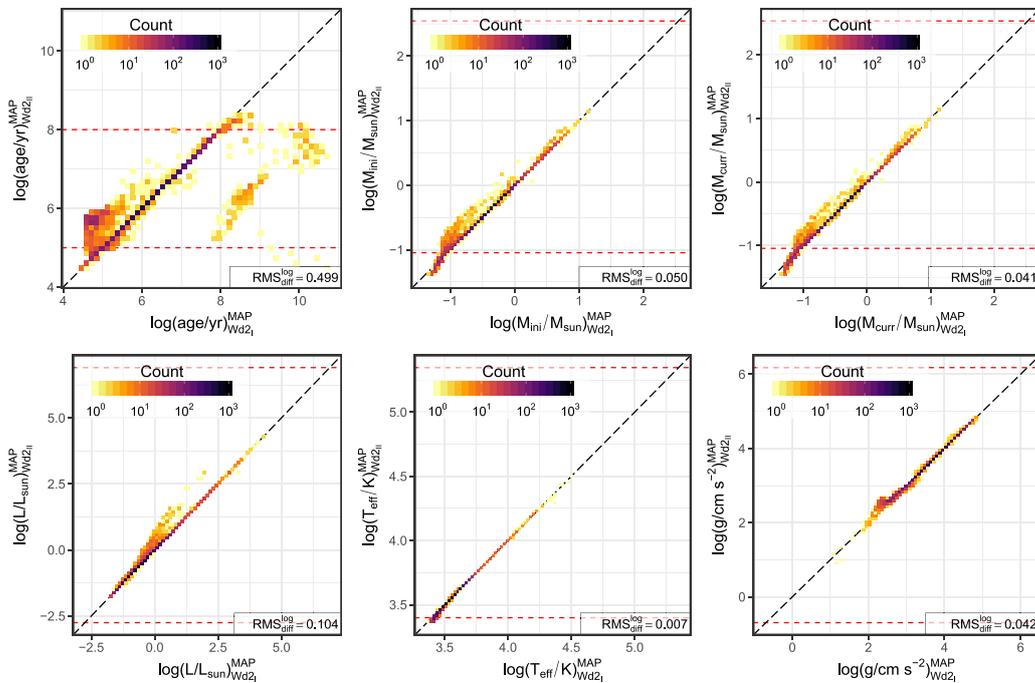
Figure C5. 2D histogram of the comparison between the MAP estimates for the physical parameters of the Wd2 stars between the two cINN models trained on 'Wd2_I' and 'Wd2_II', respectively. Note that the black dashed line indicates a perfect 1-to-1 correlation, while the red dashed lines indicate the limits of the 'Wd2_II' training set for each parameter.
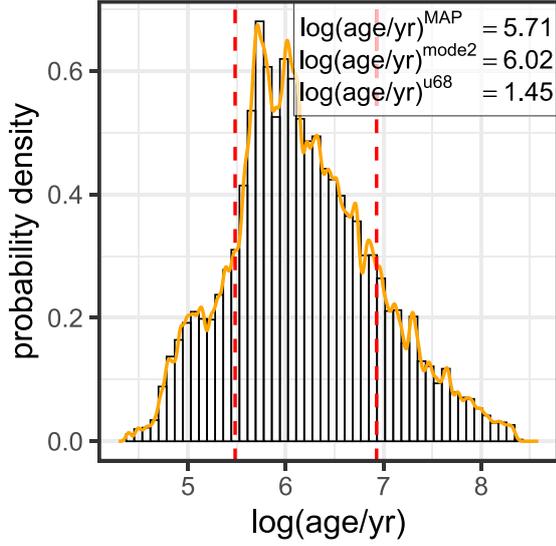
Figure C6. Histogram of the sum of the age posterior distributions of all Wd2 cluster stars as predicted by the 'Wd2_II' cINN model. The orange line indicates a kernel density fit to this cumulative posterior distribution to determine the most likely cluster age. The red dashed lines mark the width of the 68 per cent confidence interval.

APPENDIX D: 'NGC 6397_I'

Appendix D provides complementary diagrams and discussion for the 'NGC 6397_I' cINN results.

Fig. D1 provides the MAP versus true diagram for 'NGC 6397_I' analogous to Fig. 9 and Fig. C1. Fig. D2 shows the corresponding posterior versus true diagram (cf. Fig. 10 and Fig. C2). These two figures show that the 'NGC 6397_I' model delivers overall similar results to 'Wd2_I', except for slight improvements in prediction accuracy. These can likely be accredited to the larger coverage of five filters in 'NGC 6397_I' over the two filters in 'Wd2_I' as a comparison of Fig. D1 to the five filter experiment for 'Wd2_I' in Fig. B2 confirms. The same comparison holds for Fig. D2 versus the five filter results for 'Wd2_I' displayed in Fig. 11.

Figs D3–D6 present the 'NGC 6397_I' cINN model prediction results for the MIST and Dartmouth synthetic isochrones, respectively. They correspond to Figs 12 and 15, shown in the main paper. As already discussed in Section 4.4, these diagrams here show that the 'NGC 6397_I' is similarly successful on the MIST data set as 'Wd2_I', although suffering from overall larger prediction errors, but fails quite severely on the Dartmouth data due to the significant model discrepancies between Dartmouth and PARSEC.
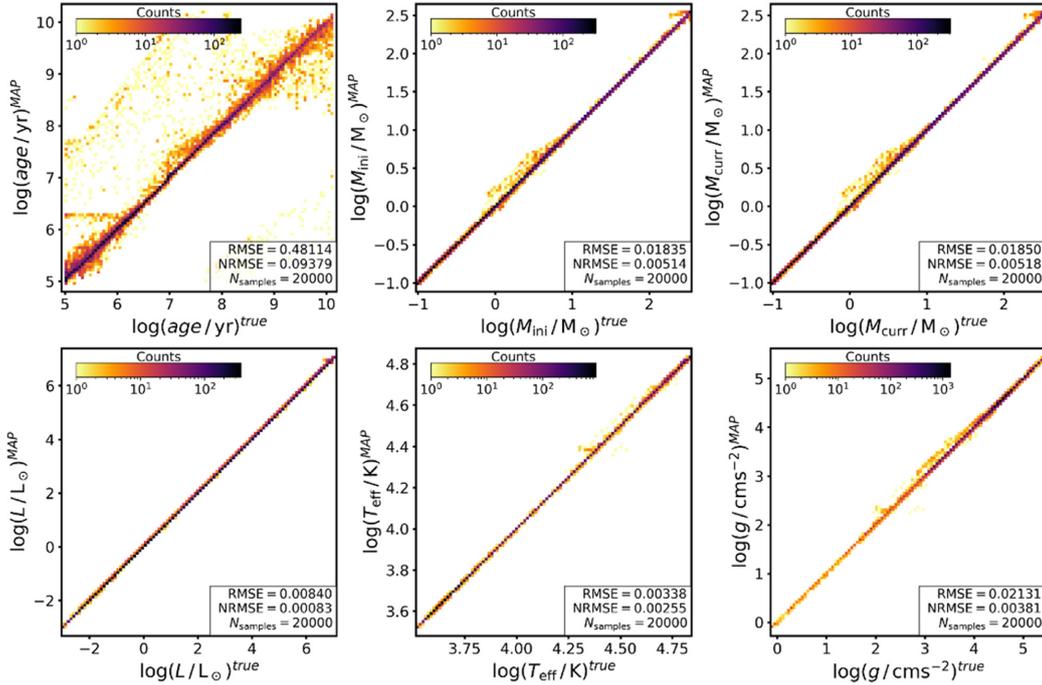
Figure D1. 2D histograms of the MAP estimates plotted against the true values for all six predicted physical parameters as given by the cINN trained on 'NGC 6397_I' for 20 000 test set observations.
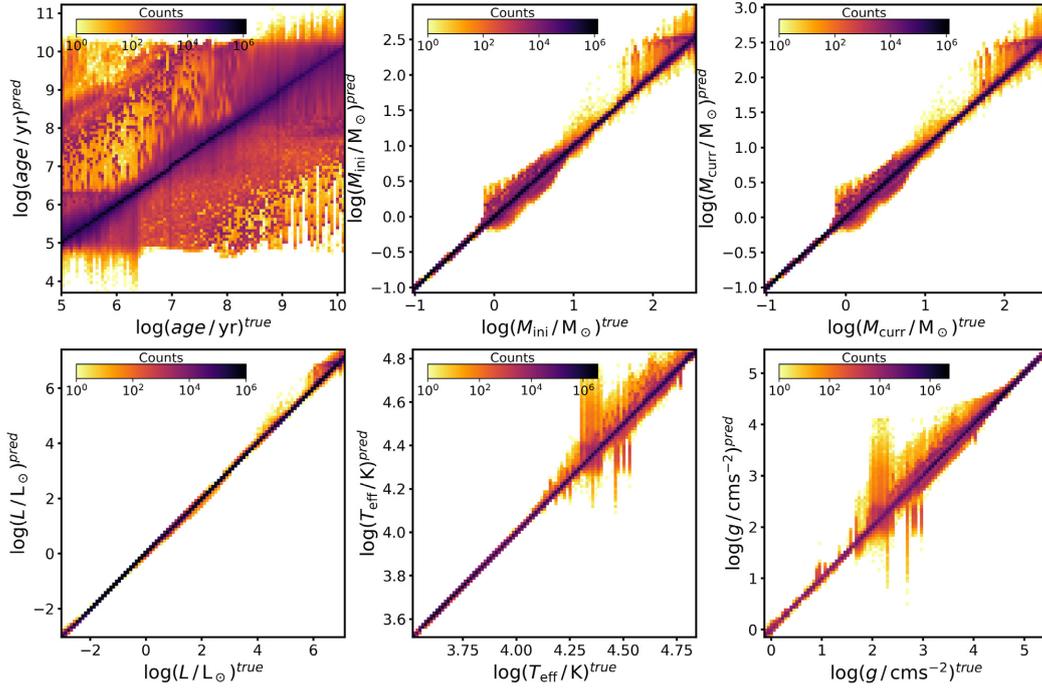
Figure D2. 2D histograms of the entire predicted posteriors plotted against the true value of the six physical parameters given by the cINN trained on 'NGC 6397_I' for 20 000 cases from the respective test set. From top left to bottom right age, M_{ini}, M_{curr}, L, T_{eff}, and g are shown.
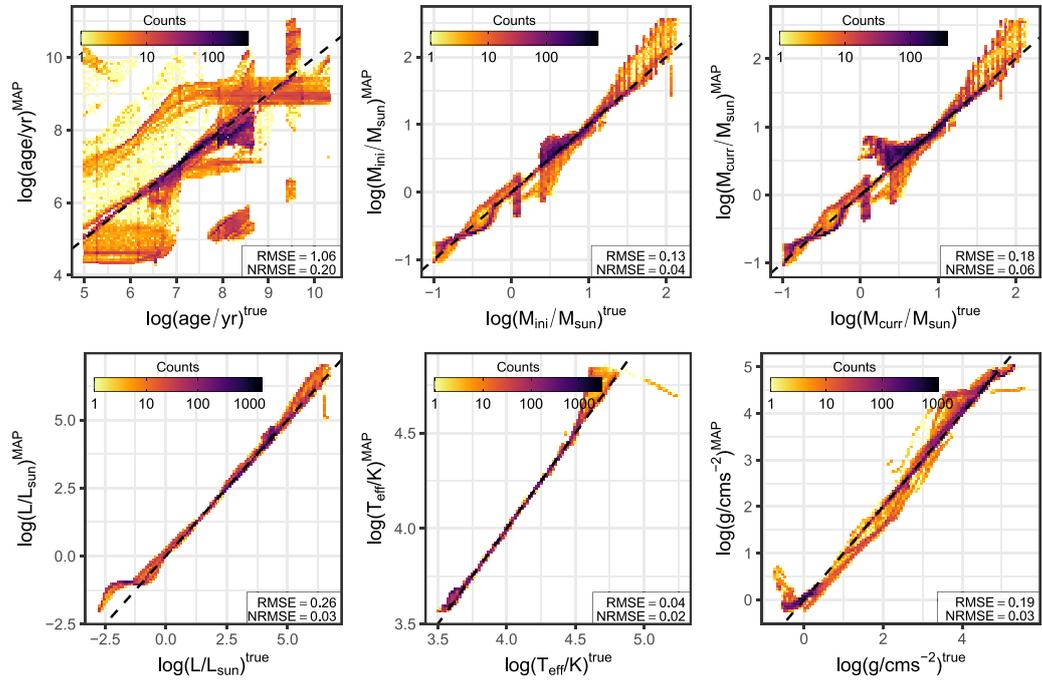
Figure D3. 2D histograms of the MAP estimates for the physical parameters of 40 000 samples of the synthetic MIST isochrone tables as predicted by the 'NGC 6397_I' cINN model. Note that the nrmses are normalized to the physical parameter ranges of the MIST ground truth here instead of the 'NGC 6397_I' PARSEC training data.
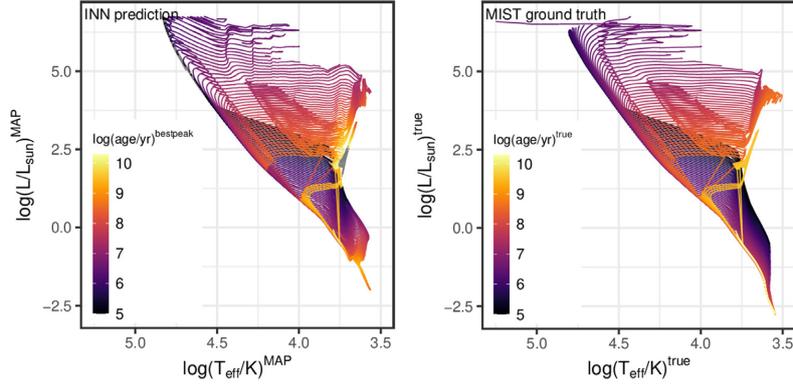
Figure D4. Left: cINN prediction for the HRD of the MIST isochrones using model 'NGC 6397_I'. Note that MAP estimates are used for luminosity and effective temperature here, but the colour code that indicates the predicted age does not correspond to the MAP age prediction but rather the best-fitting peak of the predicted age posterior. The latter is done to take multimodal age posteriors into account. Right: ground truth HRD of the MIST isochrones, colour coded according to their age.
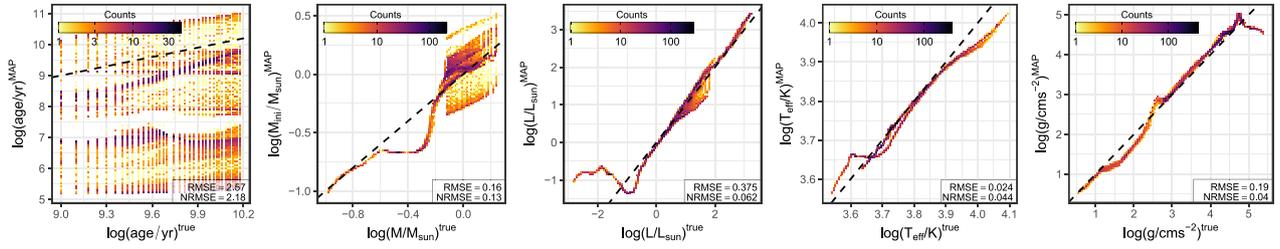
Figure D5. 2D histograms of the MAP estimates for the physical parameters of 10 000 samples of the synthetic Dartmouth isochrones as predicted by the 'NGC 6397_I' cINN model. Note that the nrmses are normalized to the physical parameter ranges of the MIST ground truth here instead of the 'NGC 6397_I' PARSEC training data.
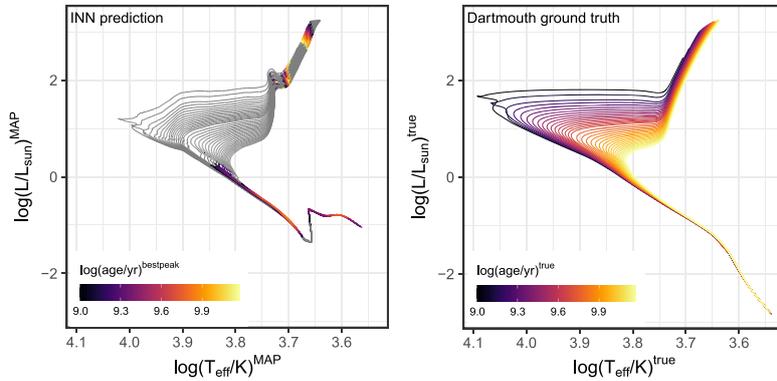
Figure D6. Left: cINN prediction for the HRD of the Dartmouth isochrones using model 'NGC 6397_I'. Note that MAP estimates are used for luminosity and effective temperature here, but the colour code that indicates the predicted age does not correspond to the MAP age prediction but rather the best-fitting peak of the predicted age posterior. The latter is done to take multimodal age posteriors into account. Right: ground truth HRD of the MIST isochrones, colour coded according to their age.

APPENDIX E: 'NGC 6397_II'

This appendix presents additional diagrams and discussion concerning the results of the 'NGC 6397_II' cINN model.

Figs E1 and E2 show the MAP and posterior versus true diagrams for 'NGC 6397', respectively, corresponding to Figs 9 and 10 for 'Wd2_I', Figs C1 and C2 for 'Wd2_II', and Figs D1 and D2 for 'NGC 6397_I'.

Fig. E3 shows the CMD coloured according to the MAP estimates analogous to Fig. 20 for 'NGC 6397_I'. In correspondence to the top left panel of Fig. E3, and Fig. 24 in the main paper provides a breakdown of the age prediction results in the CMD, distinguishing under- and overestimates from the reasonable outcomes. Fig. 24 also indicates that the elimination of all pre-main-sequence examples in 'NGC 6397_II' helps the cINN to recognize the turn-off and RGB stars as old objects (cf. Fig. 21 for 'NGC 6397_I'). As the right-hand panel demonstrates, discrepancies between the observed and modelled RGBs are again a likely cause for the cINN age overestimates for a number of RGB constituents. While the age predictions have arguably somehow improved, the prediction of effective temperature and luminosity appear to suffer slightly with the 'NGC 6397_II' cINN. Fig. E4 shows the corresponding predicted HRD in the left-hand panel, indicating about 650 outliers (orange points) to the right of the 13 Gyr isochrone, which do not appear in the corresponding 'NGC 6397_I' diagram. The right-hand panel in the same figure suggests that these outliers are, again, mainly LMS stars located where model and observations disagree the most.

Fig. E5 presents the cluster age derivation, analogous to Fig. 22 shown in the main paper for 'NGC 6397_I'. Here, it is worth mentioning that a cluster age of $13.4^{+3.4}_{-13.2}$ Gyr, determined from the most likely value of the sum of all individual age posteriors (excluding the posteriors of the total failure cases with $\log(\mathrm{age/yr})^{\mathrm{MAP}} < 5$), is actually fairly plausible despite the described issues with the age prediction. Even neglecting these problems, however, the large uncertainties of this estimate make this outcome unsatisfactory.

The last Fig. E6 exhibits the 2D histogram comparing the MAP estimates for the physical parameters between the 'NGC 6397_I' and 'NGC 6397_II' predictions, analogous to the comparison between 'Wd2_I' and 'Wd2_II' presented in Fig. C5 in Appendix C. Overall the differences are more significant than those resulting from the comparison of 'Wd2_I' and 'Wd2_II'. The predictions of L, T_{eff}, and g appear to be the least affected by the change in model, being quite close to a 1-to-1 correlation, although we find a median relative deviation of 20 per cent for L and 18 per cent for g. At a first glance, the predictions of M_{ini} and M_{curr} look more scattered around the 1-to-1 correlation, but with a median relative deviation of about 18.6 per cent the difference is of similar magnitude.
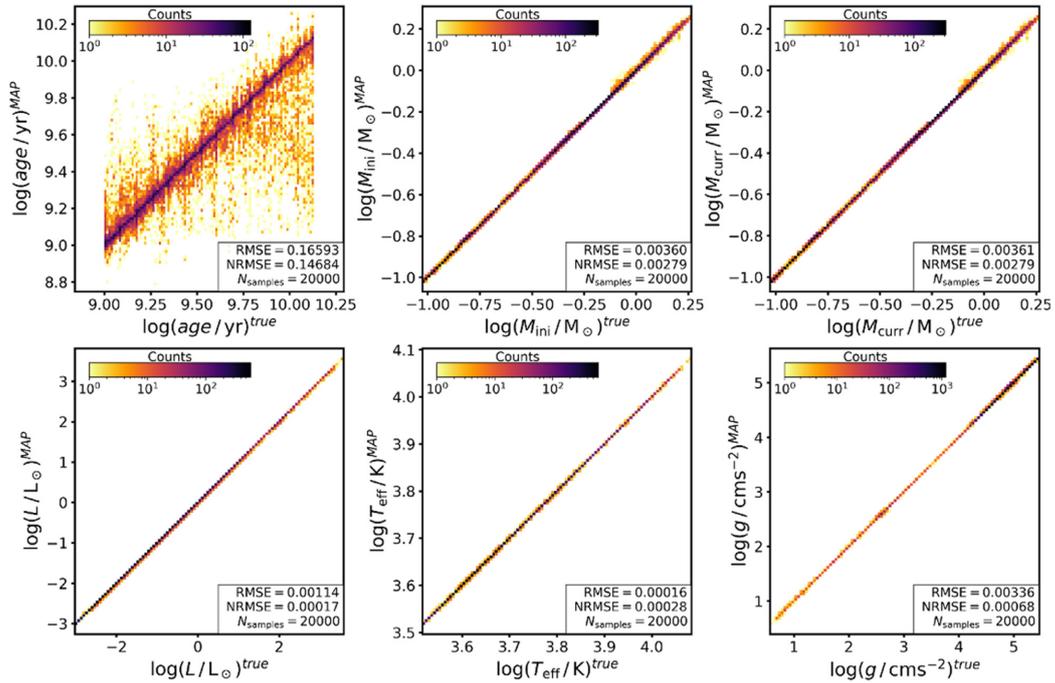
Figure E1. 2D histograms of the MAP estimates plotted against the true values for the six physical parameters we predict with the cINN trained on 'NGC 6397_II' for 20 000 cases from our test set. From top left to bottom right, we show age, M_{ini}, M_{curr}, L, T_{eff}, and g.
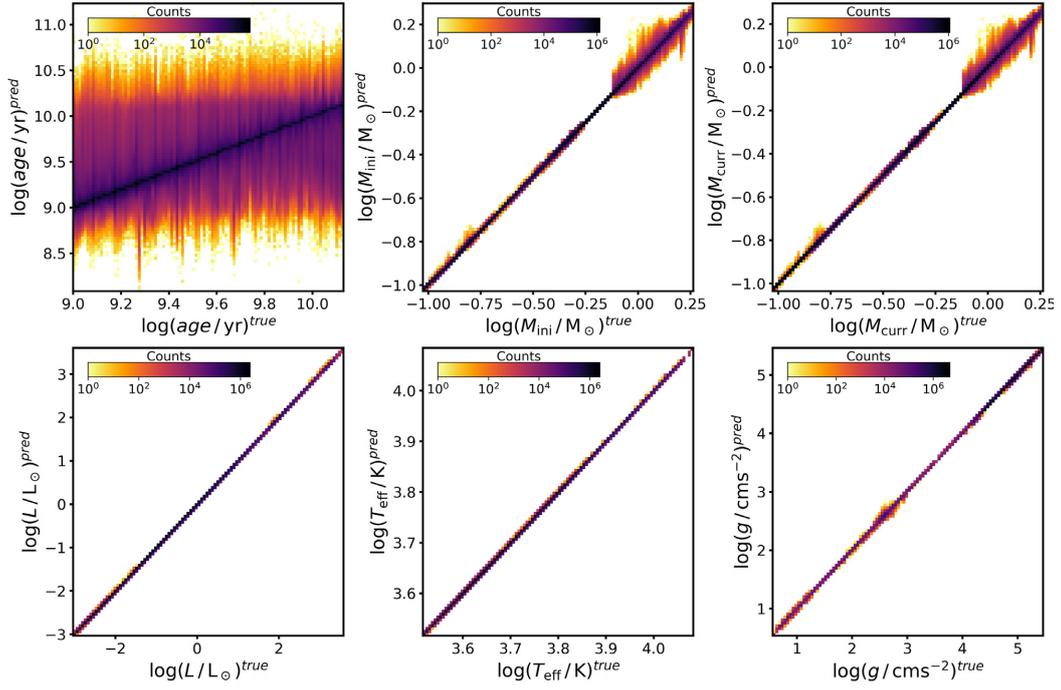
Figure E2. 2D histograms of the entire predicted posteriors plotted against the true value of the six physical parameters provided by the cINN trained on 'NGC 6397_II' for 20 000 cases from the respective test set. From top left to bottom right age, M_{ini}, M_{curr}, L, T_{eff}, and g are shown.
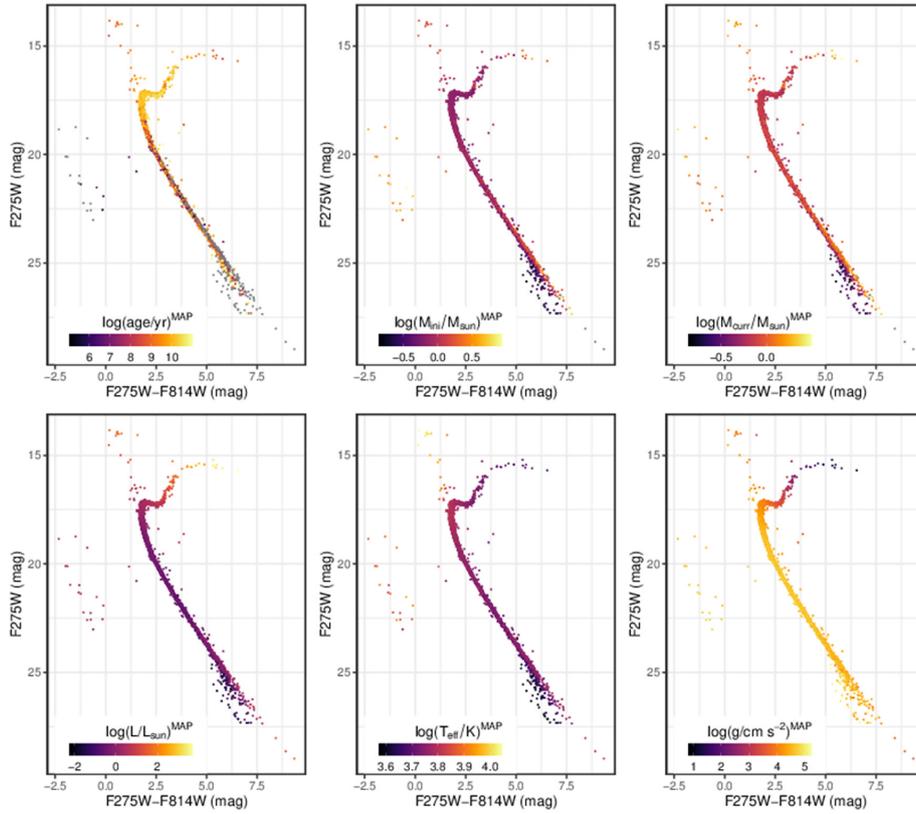
Figure E3. UV-I CMDs of NGC 6397, colour coded according to the MAP estimates for the six physical parameters predicted with the cINN trained on 'NGC 6397_II'. Note that the grey points are those stars for which the prediction falls outside of the range provided by the respective colour bar of each panel.
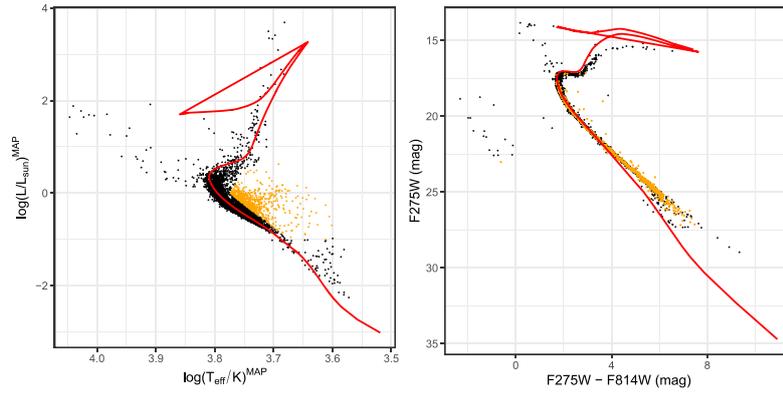
Figure E4. The left-hand panel shows the predicted HRD of NGC 6397 based on the MAP estimates of $\log(L)$ and $\log(T_{eff})$ by the cINN trained on 'NGC 6397_II'. Highlighted in orange are stars which we deem outliers. The right-hand panel shows the corresponding UV-I CMD of NGC 6397 indicating the position of these outliers in relation to the remaining observational data. The red line in both diagrams marks a 13 Gyr isochrone, the supposed age of NGC 6397, for comparison. These two diagrams show that the outlier predictions in the HRD are predominantly located in the CMD where the model and observational data deviate the most from each other.
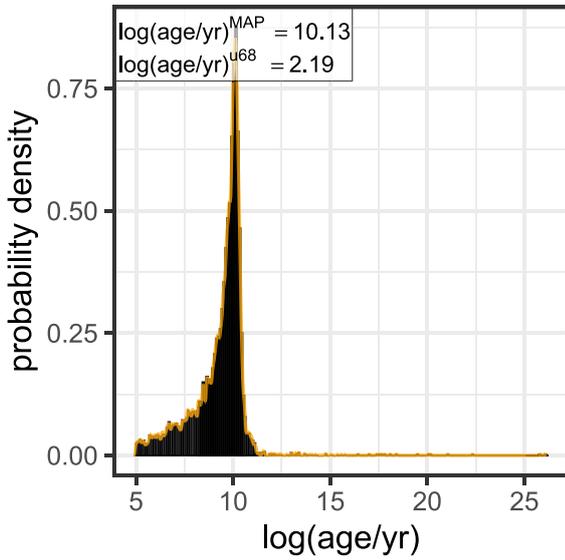
Figure E5. Histogram of the sum of the age posterior distributions of all NGC 6397 stars as predicted by the 'NGC 6397_II' cINN model. The orange line indicates a kernel density fit to this cumulative posterior distribution to determine the most likely cluster age.
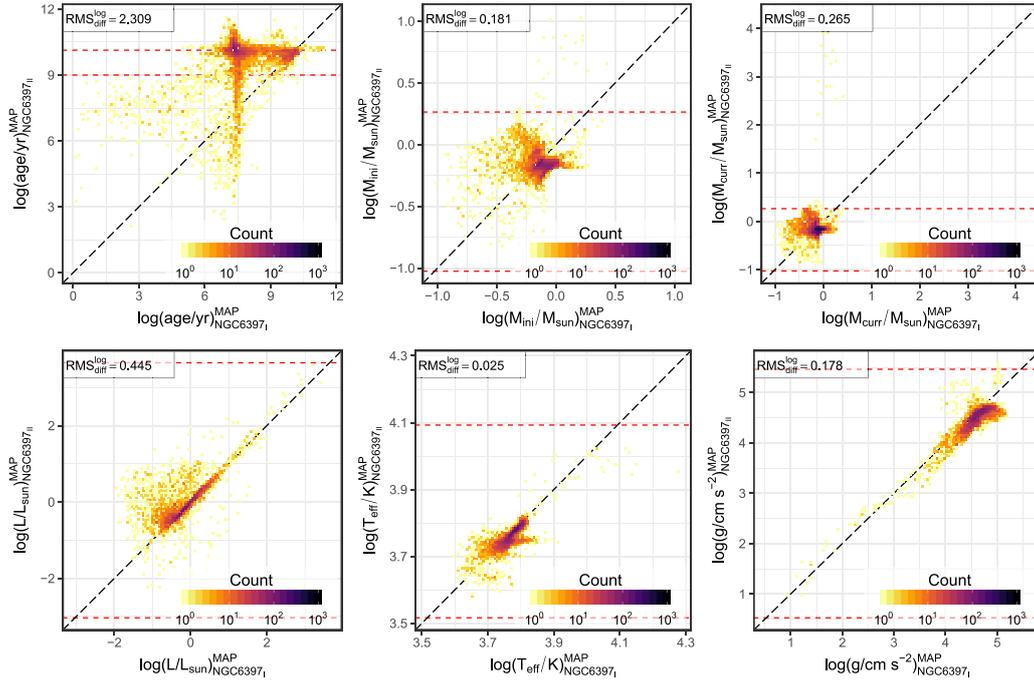
Figure E6. 2D histogram of the comparison between the MAP estimates of the physical parameters as predicted by the cINN models trained on 'NGC 6397_I' and 'NGC 6397_II', respectively. The black dashed line indicates a perfect 1-to-1 correlation, while the red dashed lines mark the limits in parameter space of the training set 'NGC 6397_II'.

This paper has been typeset from a TEX/LATEX file prepared by the author.

3.3 Measuring Young Stars in Space and Time - I. The Photometric Catalog and Extinction Properties of N44

Ksoll et al. (2021a) presents the MYSST survey (see also Section 2.5) of the star-forming complex N44 located in the LMC, its observing strategy and photometric catalogue, and investigates extinction properties of N44.

Contributions: Dimitrios Gouliermis initiated the MYSST collaboration and led the HST survey as PI of the MYSST project. Jenna Ryon, Elena Sabbi and Massimo Robberto led the data reduction and completeness analysis, and prepared the photometric catalogue with additional assistance provided by Andrew Dolphin. Ullrich Koethe suggested the RANSAC algorithm for fitting the reddening vector. I conducted all analysis on the final photometric catalogue, including the implementation of the RANSAC approach for the derivation of the reddening properties and subsequent construction of extinction maps for N44. I also created all figures (except Fig. 3-5) and wrote the text for most of the paper, excluding the introduction provided by Massimo Robberto and Section 2, penned by Jenna Ryon. Additionally, all co-authors assisted with feedback and proofreading during the draft stages of the manuscript.

Measuring Young Stars in Space and Time - I. The Photometric Catalog and Extinction Properties of N44

VICTOR F. KSOLL [ID],[1,2] DIMITRIOS GOULIERMIS [ID],[1,3] ELENA SABBI [ID],[4] JENNA E. RYON,[4] MASSIMO ROBBERTO [ID],[4,5] MARIO GENNARO [ID],[4,5] RALF S. KLESSEN [ID],[1,2] ULLRICH KOETHE [ID],[6] GUIDO DE MARCHI [ID],[7] C.-H. ROSIE CHEN [ID],[8] MICHELE CIGNONI,[9,10,11] AND ANDREW E. DOLPHIN[12,13]

[1] Universität Heidelberg, Zentrum für Astronomie, Institut für Theoretische Astrophysik,
Albert-Ueberle-Str. 2, 69120 Heidelberg, Germany
[2] Universität Heidelberg, Interdisziplinäres Zentrum für Wissenschaftliches Rechnen,
Im Neuenheimer Feld 205, 69120 Heidelberg, Germany
[3] Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany
[4] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
[5] Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA
[6] Universität Heidelberg, Heidelberg Collaboratory for Image Processing, Visual Learning Lab,
Berliner Str. 43, 69120 Heidelberg, Germany
[7] European Space Research and Technology Centre, Keplerlaan 1, 2200 AG Noordwijk, Netherlands
[8] Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-53121 Bonn, Germany
[9] Department of Physics - University of Pisa, Largo B. Pontecorvo, 3 Pisa, 56127, Italy
[10] INFN, Largo B. Pontecorvo 3, 56127, Pisa, Italy
[11] INAF-Osservatorio di Astrofisica e Scienza dello Spazio, Via Gobetti 93/3, 40129, Bologna, Italy
[12] Raytheon, 1151 E. Hermans Road, Tucson, AZ 85706, USA
[13] Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721, USA

ABSTRACT

In order to better understand the role of high-mass stellar feedback in regulating star formation in giant molecular clouds, we carried out a Hubble Space Telescope (HST) Treasury Program Measuring Young Stars in Space and Time (MYSST) targeting the star-forming complex N44 in the Large Magellanic Cloud (LMC). Using the F555W and F814W broadband filters of both the ACS and WFC3/UVIS, we built a photometric catalog of 461,684 stars down to $m_{\rm F555W} \simeq 29\,{\rm mag}$ and $m_{\rm F814W} \simeq 28\,{\rm mag}$, corresponding to the magnitude of an unreddened 1 Myr pre-main-sequence star of $\approx 0.09\,M_{\odot}$ at the LMC distance. In this first paper we describe the observing strategy of MYSST and the data reduction procedure and present the photometric catalog. We identify multiple young stellar populations tracing the gaseous rim of N44's superbubble, together with various contaminants belonging to the LMC field population. We also determine the reddening properties from the slope of the elongated red clump (RC) feature by applying the machine-learning algorithm *RANSAC*, and we select a set of upper-main-sequence (UMS) stars as primary probes to build an extinction map, deriving a relatively modest median extinction $A_{\rm F555W} \simeq 0.77\,{\rm mag}$. The same procedure applied to the RC provides $A_{\rm F555W} \simeq 0.68\,{\rm mag}$.

1. INTRODUCTION

The physical processes leading to star formation (SF) in the dynamically evolving multiphase interstellar medium (ISM) are largely regulated by massive stars. In a star-forming region, the momentum and energy feed-

Corresponding author: Victor F. Ksoll
v.ksoll@stud.uni-heidelberg.de

back from the few massive newborn stars is expected to terminate SF locally, trigger new SF remotely, and through gas expulsion modulate the gravitational potential and therefore stellar dynamics and cluster survival (see e.g. the reviews by Mac Low & Klessen 2004; Zinnecker & Yorke 2007; McKee & Ostriker 2007). Giant star-forming regions, aggregates of stellar nurseries spread over the scales of molecular clouds, are thus complex ecosystems where different stellar populations are born and interact with each other and their ambient ISM (for further discussions of the physical processes influencing ISM dynamics, see e.g. Klessen & Glover 2016 or Girichidis et al. 2020). Recent studies of high-mass star-forming regions show significant substructure and hierarchical SF (Bik et al. 2012; Gouliermis et al. 2014; Adamo et al. 2015; Sabbi et al. 2016; Cignoni et al. 2016; Gennaro et al. 2012, 2017; Nayak et al. 2016, 2018; Sun et al. 2017; Getman et al. 2018; Dib & Henning 2019; Grasha et al. 2019), with star formation rates (SFRs) and star formation efficiencies (SFEs) that vary within the same molecular cloud (Hony et al. 2015). Numerical modeling qualitatively reproduces this behavior (e.g. Bonnell et al. 1997, 2001; Klessen et al. 2000; Offner et al. 2009; Girichidis et al. 2011, 2012; Federrath & Klessen 2012; Federrath 2013; Parker & Wright 2016; Hennebelle 2018; Padoan et al. 2020), but our current understanding lacks the quantitative study of two critical measures needed to parameterize clustered SF: length scale and timescale.

Two main theories of SF on molecular cloud scales have been proposed, where the traditional approach postulates that stellar birth occurs in a slow quasi-static manner, with supporting mechanisms prolonging the cloud lifetime by many tens of dynamical times (e.g. Shu et al. 1987a; Krumholz & Tan 2007), while the more modern dynamical theory of SF acknowledges the complex morphological and kinematic structure of star-forming clouds and sees stellar birth as a highly dynamical, albeit inefficient, process. In this picture the formation of stars begins while the cloud is still forming and never reaches an equilibrium state before dispersing as a result of to feedback (e.g. Hartmann et al. 2001; Ballesteros-Paredes et al. 2007; Clark et al. 2012; Chevance et al. 2020). These two scenarios can be observationally tested, both in terms of the morphological and kinematic properties of molecular clouds and in possible local variations of the stellar initial mass function (IMF), which on average exhibits remarkably uniform behavior (Kroupa 2002; Chabrier 2003; Bastian et al. 2010; Offner et al. 2014). In quasi-static models, molecular clouds are globally gravitationally bound and well supported, allowing for slow SF, resulting in large age spreads (on

the order of several dynamical timescales; Shu et al. 1987b; Tan et al. 2006). In contrast, if the clouds are dynamically evolving and not necessarily globally gravitationally bound or very long-lived, there will be a large variety of physical conditions including bound and unbound regions that will produce stars at both high and low efficiency, respectively. Furthermore, the subsolar IMF can be significantly different as a function of the stellar clustering with a deficit of low-mass stars in the unbound, low-SFE regions (Bonnell et al. 2011).

To test these models, one has to carry out a census of newly born stars across a giant star-forming complex to identify and characterize each individual star-forming region over the whole field. In particular, one would like to analyze the distribution of stellar ages, subsolar IMF and SFE, and how these depend on the local gas properties. While young stellar clusters are typically dominated by a handful of early-type stars already on the main sequence (MS), their main stellar population is largely composed by a multitude of intermediate- and low-mass stars in the pre–main-sequence (PMS), i.e. stars still in gravitational contraction toward the MS (e.g. Nota et al. 2006; Sabbi et al. 2008; Cignoni et al. 2009; Vallenari et al. 2010; Gouliermis et al. 2007, 2011). Due to their relatively slow evolutionary time-scales (a $1\,M_\odot$ star contracts to the MS in ~ 50 Myr), they can be utilized as chronographs of the SF history of the entire region. Therefore, while high-mass MS stars provide us with the signposts of ongoing star formation, it is the population of intermediate-mass ($3 \lesssim M/M_\odot \lesssim 8$) Herbig Ae/Be and low-mass ($M \lesssim 3M_\odot$) T Tauri PMS stars that can provide us a direct measure of its youthfulness. In particular, by analyzing and comparing the different Hertzsprung-Russell diagrams one can chronologically sequence the recent star formation events, their duration, their mutual relations, and the possible differences between their stellar populations.

Our neighboring galaxy, the Large Magellanic Cloud (LMC), provides the ideal environments for this study. The LMC is a well-established laboratory to study SF because of its low metal abundance (with a metallicity $Z \simeq 1/3$ $Z_\odot$, the LMC is a proxy of the early universe conditions at the cosmic noon of SF history, $z \sim 1.5$; e.g. Madau et al. 1996), low interstellar extinction (Gordon et al. 2003), and high SF activity. We have focused our attention on the LMC H II complex LHα 120-N44 (Henize 1956, in short N44), with its rich ensemble of H II regions, bubbles, and young stellar clusters. The massive stars of the OB association LH 47 (Lucke & Hodge 1970), located in the central super-bubble of N44, are the primary drivers of the expansion of the main bubble (Oey & Massey 1995). X-ray observations reveal

$T \sim 10^6$ K gas heated by fast stellar winds and supernova explosions (Jaskot et al. 2011). The effects of stellar energy feedback, in particular along the western rim of the bubble where SF may have been triggered by its expansion, are also evident through its Hα and Spitzer images (Chen et al. 2009; Carlson et al. 2012). Herschel dust mass maps reveal the complex hierarchical ISM structure of N44 (Hony et al. 2010), and CO surveys show that SF activity arises from one molecular cloud complex (Fukui et al. 2001; Wong et al. 2011), which can be analyzed for the process of hierarchical SF. N44 has a total Hα luminosity that places it between 30 Dor, an exceptional starburst event also in the LMC, and M42 in Orion, our closest example of ongoing massive SF (30 Dor:N44:Orion = 20:1:0.04; Kennicutt & Hodge 1986). In conclusion, with multiple star-forming "hot spots" at different evolutionary stages, this complex provides the best paradigm of a "quiescently active" star-forming ecosystem.

In this paper we present the first results from a new Hubble Space Telescope (HST) Treasury Program Measuring Young Stars in Space and Time (MYSST, GO-14689, P.I. D. Gouliermis). In Section 2 we summarize the observational parameters of the MYSST survey and the data processing strategy leading to the construction of the photometric catalog. In Section 3 we introduce the different stellar populations that can be isolated in the dataset. In Section 4 we derive the optical extinction properties of N44 from the MYSST data by evaluating the slope of the reddened red clump (RC) feature in the color-magnitude diagram (CMD) and our method to assign a value of extinction to each source. Lastly, in Section 5 we discuss and summarize our findings, concluding with an outlook on our future follow-up studies.

2. OBSERVATIONS

Complementing the HST Treasury Programs on 30 Dor (GO-12939, P.I. E. Sabbi Sabbi et al. 2013, 2016) and M42 (GO-10246, P.I. M. Robberto, Robberto et al. 2013, and GO-12825, P.I. J. Shull (no refs yet)), MYSST (GO-14689, P.I. Gouliermis) is a deep, high spatial resolution HST survey of the star-forming complex N44 (Henize 1956) located in the Large Magellanic Cloud. It covers the large superbubble of N44, as well as the region south of it, with a field of view (FOV) of $12.2 \times 14.7 \, \mathrm{arcmin}^2$, which translates to about $180 \, \mathrm{pc} \times 215 \, \mathrm{pc}$ at the distance of the LMC assuming $(m - M)_0 = 18.55 \pm 0.05$ (Panagia et al. 1991; De Marchi et al. 2016). The survey provides observations in two broadband filters, F555W and F814W, with the Advanced Camera for Surveys (ACS) and Wide Field Camera 3 (WFC3, UVIS channel) instruments of the HST.

The N44 region was tiled in a grid pattern of three rows by four columns. Observations were taken in parallel, such that WFC3 covered the northern part of N44 and ACS the southern part, with a region of overlap in the middle. Table 1 lists details of the observations we describe here. Each grid point was visited twice. Each visit consisted of two orbits, the first utilizing F555W and the second F814W, with both cameras reaching down to 29 mag in F555W and 28 mag in F814W. In each orbit, four exposures were obtained using a subpixel box dither pattern. Two short (35 s) exposures in each filter were obtained during the second visit to each grid point. Two additional fields to the east of the main mosaics were obtained by a single pointing with ACS and WFC3 observing in parallel. This pointing was visited three times with the same two-orbit, four-exposure setup. Two short (35 s) exposures were obtained in each filter in the second visit. The HST two-color composite image of N44 is shown in Figure 1.

2.1. Data Processing

Bias, dark, flat-field, and charge transfer inefficiency corrected images, known as FLCs (`*_flc.fits`), were downloaded from the Mikulski Archive for Space Telescopes (MAST).[1] These processing steps were performed by the standard calibration pipelines CALWF3[2] version 3.4.1 and CALACS[3] version 9.2.0. The images were aligned to the Gaia reference frame (Gaia Collaboration et al. 2018) using `TweakReg`, part of the Drizzlepac software package.[4] The Gaia catalog was queried within the R.A, decl. bounds of the combined footprint of the FLC images, and the resulting sources were provided as a reference catalog to TweakReg to improve the absolute astrometry of our data. The coordinates of the Gaia sources span 81.1533 to 80.3041 degrees in R.A. and -68.1254 to -67.8355 degrees in decl. The FLC images were aligned to better than 0.008" (maximum root mean squared error).

The long (>400 s) FLC images were then combined using `AstroDrizzle` (Hack et al. 2012) to create reference frames for each camera and filter for photometry. We used `resetbits = 4096` to ignore existing cosmic-ray flags, `skymethod = localmin` for sky subtraction,

---

[1] http://archive.stsci.edu/
[2] https://wfc3tools.readthedocs.io/en/latest/wfc3tools/calwf3.html
[3] https://www.stsci.edu/hst/instrumentation/acs/software-tools/calibration-tools
[4] https://www.stsci.edu/scientific-community/software/drizzlepac.html
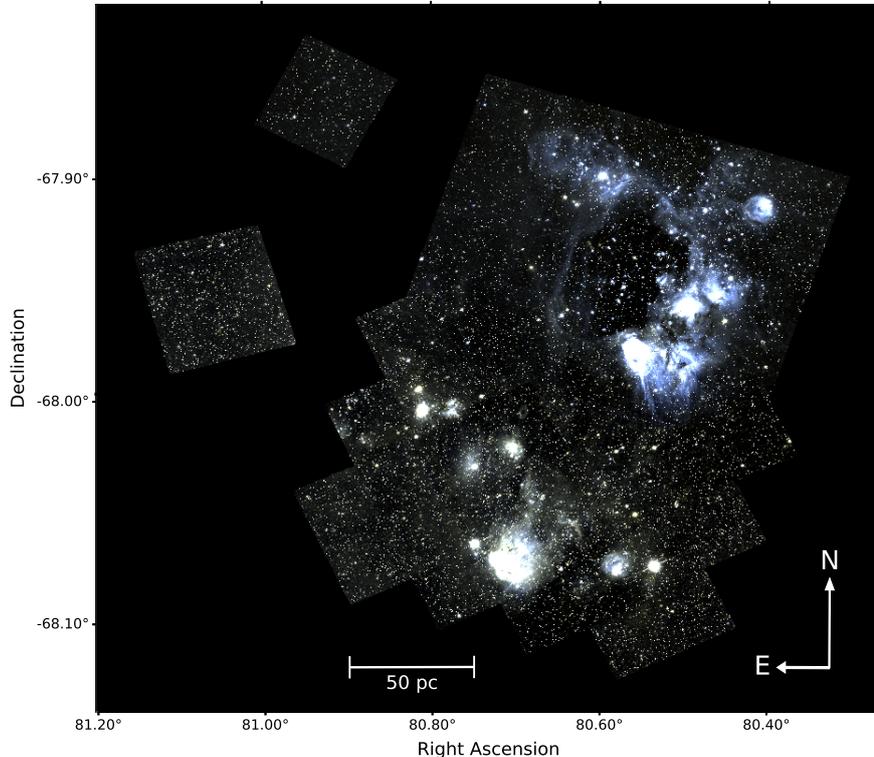
Figure 1. Two-color composite image of N44 from the MYSST HST survey with the observations in F555W in blue and those in F814W in green. N44's characteristic superbubble can be seen in the north.

and `combine_type = imedian` to avoid flagging saturated stellar cores as cosmic rays. The final drizzled images are sky subtracted and normalized by exposure time (units of $e^{-}s^{-1}$). The final pixel scales are native to each instrument, i.e. 0.05×0.05 arcsec^2 pixel^{-1} and 0.04×0.04 arcsec^2 pixel^{-1} for ACS and WFC3, respectively. The cosmic-ray flagging performed by `AstroDrizzle` was propagated back to the data quality (DQ) extensions of the input FLC images.

2.2. *Point-spread Function (PSF) Photometry*

Photometry was performed with DOLPHOT (version 2.0, downloaded on 2018 March 2;[5] see Dolphin 2000), which is capable of running photometry on multiple

images and cameras simultaneously. Because the full dataset is quite large, we split the visits into seven groups to maximize the photometric depth while minimizing the required computing resources and number of catalogs to merge. Table 2 lists the visits contained in each group. The separated field visits were grouped by camera because they do not overlap the main mosaics, and they were called Field 0 and Field 1. The main mosaic was divided into horizontal "Strips", called Strips 0 through 4. Strip 2 contains the region of overlap between the two cameras.

The N44 star-forming region is fairly crowded, requiring PSF-fitting photometry with DOLPHOT. TinyTim PSFs (Krist et al. 2011) included in the DOLPHOT download were used for both cameras. The simultaneous, iterative fitting and subtraction of stars by

---

[5] http://americano.dolphinsim.com/dolphot/

Table 1. MYSST Observations - Program 14689

Camera	Filter	Tot. EXPTIME	Short Exp.?	Visit Numbers
WFC3/UVIS	F555W	2715s	No	01-12
		2532s	No	25, 27
		2396s	Yes	13-24
		2643s	Yes	26
WFC3/UVIS	F814W	2532s	No	01-12
		2808s	No	25, 27
		2040s	Yes	13-24
		2146s	Yes	26
ACS/WFC	F555W	2558s	No	01-12
		2522s	No	25, 27
		2361s	Yes	13-24
		2517s	Yes	26
ACS/WFC	F814W	2522s	No	01-12
		2682s	No	25, 27
		2030s	Yes	13-24
		2020s	Yes	26

Table 2. Photometry Groups

Group	ACS Visits	WFC3 Visits
Strip 0	–	05-12, 17-24
Strip 1	–	01-08, 13-20
Strip 2	09-12, 21-24	01-04, 13-16
Strip 3	05-12, 17-24	–
Strip 4	01-08, 13-20	–
Field 0	–	25-27
Field 1	25-27	–

Table 3. DOLPHOT Parameters for PSF Photometry

DOLPHOT Parameters	
img_rchi = 2.0	FSat = 0.999
img_raper = 3	PSFPhot = 1
img_rsky = 15 35	FitSky = 2
img_rsky2 = 4 10	SkipSky = 2
img_rpsf = 15	SkySig = 2.25
img_apsky = 20 35	MaxIT = 25
UseWCS = 2	NoiseMult = 0.10
Align = 2	SigPSF = 3.0
aligntol = 4	CombineChi = 1
alignstep = 2	DiagPlotType = PS
Rotate = 1	ApCor = 1
img_shift = 0 0	Force1 = 1
img_xform = 1 0 0	FlagMask = 4
SecondPass = 5	ACSuseCTE = 0
RCentroid = 1	WFC3useCTE = 0
SearchMode = 1	ACSpsfType = 0
SigFind = 3.0	WFC3IRpsfType = 0
SigFindMult = 0.85	WFC3UVISpsfType = 0
SigFinal = 3.5	InterpPSFlib = 1
PosStep = 0.1	PSFres = 1
dPosMax = 2.5	psfoff = 0.0
RCombine = 1.415	

DOLPHOT refines the PSF model. The parameters used are given in Table 3.

Prior to running photometry, the SCI extensions of the drizzled reference frames and FLCs were masked according to the WHT and DQ extensions, respectively. The SATURATE header keyword was set to 71,000 e$^-$ for ACS FLCs and 55,000 e$^-$ for WFC3 FLCs because initial DOLPHOT runs were impacted by the presence of saturated pixels with values below the limits provided in the pipeline-processed files from MAST. Sky images were calculated for each FLC with $\texttt{step} = -64$, $\sigma_{\mathrm{low}} = 2.25$, and $\sigma_{\mathrm{high}} = 2.0$. The F814W drizzled frame from the appropriate camera was used as the reference image for DOLPHOT alignment of the images in each strip and field; the WFC3 drizzled frame was used for Strip

2. Among all FLCs, the long exposures aligned to within 0.008" and the short exposures to within 0.02".
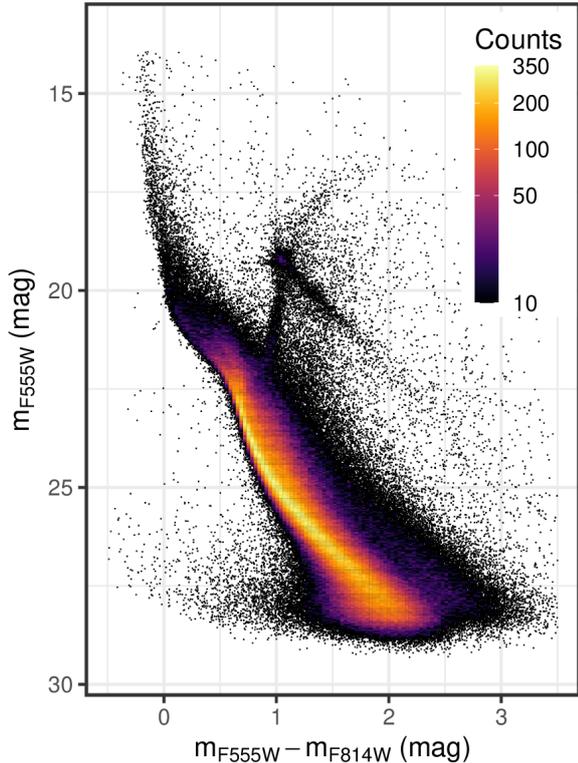
Figure 2. Optical CMD of the MYSST photometric catalog of N44. To highlight the structure of the CMD, a 2D histogram with square bins of size 0.032 mag is overlaid on the scatter plot where the number of stars per bin exceeds 10.

Stellar sources, object types 1 and 2 (object type indicates a DOLPHOT internal source classification; values ≥ 3 mark extended or single pixel sources), with signal-to-noise ratio (S/N) ≥ 5 in both filters, were selected from the full photometric catalogs to create intermediate so-called "st" catalogs. In DOLPHOT object type 2 denotes 'star too faint for PSF determination'. This refers only to a position refinement procedure, as DOLPHOT uses different methods to measure position and fluxes of the detected sources. For type 1 the PSF is used to measure both flux and position; for type 2 the position from the initial finding stage is used instead.

The ACS photometric system (VEGAMAG) and the WFC3 image coordinate system were chosen to be the survey standards. Well-measured stars in the region of overlap between the two cameras (Strip 2) were used to determine an empirical conversion from WFC3 to ACS magnitudes. In the expressions below, m_{W555} (m_{W814}) are the WFC3 F555W (F814W) magnitudes and m_{A555} (m_{A814}) are the ACS F555W (F814W) mag-

nitudes. Given the WFC3 color,

$$C = m_{\mathrm{W555}} - m_{\mathrm{W814}}, \qquad (1)$$

the ACS magnitudes are taken as

$$m_{\mathrm{A555}} = \begin{cases} m_{\mathrm{W555}} - 0.071 - 0.01(C - 1.45) \\ \qquad + 0.019(C - 1.45)^2 & C < 1.45 \\ m_{\mathrm{W555}} - 0.07 & C > 1.45 \end{cases}$$
$$(2)$$

$$m_{\mathrm{A814}} = m_{\mathrm{W814}} - 0.008. \qquad (3)$$

For the st catalogs of Strips 0 and 1 and Field 0 and the full catalog of Strip 2, the WFC3 photometry was converted to the ACS system. For Strip 2, the converted WFC3 photometry was combined with the ACS photometry with the same DOLPHOT technique used to combine multiple photometry blocks from individual FLCs (CombineChi = 1). The st catalog criteria for object type and S/N were then applied to the full Strip 2 catalog. For Strips 3 and 4 and Field 1, the source coordinates in the st catalog were converted from the ACS reference image coordinate system to the WFC3 system using astropy.wcs, and corrected for a small residual offset (-0.13 pixels in x and -0.11 pixels in y).

The st catalogs containing ACS-system photometry and WFC3-system coordinates were merged by defining dividing lines between the strips and fields. Sources were retained from an individual st catalog if they fell in a given region defined by the dividing lines. Within ± 5 pixels of each dividing line, stars were matched between the two catalogs if the distance between their centers was <1 pixel and both magnitudes were within 0.25 mag. If a match was found, the coordinates were averaged and the photometry from the strip appropriate for the average position was retained. Finally, sharpness (within ± 0.3) and crowding (≤ 0.25 mag) criteria in both filters were applied to the combined st catalog to create the final PSF photometry catalog.

2.3. *Aperture Photometry*

Finalizing the PSF photometry catalog, we found that the brightest stars are saturated even in the short exposures. To recover their flux, we performed aperture photometry with DOLPHOT. Saturated stars exhibit a "bleed-out" effect into the neighboring pixels, but in the HST CCD detectors the photogenerated charges are conserved, and using gain=2, the dynamic range of the CCDs is fully sampled by the Analog to Digital Converters. Thus, using a reasonably sized aperture, the total generated flux can be measured. To take advantage of the superior astrometry of HST and its well-understood PSF residuals, PSF photometry was first run using identical parameters to the previous runs, but

Table 4. DOLPHOT Parameters
Updated for Aperture Photometry

Updated DOLPHOT Parameters
img_raper = 6
img_rsky2 = 7 12
SecondPass = 1
RCombine = 9
PSFPhot = 0
FlagMask = 0

only on the short exposures. For the aperture photometry, saturation and cosmic-ray flags were ignored during the SCI extension masking step, and several DOLPHOT parameters were changed (listed Table 4). For each field and strip, the PSF photometry catalog from the short-exposure run was specified in the UsePhot option for the aperture photometry run.

The same `st` selection criteria, magnitude and coordinate conversions, and catalog merging steps applied to the PSF photometry were applied to the aperture photometry. Less stringent sharpness (within ±0.7) and crowding (≤0.5 mag) criteria in both filters were applied to the combined catalog to create the final aperture photometry catalog.

2.4. Combined Photometry Catalog

Lastly, the final PSF and aperture photometry catalogs were merged together and assigned flags (f_{po}) as described in the following steps. These steps were applied to stars satisfying $13.9\,\mathrm{mag} \leq \mathrm{F555W} \leq 18.3\,\mathrm{mag}$ and $12.9\,\mathrm{mag} \leq \mathrm{F814W} \leq 18.0\,\mathrm{mag}$. Brighter stars had saturated pixels extending beyond the 6-pixel aperture radius and were eliminated from the final, merged catalog, while fainter stars were better measured by PSF photometry (flag 4).

First, magnitude offsets between PSF and aperture photometry were calculated from matching high-S/N stars and applied to the aperture photometry: $\Delta m_{\mathrm{F555W}} = 0.042\,\mathrm{mag}$, $\Delta m_{\mathrm{F814W}} = 0.035\,\mathrm{mag}$. Then, blends in the final aperture photometry catalog, i.e. multiple bright stars falling in the aperture, were identified. For each star, stars within 6 pixels in the `st` PSF catalog were found, and the potential aperture photometry contribution from neighbors was calculated from the difference between the combined brightness of all PSF photometry stars in the aperture and the brightest PSF magnitude. The star was determined to be a blend if the following criteria were satisfied for either filter:

1. the potential aperture photometry contribution from neighbors was more than 0.03 mag,

2. the aperture magnitude is brighter than the PSF magnitude of the brightest star and fainter than the brightest star minus twice the potential aperture photometry contribution from neighbors.

This 0.03 mag threshold corresponds approximately to the apparent broadening of CMD features in the PSF-fitting photometry, i.e. the error introduced in the PSF magnitudes by uncertainties in the precise PSF shape. The rationale of the other criteria is that contamination is given if the aperture photometry is consistent with the sum of the brightnesses of multiple objects within the 6-pixel radius.

Lastly, for the final, merged catalog, aperture photometry was used for stars without matching sources in the final PSF catalog (flag 0) and for stars with a single match in the final PSF catalog (flag 1). PSF photometry was employed for blended stars in the aperture catalog (flag 2) and stars with no aperture detection (flag 3).

In total the MYSST photometric catalog consists of 461,684 sources across the observed FOV of N44, as well as two smaller reference fields in the LMC. Figure 2 shows the optical CMD of the survey. Due to saturation, the catalog does not include objects brighter than 14 mag in F555W and 13 mag in F814W. Consequently, some of the most massive O-type stars in the region are not part of the catalog. The faintest detected objects in the catalog reach down to about 29 mag in F555W and 28 mag in F814W. The noticeable broadening of the upper main-sequence (UMS) and the striking diagonal elongation of the RC indicate that N44 is subject to a substantial amount of differential reddening.

The MYSST photometric catalog and the four individual mosaics (F555W-ACS, F814W-ACS, F555W-WFC3 and F814W-WFC3) are available at the MAST archive as High Level Science Products via 10.17909/t9-p5vg-ke50[6]. The catalog (see Table D1 for an excerpt in the Appendix) lists for each source the survey internal ID (i.e. survey name combined with sexagesimal coordinates), pixel coordinates X and Y, celestial coordinates R.A. and decl., DOLPHOT object *type*, and magnitude m, photometric error σ, and the (DOLPHOT) photometry flag f[7] for both the F555W and F814W filters. Also

---

[6] https://archive.stsci.edu/hlsp/mysst

[7] These are bit flags; relevant for this catalog are 0, "Star well recovered in the image"; 1, "photometry aperture extends off chip"; and 2, "too many bad or saturated pixels" (see also DOLPHOT manual, http://americano.dolphinsim.com/dolphot/dolphot.pdf).

provided are the following DOLPHOT output parameters: the PSF fit quality parameter χ^2, the S/R, sharpness *shrp*, roundness *rnd* and crowding *crwd*, once for each filter individually and once as a combined value over both. In the latter case also the major axis (if source is not round) *mjaxdir* is listed.

2.5. *Artificial-star Tests*

Artificial-star tests were run with DOLPHOT to measure completeness in the N44 region. The input artificial-star list generated with `fakelist` largely matches the distribution within color-magnitude space of stars in the final PSF photometry catalog. The bounds in color-magnitude space were $14\,\mathrm{mag} < \mathrm{F555W} < 36\,\mathrm{mag}$ and $-2\,\mathrm{mag} < \mathrm{F555W} - \mathrm{F814W} < 6\,\mathrm{mag}$. To reach $\mathrm{F555W} \sim 36\,\mathrm{mag}$, the dimmest region of the real color-magnitude distribution is extended to dimmer magnitudes while retaining the same color distribution. Note that saturation effects are not modeled in DOLPHOT's artificial stars. Spatially, the artificial stars were randomly placed around the image by `fakelist` and then manually separated for each strip and field according to the dividing lines described in Section 2.2. ACS coordinates and magnitudes were converted to WFC3 coordinates and magnitudes for Strips 0, 1, and 2 and Field 0. DOLPHOT was run for each Strip and Field with the original PSF photometry parameters, but with the additional parameter FakeStars set to the appropriate input artificial-star list. The output lists (`.fake` files) for Strips 0, 1, and 2 and Field 0 were converted back to ACS magnitudes.

2.6. *Completeness*

To measure completeness, we first determine which input artificial stars were recovered by DOLPHOT. Stars were considered to be recovered if they met the same criteria in both filters for object type, S/N, sharpness, and crowding as the PSF photometry. A further recovery requirement on the artificial-star photometry was that Input − Measured < 0.75 mag in both filters. This ensures that the artificial star was not colocated with a real star of equal or brighter magnitude. In Figure 3, we plot the difference between input and measured magnitudes of the artificial stars as a function of measured magnitude. Recovered stars are highlighted in orange, and the Input − Measured < 0.75 mag requirement is shown as a horizontal line. The shaded regions indicate the magnitude ranges over which the PSF and aperture photometry catalogs were combined to capture saturated stars. The lower thresholds are $m_{\mathrm{F555W}} = 18.3\,\mathrm{mag}$ and $m_{\mathrm{F814W}} = 18.0\,\mathrm{mag}$. Because saturation effects are not properly modeled in DOLPHOT's artificial stars, we cannot study completeness effects in these regions.
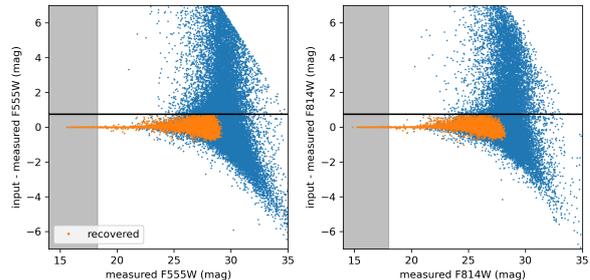
Figure 3. Difference between input and measured magnitudes as a function of measured magnitude of artificial stars for both F555W and F814W. The blue points are all artificial stars, and the orange points are those that satisfy our selection criteria for robust recovery, similar to the criteria used to accept stars in the data catalog. The horizontal line shows the Input − Measured < 0.75 mag requirement to prevent colocation of recovered artificial stars with real stars of equal or brighter magnitude. The shaded regions indicate the magnitude ranges over which saturation prevents completeness measurements.

Completeness was measured in 1000×1000 pixel spatial bins containing more than 500 stars across the main mosaics and offset fields. Within each spatial bin, we calculate the fraction of artificial stars recovered in 1 mag wide bins in measured magnitude for each filter. We used linear interpolation of the recovery fraction over a finely sampled magnitude range to find the 50% and 80% completeness magnitudes for each filter in each spatial bin. Figures 4 and 5 show the 50% and 80% completeness maps of the N44 region, respectively. Spatial bins with brighter completeness limits correspond to regions of high stellar density and, in a few cases, saturated stars. The offset field and portion of the main mosaic covered by ACS/WFC show systematically brighter completeness limits. Over the entire N44 region, the average 50% completeness is 28.1 mag for F555W and 26.7 mag for F814W, and the average 80% completeness is 27.3 mag for F555W and 25.7 mag for F814W.

Comparing the completeness in F555W to isochrones from the PARSEC stellar evolution models (Bressan et al. 2012) of the appropriate metallicity for the LMC ($Z = 0.008$), we find that the 50% and 80% limits correspond to the brightness of unreddened 0.14 and 0.18 $M_\odot$ 1 Myr old PMS stars at the distance of the LMC. The F555W detection limit of about 29 mag implies even a lowest mass limit of 0.09 $M_\odot$. Alternatively, for 10 Gyr[8]

---

[8] Note that these low-mass objects evolve so slowly that the mass limits derived from the 10 Gyr isochrone do not vary significantly from e.g. either the 1 Gyr or 100 Myr isochrone, so they are practically identical to the low-mass ZAMS value.
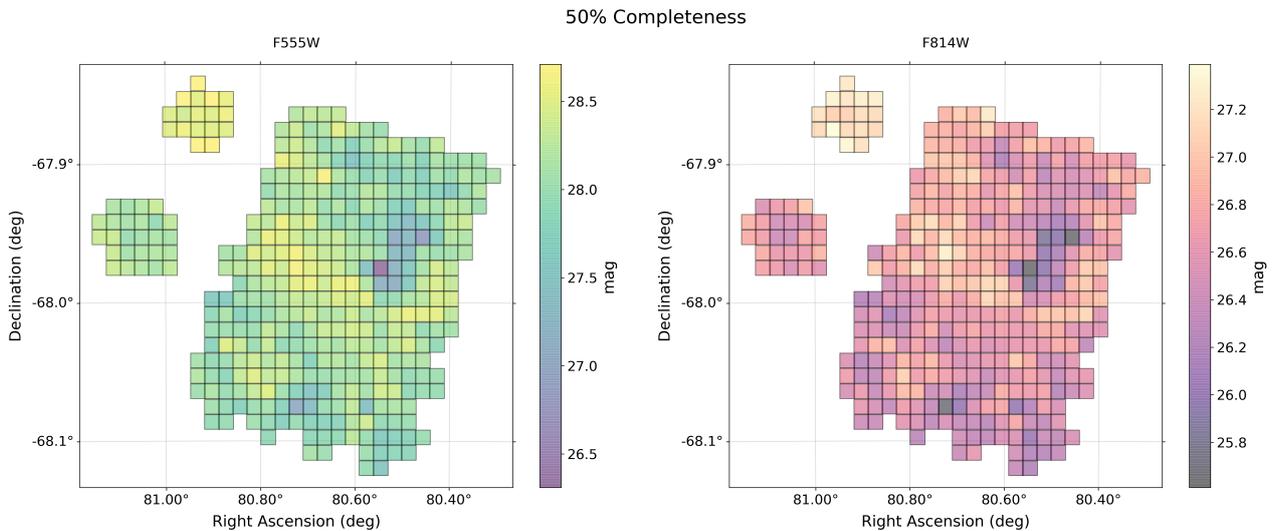
Figure 4. 50% completeness magnitude map for F555W (left) and F814W (right). Completeness magnitudes were measured in spatial bins of 1000×1000 pixels.
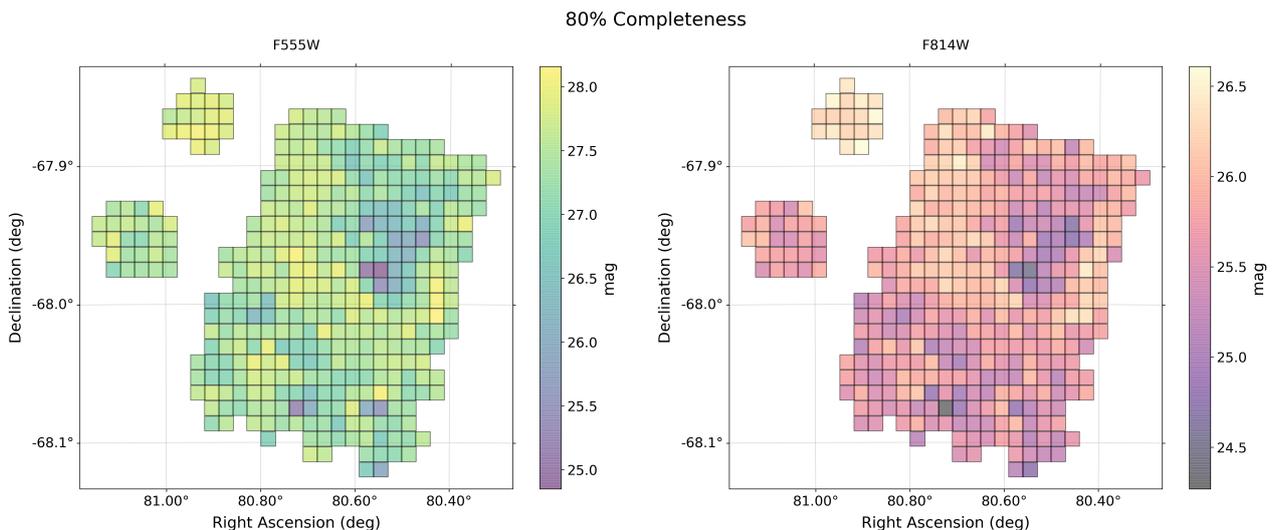
Figure 5. Same as Figure 4 for 80% completeness.

old, nonextinguished, low-mass MS objects the 80% and 50% completeness limit and the detection limit imply, mass thresholds of about 0.55, 0.5 and 0.4 $M_\odot$, respectively.

3. STELLAR POPULATIONS

In this section we shall briefly illustrate the stellar populations revealed by the MYSST survey. It is, however, not supposed to provide an exhaustive characterization. We further discuss the UMS and RC stars in Section 4 and dedicate a follow-up study (Ksoll et al. 2020, hereafter Paper II) to the PMS population of N44.

The rich CMD of the MYSST catalog (Figure 6, left) immediately reveals that the survey captures a complex collection of different stellar populations. In the left panel of Figure 6 we highlight the loci of the most prominent stellar types (from an evolutionary standpoint). In addition, the right panel of this figure indicates the spatial distributions of these roughly selected stellar populations across the observed FOV.

The UMS stars are young massive objects that trace the centers of SF in large molecular clouds and rearrange their surrounding material through their powerful feedback in the form of winds and radiation (i.e. Elmegreen
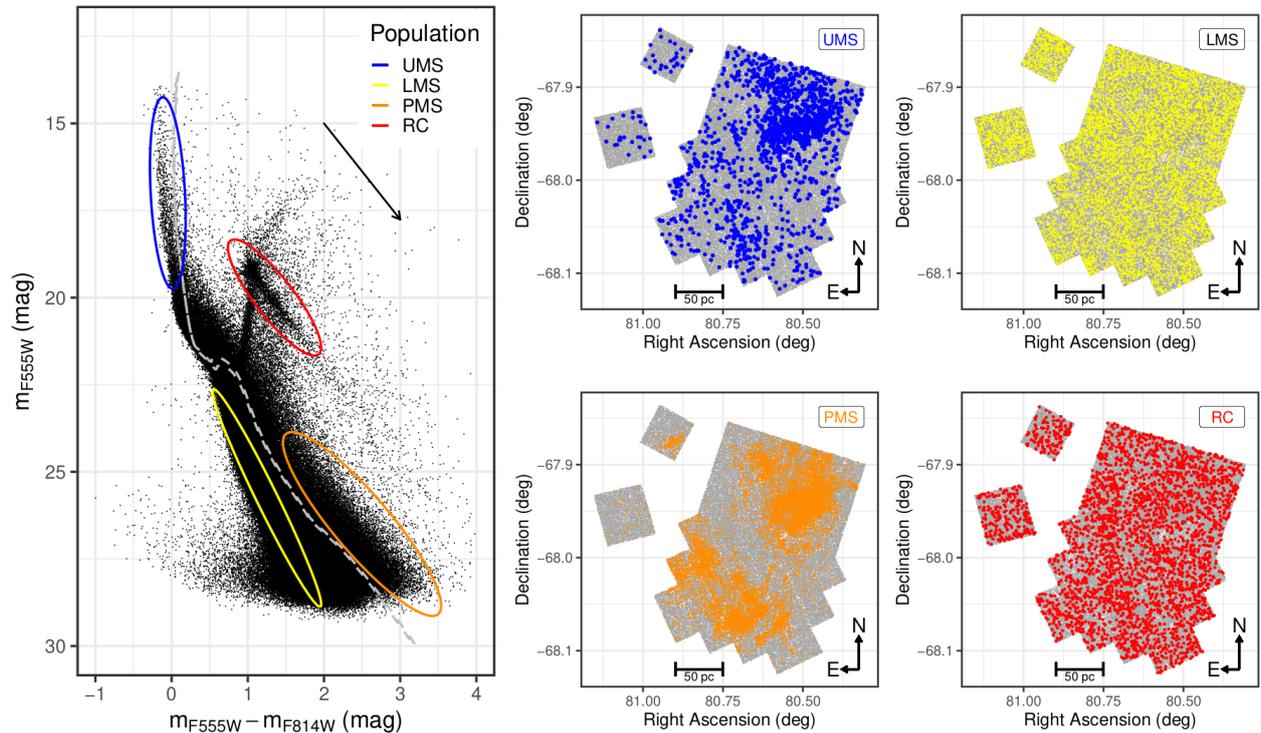
Figure 6. Left: optical CMD of the MYSST photometric catalog. Highlighted by the colored ellipses are the (rough) locations of different stellar populations found in the survey. For comparison, the gray dashed line indicates a 10 Myr PARSEC isochrone, corrected for the LMC distance modulus and the median UMS extinction derived in Section 4.2. The black arrow indicates the reddening vector as derived in Section 4.1.1. Right: spatial scatter plots of the roughly selected stellar populations shown in the CMD on the left in comparison to the total MYSST FOV. The population of LMS stars is subsampled to 10,000 examples for the diagram in the top right, as the LMS selection in the CMD contains more than 200,000 stars.

& Lada 1977; Bisbas et al. 2011; Dale et al. 2013; Walch et al. 2013). Several UMS stars show a color excess, likely being affected by significant reddening, although we cannot exclude that some of the most massive ones are rotating. In the latter case it is possible that they exhibit lower effective temperatures because of the lower internal pressure with respect to the nonrotating ones (see e.g. Meynet & Maeder 2000) or appear more luminous and hotter owing to rotation-induced internal mixing processes (see e.g. Brott et al. 2011). In both scenarios rotation will induce an additional broadening of the UMS (and actually all stars with types earlier than F).

Due to the previously mentioned saturation issues, the MYSST catalog is likely missing the highest-mass O-type stars, such that our roughly indicated UMS population consists of late O-type, B-type and early A-type objects. As the spatial distribution indicates, we find these sources predominantly in and around the massive superbubble of N44 in the northern half of the FOV. Additionally, there are a few compact clusterings south

of the bubble indicating the presence of additional star-forming groupings.

The lower-MS (LMS) sources consist mostly of old, low-mass stars that belong to the LMC field population, lying in either the foreground or background of N44, contaminating the FOV of the MYSST survey. Most of these sources are too old to belong to the young star-forming centers of N44 but make it difficult to determine the population of still-forming stars in the CMD. As extinction effects can dislocate their CMD position, they may overlap with the CMD regions reserved to objects in the formation process. Note that differential distance effects are negligible for the LMC given its low scale height of 500 pc (van der Marel & Cioni 2001) in comparison to its distance of more than 50 kpc. Therefore, even a separation of e.g. about 1 kpc between a star in front and one behind the LMC results in a magnitude difference smaller than 0.05 mag.

If there have been previous star forming events in N44, before the currently observed star forming activity, it is also possible that older (e.g. > 50 Myr) low-mass PMS

cluster stars, which are close to joining the MS, fall into the indicated LMS region in the CMD. Without additional measurements of e.g. the Hα excess, these still-forming objects are notoriously difficult to distinguish from the old field LMS stars. The LMS sources are numerous, our rough selection containing already more than 200,000 objects, and as expected for field populations, they are almost uniformly distributed (bar completeness and extinction modulation) as the subsampled (to 10,000 examples) spatial scatter plot reveals.

The RC population consists of old stars, already in their post-MS evolution, that are in the process of burning helium in their cores. The luminosity of RC stars is almost insensitive to their age (at least for RC stars older than about 2 Gyr); thus they have often been used as standard candles to determine distances (e.g. Stanek et al. 1998; Girardi & Salaris 2001) and reddening (Udalski et al. 1999a,b; Zaritsky et al. 2004; Haschke et al. 2011; De Marchi & Panagia 2014; De Marchi et al. 2016). Like the LMS sources, these, too, belong predominantly to the LMC field being foreground and background objects projected into the line of sight of N44. Under ideal observational circumstances, i.e. in the absence of reddening, these old post-MS objects form an almost circular overdensity in the CMD. Given that a fraction of the observed RC sources should be behind N44, we can infer from the notable elongation of the RC feature in the observed CMD that N44 is host to a substantial amount of obscuring gas and dust that significantly reddens background sources and N44's constituents (Dalcanton et al. 2015). As the LMS objects, the RC stars are also mostly uniformly distributed across the MYSST FOV, confirmed by the spatial scatter plot in Figure 11.

Lastly, the PMS population is composed of very young objects that are still contracting under self-gravity, sometimes still accreting gas from their circumstellar disks and envelopes (Manara et al. 2012), and not yet dense and hot enough to ignite hydrogen burning in their cores. A comparison with the isochrone traced in the left panel of Figure 6 shows that our contour selects PMS stars likely younger than approximately 10 Myr. These actively forming stars make up the bulk of N44's young stellar clusters and characterize the SF environment of this complex. As the survey title suggests, they are one of the primary targets of MYSST for recovering N44's SF history. Our rough selection of the PMS population already highlights that N44 is host to a large number of young forming stars, which are predominantly located in and at the edges of the region's characteristic superbubble. Given that the bubble traces N44's gas reservoirs, this arrangement is not surprising, but we also find additional compact clustered structures in the southern part of the survey. Our rough selection of PMS candidates only represents the most recent SF activity in N44, and it is plausible that older (e.g. ∼ 20-50 Myr) and less luminous PMS objects may be more spatially diffuse and confused with the LMS field population. In the immediate follow-up study to this introductory paper we quantify the young PMS population of N44 and characterize their clustering behavior (Ksoll et al. 2020).

4. EXTINCTION OF N44

In this section we study the extinction properties of N44 based on the photometric observations and construct an extinction map for the observed FOV. To achieve the latter, we estimate the extinction for a selection of UMS stars by projecting them along the direction of reddening onto their theoretical zero-age MS (ZAMS) locus in the CMD. Afterward, following a procedure we have outlined in Ksoll et al. (2018), we assign a distance-weighted average extinction of the 20 nearest UMS sources to the remaining stars in the catalog (see also De Marchi et al. 2016). For comparison we repeat the same procedure with RC extinction probes to further characterize the reddening profile of N44.

4.1. Extinction Properties

One way to constrain the extinction properties from our photometric observations is to utilize the RC feature in the CMD. Under perfect conditions, i.e. without extinction and photometric errors, the RC is a well-defined and easy-to-identify CMD feature. It is an overdensity caused by core helium-burning post-MS stars. Subject to differential extinction, however, the RC appears elongated in the CMD tracing the reddening vector, i.e.

$$R_{\mathrm{F555W-F814W}}(\mathrm{F}) = \frac{A(\mathrm{F})}{E(\mathrm{F555W - F814W})}, \quad (4)$$

where F indicates one of the two filters F555W or F814W.

Therefore, the underlying extinction properties can be determined by measuring the slope of this CMD feature (see e.g. De Marchi et al. 2014, 2016, 2020). To accurately retrieve this slope, the constituents of the RC need to be quantified first. Doing so by performing an arbitrary rectangular or elliptic selection around the RC has the major drawback of inducing an uncontrolled range of possible solutions. While this strategy may return a value that is not far from the true value, this approach is prone to nonobjective selection effects. To circumvent this issue, e.g. De Marchi et al. (2016) applied unsharp masking to constrain the RC feature. In this paper we follow a different approach, fitting the slope of the red clump while accounting for the presence of many

outliers using a well-established machine-learning algorithm, RANSAC.

4.1.1. *RANSAC*

The RANSAC (RANdom SAmple Consensus; Fischler & Bolles 1981) algorithm can perform robust fits to datasets that suffer from outliers. The underlying assumption of this approach is that the data consist primarily of a set of inliers and a few outliers. Furthermore, this set of inliers is explained by a single model and the corresponding set of parameters, which we want to find by fitting the model to the data. To derive these parameters, RANSAC selects a series of random subsets of the data and fits the model to each of these, deriving multiple sets of fit parameters. The idea is that, as the outliers are a minority in the data, most of these subsets consist only of inliers returning the same (or similar) fit parameters. In contrast, the sets that contain random outliers will not agree on any given fit parameters. Therefore, by simply "counting the votes" for the parameters of all the random subsets, the underlying data generating model parameters are revealed.

In practice, the algorithm first draws a random set of n points, a minimum required to fit the desired model (e.g. 2 for a line), and performs a fit to that random subset. Then, it determines the amount of remaining data that agree with this fit, the inliers to this specific set of model parameters. To do so for every data point, a distance to the fitted model is determined and compared to a preset acceptance threshold. If there are enough inliers to a model, it is accepted as a good fit. This identified "consensus" set is then employed to refine the fit by using all inlier points to reestimate the model parameters. This procedure is repeated a number of times, k, to determine the best model as given by the fitting error. The number of random samples k of size n that have been drawn is chosen such that it has a low probability p of containing only bad samples (in this study we use $p = 0.01$) and is given by the equation

$$k = \frac{\log(p)}{\log(1 - w^n)}, \tag{5}$$

where w is the fraction of inliers in the data. The parameter w is, of course, often unknown, but by starting with a low estimate of w the number of samples k can be iteratively determined by updating the current guess of w after every random sample with the actual determined fraction of inliers. This procedure stops when the number of samples that have been drawn exceeds the latest estimate of k (Fischler & Bolles 1981; Forsyth & Ponce 2003).

In order to make use of the RANSAC algorithm to determine the extinction law of N44, we first need to make an initial selection in the CMD of the region where the elongated RC is the dominant feature. The left panel of Figure 7 shows our selected region. We make this area large enough to be as agnostic as possible to our prior knowledge of where the RC is located. In particular, the axes of the rectangle do not prescribe a slope. At the same time, we make sure that the RC is the major feature within the selection, while we exclude a large portion of the red giant branch (RGB) leading up to the RC to facilitate the RANSAC inlier search.

Since our ultimate goal is to fit a line in the CMD, the RANSAC algorithm will draw samples with the minimum amount of necessary points, i.e. $n = 2$ in our application. From a series of initial tests we determine that an acceptance threshold of 0.26 in the absolute error returns inlier selections and model fits that trace the elongated RC as we would expect it. To minimize any influence from the random sampling on the final fit of the slope of the reddening vector, we repeat the RANSAC algorithm 5000 times. Note that this is not the k parameter: each of these 5000 runs will draw k random samples, with k automatically determined according to Eq. (5).

As a final value of the slope we take the average of the predicted slopes of these 5,000 RANSAC runs and use their standard deviation as the associated uncertainty. The right panel of Figure 7 presents the result of this procedure for the slope of the reddening vector, $R_{F555W-F814W}(F555W) = 2.8 \pm 0.3$. Note that we color-coded every star in this diagram according to the fraction of times across the 5000 RANSAC runs that it was selected as an inlier to the final model. The plot shows that the RANSAC algorithm is able to accurately recognize the constituents of the elongated RC feature. In fact, out of 5000 runs the RGB leading up to the red clump is never identified as the dominant feature in our selection window. Using the same data selection and RANSAC procedure, the corresponding slope in our other filter is $R_{F555W-F814W}(F814W) = 1.8 \pm 0.3$. For comparison, the Cardelli et al. (1989) galactic reddening law, with $R_V = 3.1$, returns $R_{F555W-F814W}(F555W) = 2.4$ and $R_{F555W-F814W}(F814W) = 1.4$, respectively, while the Fitzpatrick (1999) model gives $R_{F555W-F814W}(F555W) = 2.2$ and $R_{F555W-F814W}(F814W) = 1.2$. Hence, overall we find slightly larger R values in N44, indicating a more "gray" reddening.

4.2. *UMS Extinction*

With the reddening vector constrained, we now measure extinction in F555W for UMS stars by reprojecting them along the reddening vector onto their theoret-
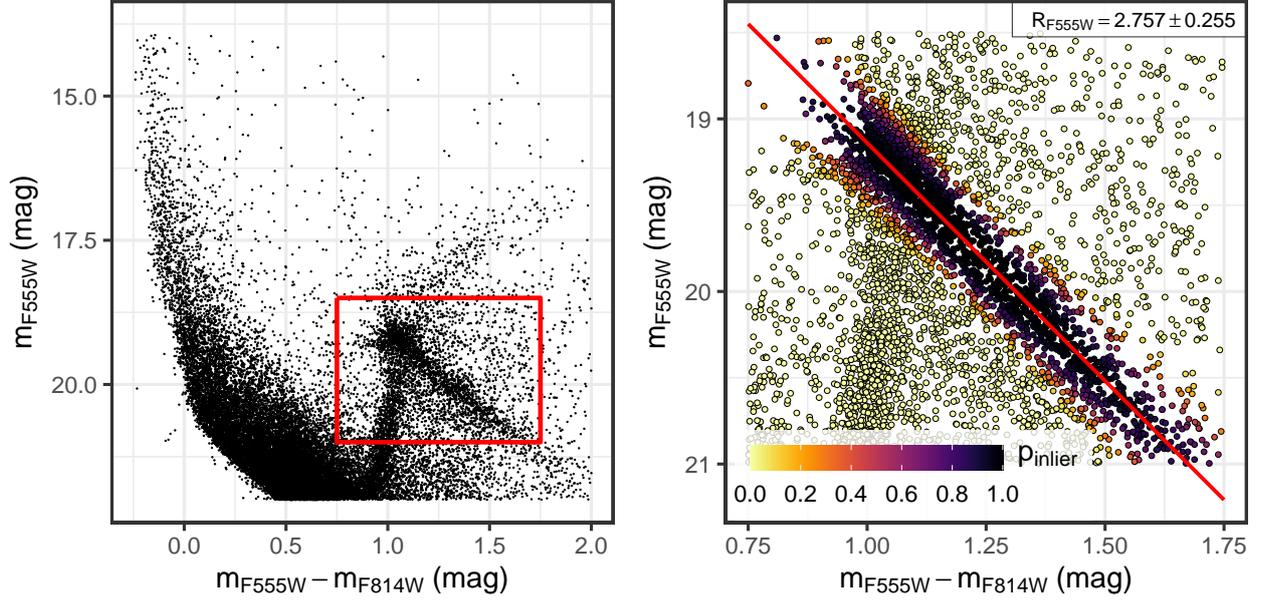
Figure 7. Left: zoom-in on the bright end of the optical CMD of the MYSST N44 data. Highlighted in red is the region used to derive the optical extinction law of N44 from the slope of the RC elongation. Right: zoom-in on the selected candidate RC region from the left panel. The stars are color-coded according to the fraction of times they were considered as an RC inlier across the 5000 RANSAC runs. The red line indicates the averaged resulting slope across all RANSAC fits. For the reddening vector in F555W we find a value of $R_{F555W} = 2.8 \pm 0.3$.
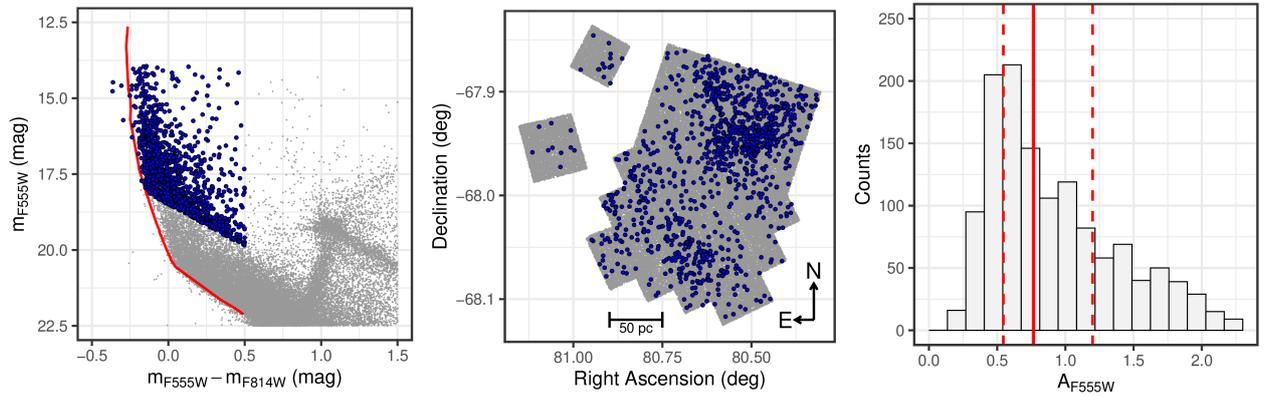
Figure 8. Left: zoom-in on the optical CMD of N44. Highlighted in blue are UMS stars selected to estimate individual stellar extinctions. The red line indicates the position of the ZAMS, corrected for the LMC distance modulus and MW foreground extinction, which serves as the target position for the UMS extinction measurement. Middle: spatial distribution of the MYSST photometric catalog of N44. Highlighted in blue are the positions of our selected UMS extinction probes. Right: histogram of the extinction measurements (including the MW foreground) in F555W of the UMS probes. The solid red line indicates the median extinction of 0.77 mag, while the dashed red lines mark the 25% and 75% quantiles.
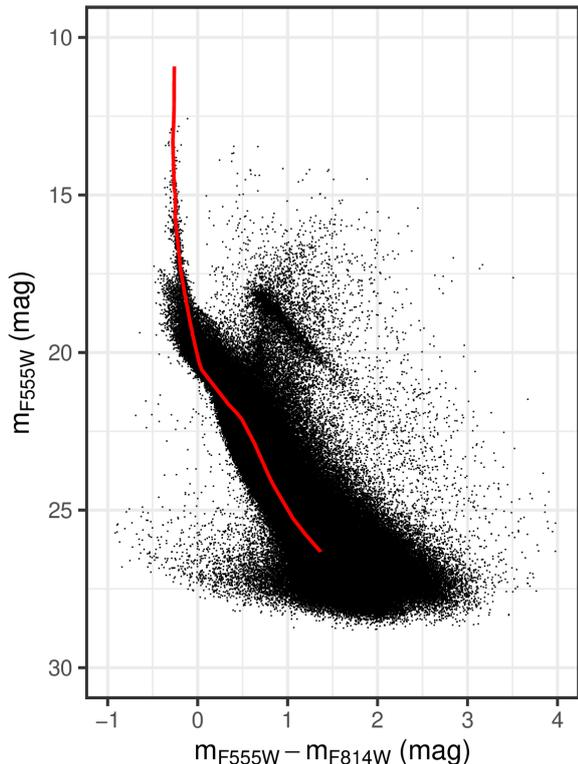
Figure 9. Optical CMD of the N44 data corrected for extinction. Each star is corrected by the distance-weighted average extinction of their 20 nearest UMS neighbors. For visualization purposes the extinction of each UMS star is sampled within its measurement uncertainty. For reference the red line indicates the ZAMS, corrected for the LMC distance and MW foreground extinction, that is used to determine the extinction of the UMS probes.

ical optical CMD location, assuming that they should be on the ZAMS. Note that we derive the ZAMS locus from PARSEC isochrones (Bressan et al. 2012) with a metallicity of $Z = 0.008$ for the LMC and correct the ZAMS position for Milky Way (MW) foreground extinction prior to the measurement. Here we adopt a value of $A_V^{mw} = 0.22$ mag ($A_{F555W}^{mw} \approx 0.223$ mag) toward the LMC (De Marchi & Panagia 2014) and the Cardelli et al. (1989) Galactic reddening law. The measured value and the 0.223 mag MW foreground offset are then summed up to provide the total extinction.

For the extinction measurements we make a selection of the brightest UMS stars in the MYSST photometric catalog, accounting for the slope of the reddening vector and field contamination. This selection, depicted in the left panel of Figure 8, consists of 1291 stars and represents a trade-off between retaining enough sources for good statistics and minimizing potential contamination

from old field sources. Scaling the source density estimated in the reference fields to our CMD, we find that our selection criterion entails about 15% field contamination, i.e. specifically we expect 194 ± 14 field stars in our sample. For more details on the UMS selection, see Appendix A. As previously mentioned, due to saturation issues, our selection of UMS stars is likely missing some of the most massive objects of N44. A comparison with the ZAMS, shifted along the reddening vector, puts our UMS sources in an approximate mass range between ~ 6 and $\sim 30\,M_\odot$, indicating late O- to early B-type stars. Note that there are 13 UMS stars in our selection with a CMD position that falls to the left of the target, foreground-corrected, ZAMS locus. Given their close proximity to the ZAMS, our procedure assumes that these sources have zero LMC extinction, so they are only subject to the MW foreground.

The middle panel of Figure 8 marks the positions of our UMS extinction probes in relation to the rest of the survey. As we can see, we find most of our UMS stars in and around the superbubble of N44 located in the northern part of the survey. With the fewer available probes in the middle and southern part of the main FOV, as well as the very few UMS stars within the control fields, extinction estimates will be less precise there.

For the extinction measurement and assignment to the non-UMS stars we use a modified version of our approach presented in Ksoll et al. (2018). In order to give a measurement uncertainty for the UMS extinctions, we now sample the slope of the reddening vector within its error and derive the mean and standard deviation of the so-measured extinction values for each UMS star. Consequently, we now derive the error for the distance-weighted average extinction assigned to the non-UMS stars by propagation of the uncertainty of the UMS measurements

$$\delta A_{F555W} = \sqrt{\sum_{i=1}^{20} \left(w_i \delta A_{F555W}^{UMS} \right)^2}, \qquad (6)$$

with weights

$$w_i = \frac{1}{d_i^2 + \epsilon^2} \frac{1}{\sum_{n=1}^{20} (1/d_n^2 + \epsilon^2)}, \qquad (7)$$

where d_i is the Euclidean distance to the ith-nearest UMS neighbor in pixels and ϵ is a smoothing factor (for more details see Ksoll et al. 2018).

The distribution of the final F555W extinction measurements for the UMS stars is summarized in the right panel of Figure 8. With an overall median extinction of $0.77^{+0.42}_{-0.23}$ mag, it appears that the UMS population of N44 is for the most part only moderately attenuated.
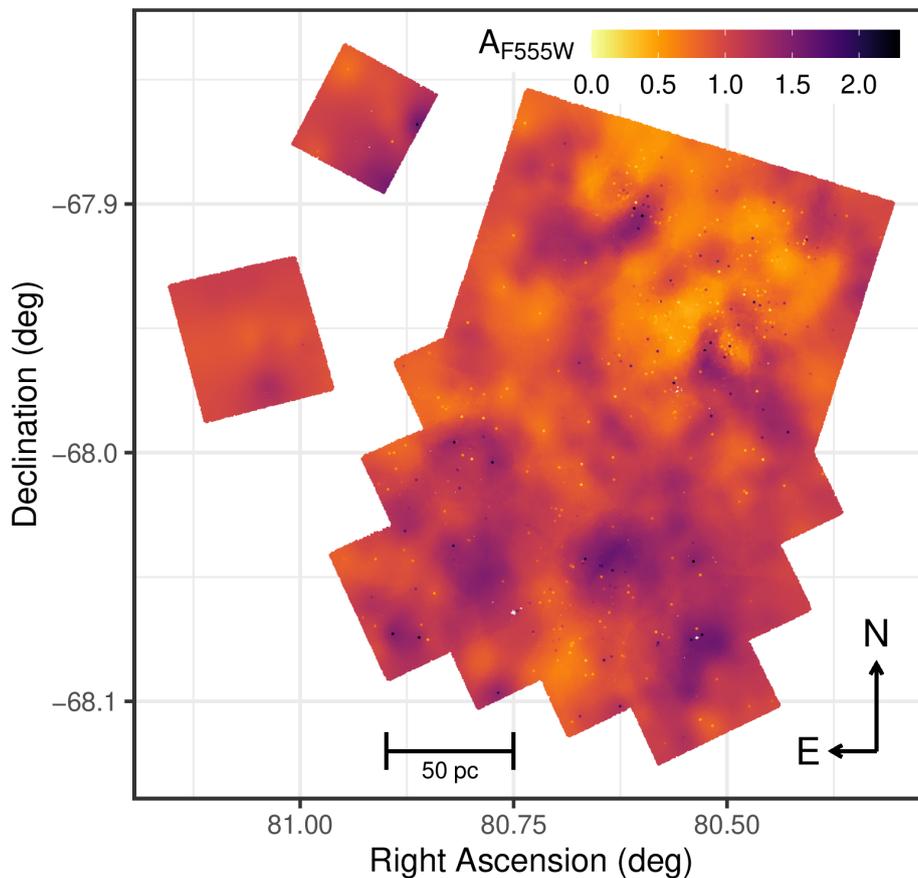
Figure 10. Spatial distribution diagram of the MYSST photometric catalog. Each star is color-coded according to the assigned distance-weighted average extinction of its 20 nearest UMS neighbors. The UMS extinction probes themselves appear as distinctly colored points in this diagram, as they are not subject to the smoothing effect of our assignment procedure for the other stars. Additionally, any white spots are simply caused by a lack of sources, as this plot shows every individual star of the catalog.

This is consistent with the fact that a notable fraction of our UMS selection is located inside of the superbubble of N44, where feedback from the most massive stars has cleared out substantial amounts of gas. Still there are about 200 UMS stars in our sample that exhibit more than 1.5 mag of extinction up to a maximum of 2.29 mag, indicating the presence of regions that are subject to substantial reddening.

For the assignment to the non-UMS sources we follow our findings in Ksoll et al. (2018) and use the distance-weighted average extinction of the 20 nearest UMS neighbors, employing a smoothing factor of $\epsilon = 500\,\mathrm{px}$ ($20''$, $\sim 5\,\mathrm{pc}$). Figure 9 displays the CMD of the MYSST photometric catalog when corrected for the assigned extinction values in comparison to the ZAMS used to measure the extinction for the UMS stars. As we have elab-

orated in Ksoll et al. (2018), this method of extinction estimation is not ideal for all types of stars, since we find some over- and underestimation for, e.g. field MS and RC stars. The main issue here is that these objects are likely foreground and background objects of the young star-forming clusters that are the primary target of this survey. While recent studies (Cignoni et al. 2015) provide strong evidence that young PMS objects tend to cluster around young massive UMS stars, so that this form of extinction estimate works well for young PMS objects (see also De Marchi et al. 2016), no such tendency for spatial colocation is given for the field contaminants. Therefore, the UMS extinction probes are less representative for these objects, and the extinction estimate is less precise (De Marchi et al. 2011, 2017).

There is an additional caveat with our UMS extinction estimate to mention here. Our core assumption is that all UMS stars are on the ZAMS. Given the fast evolution of very massive stars, this might not necessarily be true, and their actual position could be slightly different from the ZAMS, even if they are only a few megayears old. Oey & Massey (1995) find that the UMS stars interior to N44's superbubble are likely about 10 Myr old, while the ones located at the western rim of the bubble are 5 Myr old. Isochrones corresponding to these ages deviate to colors redder than the ZAMS in the high-mass regime. In these cases, dereddening to the ZAMS, i.e. extending the vector beyond the correct isochrone, will overestimate both extinction and mass. Without additional information one cannot date the UMS stars more accurately across the entire MYSST FOV. However, due to the saturation limit of the MYSST survey, the number of stars where a significant difference between the ZAMS and the actual stellar age might occur is relatively small. Taking, for instance, a 10 Myr isochrone instead of the ZAMS, as appropriate for the bubble interior, only 276 stars out of our 1291 UMS sources are massive enough to be affected. For the UMS stars located in the bubble and western rim we find a median absolute error of only $0.043^{+0.019}_{-0.014}$ mag in extinction. Further details on this error estimate are provided in Appendix B. With this caveat in mind, our ZAMS assumption allows the derivation of a self-consistent (relative to the MYSST data) extinction estimate that in some cases may just provide an upper limit to the true value.

Besides the age of the UMS, there are other effects, e.g. unresolved binarity or metallicity gradients, that can induce a broadening of the UMS in the CMD even in the absence of extinction. While often not considered for star clusters, there are findings that could support a potential metallicity gradient in N44. As previously mentioned, N44 has seen at least two known episodes of SF (Oey & Massey 1995). Additionally, a supernova remnant, SNR 0523-679 (Chu et al. 1993), that exhibits characteristics of a core-collapse supernova (Jaskot et al. 2011) is present within N44. Oey & Massey (1995) also estimate that up to four supernovae occurred in the region in the past. Lastly, Jaskot et al. (2011) find some evidence for metallicity enhancement in N44's superbubble. Thus, pollution of the formation environment of the younger population by one or multiple supernovae from the previous SF event is a possibility. We investigate the impact of these effects on our extinction estimation procedure in Appendix C.

As a final test we also look into the potential extinction error induced by neglecting stellar rotation. As mentioned in Section 3, rotation may induce additional color excess, in particular for very massive stars. To evaluate the effect, we use the MIST (Dotter 2016; Choi et al. 2016; Paxton et al. 2011, 2013, 2015) stellar evolution models (with $Z = 0.008$) to construct two ZAMS loci, one for the rotating ($v/v_{\rm crit} = 0.4$) and one for the nonrotating case. Here we find that both ZAMS loci (and e.g. the 5 and 10 Myr isochrones too) are practically identical, such that the expected extinction error caused by moderate stellar rotation is negligible for our approach. Using the SYCLIST[9] stellar evolution models (Georgy et al. 2013) we also investigate the error induced in our approach for extremely fast rotating stars, i.e. $v/v_{\rm crit} = 0.95$, in comparison to a nonrotating model. In this test we find a median absolute extinction error between the two models of $0.041^{+0.022}_{-0.017}$ mag. Consequently, the effect of rotation on our extinction estimate is minimal even for fast rotators. We have to note, however, that the SYCLIST models for $v/v_{\rm crit} = 0.95$ are only available up to a stellar mass of 15 $M_\odot$, such that the brightest 276 UMS stars in our selection could not be considered in this test. In addition, we had to employ models with $Z = 0.006$ as the closest readily available metallicity to our adapted LMC value of $Z = 0.008$.

Figure 10 shows the extinction map we derive from our estimates by color-coding each star according to its assigned value. As we can see, the region of the superbubble is indeed subject to the least amount of reddening, while we find the most extinguished areas at the western edge of the bubble, as well as predominantly in the southern part of the observed FOV. The rather prominent dark "filament", extending from the southern bubble edge to the southwest corner of the FOV, does not have a significant counterpart in the longer-wavelength observations of N44 taken with Spitzer. Pointing to relatively low extinction values, this might suggest that this feature is actually an artifact of our ZAMS assumption. We verified that there are very few UMS stars in this "filament" that would experience a reduction in extinction measure if they were dereddened to a 10 Myr isochrone instead of the ZAMS (see Figure B1 in Appendix B). Therefore, a significant systematic error could be justified only if these UMS stars are even older than 10 Myr.

It should also be noted that the extinction of the two control fields is likely not particularly precise, due to the limitations of our approach listed above and the fact that these field stars are not likely to be colocated with the few selected UMS stars within these regions.

---

[9] https://www.unige.ch/sciences/astro/evolution/en/database/syclist/

4.3. *RC extinction*

To further characterize the extinction pattern of N44, we also determine extinction measures for the RC stars. As previously mentioned, the nonextinguished RC is usually a prominent, almost circular feature in the CMD, which is smeared out in the MYSST observation due to differential reddening. Still, the tip of this smeared-out feature marks the nominal nonextinguished position of the RC. Consequently, measuring extinction for the RC sources is also straightforward. To quantify the position of the RC tip, we use a kernel density estimate (KDE) in the CMD space (using separate bandwidths for magnitude and color, each estimated by Silverman's rule) and identify the significant overdensity in the vicinity of the tip. In particular, we find a density contour at 4σ significance (above the mean density) that traces the nonextinguished end of the RC (blue contour in the left panel of Figure 11). We then determine the nominal RC position as a line perpendicular to the reddening vector anchored at the intersection point between the 4σ contour and the reddening vector going through its center. In practice, we vary the reddening slope within its uncertainty, determining a new target line for each sample slope, but keeping the same anchor point, in order to provide measurement errors for the RC probes. For comparison, the left panel of Figure 11 also indicates the theoretical position of the RC for the LMC field ($m_{\mathrm{F555W}} = 19.16$, $m_{\mathrm{F814W}} = 18.1$, corrected for distance modulus and foreground MW extinction) as determined by De Marchi & Panagia (2014) for the HST filters used in this survey. Here we find an excellent agreement with our empirically determined nominal RC position. This match allows us to easily transform these relative RC extinction measurements to total extinction values by correcting for the Milky Way extinction contribution of $A_{\mathrm{V}}^{\mathrm{mw}} = 0.22$ mag, translating to $A_{\mathrm{F555W}}^{\mathrm{mw}} \approx 0.223$ mag, assumed for the theoretical RC position in De Marchi & Panagia (2014).

From our RANSAC inlier determination, we select the bona fide RC constituents as those that reach an inlier fraction above 50%, i.e. the stars that are chosen as RC inliers at least half of the time across all 5000 RANSAC runs. This returns a sample of 1737 RC stars, which appear to be almost uniformly distributed across the MYSST FOV (see Figure 11, middle panel, red points). As these objects are most likely foreground and background field sources and not part of the star-forming clusters in N44, this is to be expected. Note that there are 145 stars in our RC inlier selection that fall above our target RC position. Given their close proximity to the target, our measurement approach assumes that these

sources are only subject to the MW foreground extinction.

The right panel of Figure 11 shows the outcome of the extinction measurements for the RC probes. With a median total extinction of $0.68^{+0.52}_{-0.32}$ mag, RC stars are also overall only subject to moderate reddening. In comparison to the UMS sources, the RC extinction distribution appears fairly similar, being only about 0.1 mag less extinguished on average. This slight difference could be due to the UMS sources being likely embedded within the star-forming centers of N44, whereas the mostly uniformly distributed RC field stars are not as obscured by N44's gas reservoirs. But with a maximum of 2.1 mag extinction, there are also a few hundred RC sources that are affected by more severe reddening. Note, however, that this maximum is certainly affected by our initially selected CMD region for the RANSAC procedure. It is possible that a few more heavily extinguished RC objects were excluded by this selection, so that this upper extinction limit should not be treated as an absolute maximum.

To make a direct spatial comparison to the UMS extinction map, we cannot follow the same procedure as for the UMS stars in Section 4.2. Assigning distance-weighted average values of the nearest RC neighbors to the other stars has little meaning, since these RC field stars are very unlikely to have a spatial correlation to the N44 constituents beyond projection effects. Being predominantly part of the LMC field population, the RC sources are, however, more likely to provide a representative extinction measure for other field sources, such as the many LMS stars captured in the MYSST FOV. To perform a spatial comparison, we instead use a 2D binning approach, computing the average measured extinction of the RC sources in square 41"×41" (11 pc×11 pc) spatial bins. The resulting low-resolution RC extinction map is shown in the left panel of Figure 12. The most extincted RC stars are located toward the eastern and northern edge of the superbubble as well as in more compact regions south of the bubble. The right panel of Figure 12 provides a direct comparison between the 2D binned RC mean extinction map and a corresponding map derived from the UMS sources, showing the difference $\Delta \bar{A}_{\mathrm{F555W}}^{\mathrm{UMS-RC}} = \bar{A}_{\mathrm{F555W}}^{\mathrm{UMS}} - \bar{A}_{\mathrm{F555W}}^{\mathrm{RC}}$ in each bin. Positive values in this map indicate where the mean extinction inferred from the UMS stars is larger, while negative values imply the opposite. As we can see, the UMS extinction exceeds that of the RC mostly in the southern half of the FOV. Notable regions where the mean RC extinction is larger are located at the eastern and northern edge of the bubble, as well as in several compact patches in the south of the FOV. These RC stars
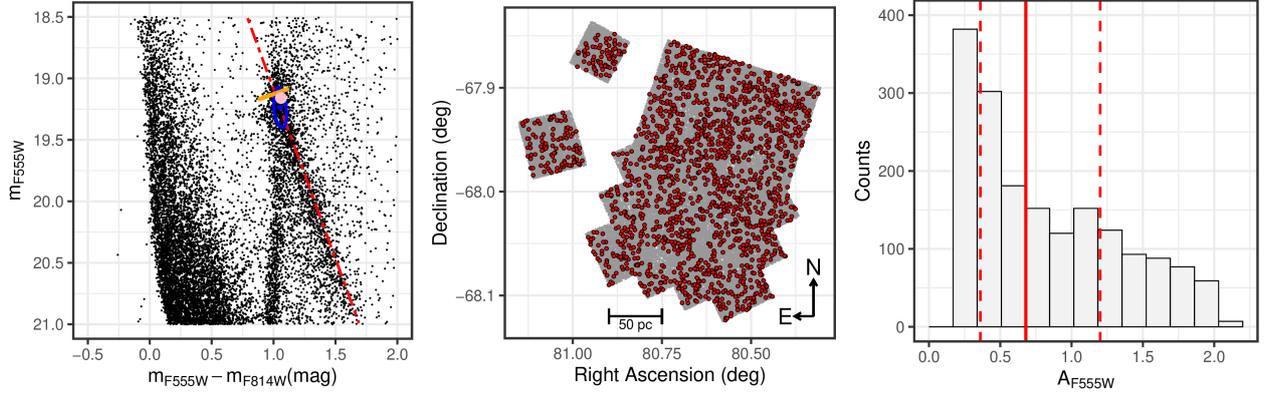
Figure 11. Left: zoom-in on the RC in the optical CMD of the MYSST catalog. The blue line represents a 4σ KDE density contour, marking the tip of the RC and indicating the nominal nonextinguished RC position in the CMD. The red dashed line marks the reddening vector quantified with the RANSAC approach. The solid orange line indicates the target true position, i.e. the tip of the RC, used to determine the extinction of the RC probes. It is perpendicular to the reddening vector and anchored to the intersection of the reddening vector and the 4σ density contour. For comparison, the pink point indicates the theoretical RC position for the LMC field as determined by De Marchi & Panagia (2014), demonstrating the excellent agreement with our empirically determined position. Middle: spatial distribution of the RC constituents (red points) identified with the RANSAC procedure (i.e. all sources with an inlier probability above 0.5). Right: histogram of the measured extinction values in F555W of the RC probes. The solid red line indicates the median extinction of 0.68 mag, while the dashed lines mark the 25% and 75% quantiles.
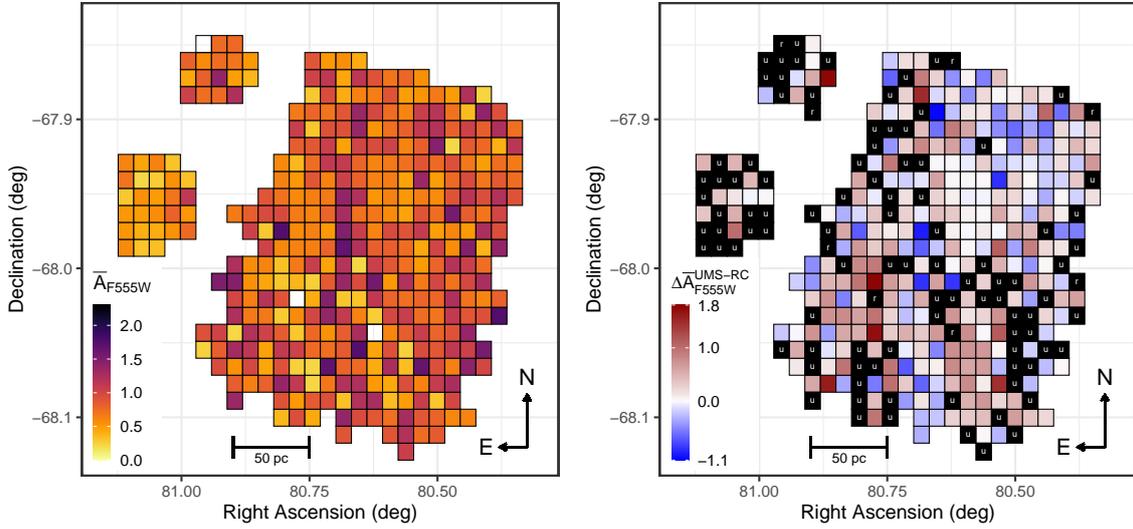
Figure 12. Left: 2D binned extinction map of the RC sources. Each 41" $\times$ 41" ($11\,\mathrm{pc} \times 11\,\mathrm{pc}$) bin is colored according to the average measured extinction of the RC sources located inside. White tiles indicate bins in which no RC stars are found. Right: The same 2D bin diagram as in the left panel, but now each bin is colored according to the difference $\Delta \bar{A}_{\mathrm{F555W}}^{\mathrm{UMS-RC}} = \bar{A}_{\mathrm{F555W}}^{\mathrm{UMS}} - \bar{A}_{\mathrm{F555W}}^{\mathrm{RC}}$ of the mean extinction in each bin between the values derived from UMS and RC extinction probes. A positive value indicates a larger mean extinction derived from the UMS sources, while a negative one implies that the RC sources experience more extinction on average. The black tiles indicate bins where the difference $\Delta \bar{A}_{\mathrm{F555W}}^{\mathrm{UMS-RC}}$ cannot be computed because either no UMS ("u") or no RC ("r") sources are present.

are likely located behind the star-forming gas of N44. On the other hand, in the regions where the mean UMS extinction dominates it is possible that background RC sources are simply not detected owing to the obscuring gas that the UMS sources are embedded in. Interestingly, we find that the RC and UMS extinction agrees very well inside N44's bubble. This provides further confirmation that the feedback of the massive stars interior to the bubble has cleared out almost all of the gas, such that barely any local obscuration is left. It is necessary to note, though, that the difference between UMS and RC extinction might be affected by stochastic variability in the number of RC sources that are in the foreground and background in a given 2D bin in the right panel of Figure 12.

5. SUMMARY

In this paper we introduce the new HST Treasury Program "Measuring Young Stars in Space and Time" (MYSST), which captures the active star-forming complex N44 with its rich collection of H II regions, young stellar clusters, and bubbles, located in the LMC. We present the observing strategy of MYSST, describe our data reduction procedure, and construct the photometric catalog of the survey. In addition, we highlight our first scientific results, briefly discussing the stellar populations found across N44 and determining the extinction properties of the region. On top of that, we infer extinction maps for N44 from reddening measurements of UMS and RC stars.

The MYSST survey observed N44 in the optical wavelength regime using the F555W and F814W broadband filters of both the ACS and WFC3 imagers on board the HST. Combining PSF-fitting and aperture photometry, the latter being needed to recover saturated bright sources, we compile a photometric catalog that comprises 461,684 stars across the MYSST FOV, going down as deep as 29 mag in F555W and 28 mag in F814W, probing even the lowest-mass stellar population of N44 (e.g. down to $0.09\,M_\odot$ for an unreddened 1 Myr PMS star). Due to saturation effects, the catalog does not contain sources brighter than 14 mag in F555W and 13 mag in F814W, likely missing the most massive O-type stars of the region. Due to stellar crowding, background, and saturation, the completeness of the catalog varies across the FOV, but reaches an excellent average of 26.7 (28.1) mag in F555W and 25.7 (26.7) mag in F814W at the 80% (50%) level.

The rich photometric catalog reveals many different stellar populations spread across the MYSST FOV. We identify numerous old LMS and RGB sources that are almost uniformly distributed across N44, likely foreground

and background contaminants belonging to the LMC field population. We also find young high-mass UMS and lower-mass PMS stars within the survey, which exhibit clustered spatial distributions, tracing e.g. the gaseous rim of N44's characteristic superbubble. These young stars mark N44's numerous active star-forming centers.

To constrain the reddening properties of N44, we measure the slope of the RC feature that appears elongated in the CMD owing to differential extinction. Here we present a new approach to jointly establish the constituents of this elongated RC feature and perform the fit of the reddening vector by applying the well-established learning algorithm RANSAC (Fischler & Bolles 1981). This algorithm is a very robust tool to fit models to data in the presence of outliers. RANSAC is an iterative process, where in each iteration first a minimal subset of the data, large enough to fit the given model, is drawn randomly from the total data. Then, the model fit is performed on that subset, and finally the number of data points within the complete data set are determined that are inliers to the fitted model. These steps are then repeated until an optimal model is found, at which point a final fit to the inliers of this model is performed for further refinement.

Selecting a window within the optical CMD of the MYSST data in which the RC is the predominant feature, we apply the RANSAC algorithm, repeating it 5000 times to negate all effects of the random seed, to determine the RC constituents and fit a line to the elongated red clump to derive the slope of the reddening vector. With this approach we find the total-to-selective extinction ratios $R_{(\mathrm{F555W-F814W})}(\mathrm{F555W}) = 2.8 \pm 0.3$ and $R_{(\mathrm{F555W-F814W})}(\mathrm{F814W}) = 1.8 \pm 0.3$ for the slope of the reddening vector in N44. These results are notably larger than the values for the standard galactic extinction law with $R_\mathrm{V} = 3.1$, returning 2.4 and 1.4 (Cardelli et al. 1989) or 2.2 and 1.2 (Fitzpatrick 1999), respectively.

With the reddening vector constrained we select a set of UMS stars as probes and measure their extinction by reprojecting their position in the CMD back to their theoretical location, assuming that they should be on the ZAMS. Afterward, we assign each non-UMS star in the MYSST photometric catalog a distance-weighted average extinction of their 20 nearest UMS neighbors. This procedure has been found to return reasonable extinction estimates for the constituents of young star-forming clusters (De Marchi et al. 2016), such as the ones we are aiming to find here, but suffers from occasional extinction over- or underestimation for field constituents. Additionally, this approach might overestimate extinc-

tion for older ($> 10\,\mathrm{Myr}$) very massive UMS sources, as their true position might slightly deviate from the ZAMS, such that at worst the estimate provides only an upper limit to the true extinction. For a subset of UMS stars with known ages inside and at the rim of N44's bubble we find, however, only a median absolute error of $0.043^{+0.019}_{-0.014}$ mag with our approach compared to using the correct ages. We make the ZAMS assumption since the UMS ages are not easily recovered across the entire FOV and because it allows a MYSST self-consistent extinction estimate that entails the same systematic error everywhere. Note that we plan to provide more precise extinction measures in a follow-up study that explores synergies with other observations of N44 (e.g. Gaia).

Following the assignment, we present an extinction map for N44 based on the measured UMS extinction. With a median extinction of $0.77^{+0.42}_{-0.23}$ mag in F555W, it appears that the UMS population of N44 is overall only moderately extincted. With about 200 UMS probes, though, exceeding 1.5 mag up to a maximum of 2.29 mag in extinction, there are still a notable number of regions subject to more severe reddening. Our extinction map confirms that the reddening of N44 is patchy and highly differential across the MYSST FOV.

For comparison we also compile a 2D binned average extinction map derived from measurements of RC stars. Showing an overall median extinction of $0.68^{+0.52}_{-0.32}$ mag, the RC stars across the MYSST FOV are similarly reddened to the UMS population. There are a few hundred of our total of $\sim$1700 RC extinction probes that also exhibit more severe reddening up to a maximum of 2.1 mag. This is, however, not an absolute maximum, as our RANSAC approach may have excluded a few heavily extinguished RC stars. A direct spatial comparison between a UMS and RC 2D binned mean extinction map reveals that the UMS sources tend to be more reddened across most of the southern half of the FOV. Notable areas where RC extinction exceeds the UMS values are the eastern and northern edge of the N44 bubble, as well as a few compact patches south of the bubble.

In conclusion, the MYSST survey provides an extraordinary view of extragalactic SF across an entire giant star-forming complex that highlights the complex interplay between high-mass stellar feedback and star-forming events. With its high resolution and deep photometry, it provides the opportunity to study length scales and timescales of the SF process on the scale of a giant molecular cloud. In our subsequent study in Pa-

per II we begin to quantify the SF history of N44 by identifying its rich PMS stellar content and analyze the complex clustering behavior of the young, still-forming PMS stars across N44.

6. ACKNOWLEDGEMENTS

Software: Astrodrizzle (Hack et al. 2012), DOLPHOT (v2.0 Dolphin 2000), PARSEC (Bressan et al. 2012), MIST (Dotter 2016; Choi et al. 2016; Paxton et al. 2011, 2013, 2015), SYCLIST (Georgy et al. 2013)

REFERENCES

Adamo, A., Kruijssen, J. M. D., Bastian, N., Silva-Villa, E., & Ryon, J. 2015, MNRAS, 452, 246, doi: 10.1093/mnras/stv1203

Ballesteros-Paredes, J., Klessen, R. S., Mac Low, M.-M., & Vazquez-Semadeni, E. 2007, in Protostars and Planets V, ed. B. Reipurth, D. Jewitt, & K. Keil (University of Arizona Press), 63–80

Bastian, N., Covey, K. R., & Meyer, M. R. 2010, ARA&A, 48, 339, doi: 10.1146/annurev-astro-082708-101642

Bik, A., Henning, T., Stolte, A., et al. 2012, ApJ, 744, 87, doi: 10.1088/0004-637X/744/2/87

Bisbas, T. G., Wünsch, R., Whitworth, A. P., Hubber, D. A., & Walch, S. 2011, ApJ, 736, 142, doi: 10.1088/0004-637X/736/2/142

Bonnell, I. A., Bate, M. R., Clarke, C. J., & Pringle, J. E. 1997, MNRAS, 285, 201

Bonnell, I. A., Clarke, C. J., Bate, M. R., & Pringle, J. E. 2001, MNRAS, 324, 573, doi: 10.1046/j.1365-8711.2001.04311.x

Bonnell, I. A., Smith, R. J., Clark, P. C., & Bate, M. R. 2011, MNRAS, 410, 2339, doi: 10.1111/j.1365-2966.2010.17603.x

Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127, doi: 10.1111/j.1365-2966.2012.21948.x

Brott, I., de Mink, S. E., Cantiello, M., et al. 2011, A&A, 530, A115, doi: 10.1051/0004-6361/201016113

Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245, doi: 10.1086/167900

Carlson, L. R., Sewiło, M., Meixner, M., Romita, K. A., & Lawton, B. 2012, A&A, 542, A66, doi: 10.1051/0004-6361/201118627

Chabrier, G. 2003, PASP, 115, 763, doi: 10.1086/376392

Chen, C. H. R., Chu, Y.-H., Gruendl, R. A., Gordon, K. D., & Heitsch, F. 2009, ApJ, 695, 511, doi: 10.1088/0004-637X/695/1/511

Chevance, M., Kruijssen, J. M. D., Vazquez-Semadeni, E., et al. 2020, SSRv, 216, 50, doi: 10.1007/s11214-020-00674-x

Choi, J., Dotter, A., Conroy, C., et al. 2016, ApJ, 823, 102, doi: 10.3847/0004-637X/823/2/102

Chu, Y.-H., Mac Low, M.-M., Garcia-Segura, G., Wakker, B., & Kennicutt, Robert C., J. 1993, ApJ, 414, 213, doi: 10.1086/173069

Cignoni, M., Sabbi, E., Nota, A., et al. 2009, AJ, 137, 3668, doi: 10.1088/0004-6256/137/3/3668

Cignoni, M., Sabbi, E., van der Marel, R. P., et al. 2015, ApJ, 811, 76, doi: 10.1088/0004-637X/811/2/76

—. 2016, ApJ, 833, 154, doi: 10.3847/1538-4357/833/2/154

Clark, P. C., Glover, S. C. O., Klessen, R. S., & Bonnell, I. A. 2012, MNRAS, 424, 2599, doi: 10.1111/j.1365-2966.2012.21259.x

Dalcanton, J. J., Fouesneau, M., Hogg, D. W., et al. 2015, ApJ, 814, 3, doi: 10.1088/0004-637X/814/1/3

Dale, J. E., Ercolano, B., & Bonnell, I. A. 2013, MNRAS, 430, 234, doi: 10.1093/mnras/sts592

De Marchi, G., & Panagia, N. 2014, MNRAS, 445, 93, doi: 10.1093/mnras/stu1694

De Marchi, G., Panagia, N., & Beccari, G. 2017, ApJ, 846, 110, doi: 10.3847/1538-4357/aa85e9

De Marchi, G., Panagia, N., & Girardi, L. 2014, MNRAS, 438, 513, doi: 10.1093/mnras/stt2233

De Marchi, G., Panagia, N., & Milone, A. P. 2020, ApJ, 899, 114, doi: 10.3847/1538-4357/aba834

De Marchi, G., Paresce, F., Panagia, N., et al. 2011, ApJ, 739, 27, doi: 10.1088/0004-637X/739/1/27

De Marchi, G., Panagia, N., Sabbi, E., et al. 2016, MNRAS, 455, 4373, doi: 10.1093/mnras/stv2528

Dib, S., & Henning, T. 2019, A&A, 629, A135, doi: 10.1051/0004-6361/201834080

Dolphin, A. E. 2000, PASP, 112, 1383, doi: 10.1086/316630

Dotter, A. 2016, ApJS, 222, 8, doi: 10.3847/0067-0049/222/1/8

Elmegreen, B. G., & Lada, C. J. 1977, ApJ, 214, 725, doi: 10.1086/155302

Federrath, C. 2013, MNRAS, 436, 3167, doi: 10.1093/mnras/stt1799

Federrath, C., & Klessen, R. S. 2012, ApJ, 761, 156, doi: 10.1088/0004-637X/761/2/156

Fischler, M., & Bolles, R. 1981, Random sample consensus (ACM), 381–395

Fitzpatrick, E. L. 1999, PASP, 111, 63, doi: 10.1086/316293

Forsyth, D., & Ponce, J. 2003, Computer vision, Prentice Hall series in artificial intelligence (Upper Saddle River, NJ: Prentice Hall), XXV, 693 S.

Fukui, Y., Mizuno, N., Yamaguchi, R., Mizuno, A., & Onishi, T. 2001, PASJ, 53, L41, doi: 10.1093/pasj/53.6.L41

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1, doi: 10.1051/0004-6361/201833051

Gennaro, M., Goodwin, S. P., Parker, R. J., Allison, R. J., & Brandner, W. 2017, MNRAS, 472, 1760, doi: 10.1093/mnras/stx2098

Gennaro, M., Bik, A., Brandner, W., et al. 2012, A&A, 542, A74, doi: 10.1051/0004-6361/201218980

Georgy, C., Ekström, S., Granada, A., et al. 2013, A&A, 553, A24, doi: 10.1051/0004-6361/201220558

Getman, K. V., Feigelson, E. D., Kuhn, M. A., et al. 2018, MNRAS, 476, 1213, doi: 10.1093/mnras/sty302

Girardi, L., & Salaris, M. 2001, MNRAS, 323, 109, doi: 10.1046/j.1365-8711.2001.04084.x

Girichidis, P., Federrath, C., Allison, R., Banerjee, R., & Klessen, R. S. 2012, MNRAS, 420, 3264, doi: 10.1111/j.1365-2966.2011.20250.x

Girichidis, P., Federrath, C., Banerjee, R., & Klessen, R. S. 2011, MNRAS, 413, 2741, doi: 10.1111/j.1365-2966.2011.18348.x

Girichidis, P., Offner, S. S. R., Kritsuk, A. G., et al. 2020, SSRv, 216, 68, doi: 10.1007/s11214-020-00693-8

Gordon, K. D., Clayton, G. C., Misselt, K. A., Land olt, A. U., & Wolff, M. J. 2003, ApJ, 594, 279, doi: 10.1086/376774

Gouliermis, D. A., Henning, T., Brand ner, W., et al. 2007, ApJL, 665, L27, doi: 10.1086/521224

Gouliermis, D. A., Schmeja, S., Ossenkopf, V., Klessen, R. S., & Dolphin, A. E. 2014, in The Labyrinth of Star Formation, Vol. 36 (Springer International Publishing), 447, doi: 10.1007/978-3-319-03041-8_89

Gouliermis, D. A., Dolphin, A. E., Robberto, M., et al. 2011, ApJ, 738, 137, doi: 10.1088/0004-637X/738/2/137

Grasha, K., Calzetti, D., Adamo, A., et al. 2019, MNRAS, 483, 4707, doi: 10.1093/mnras/sty3424

Hack, W. J., Dencheva, N., Fruchter, A. S., et al. 2012, in American Astronomical Society Meeting Abstracts, Vol. 220, American Astronomical Society Meeting Abstracts #220, 135.15

Hartmann, L., Ballesteros-Paredes, J., & Bergin, E. A. 2001, ApJ, 562, 852, doi: 10.1086/323863

Haschke, R., Grebel, E. K., & Duffau, S. 2011, AJ, 141, 158, doi: 10.1088/0004-6256/141/5/158

Henize, K. G. 1956, ApJS, 2, 315, doi: 10.1086/190025

Hennebelle, P. 2018, A&A, 611, A24, doi: 10.1051/0004-6361/201731071

Hony, S., Galliano, F., Madden, S. C., et al. 2010, A&A, 518, L76, doi: 10.1051/0004-6361/201014628

Hony, S., Gouliermis, D. A., Galliano, F., et al. 2015, MNRAS, 448, 1847, doi: 10.1093/mnras/stv107

Jaskot, A. E., Strickland, D. K., Oey, M. S., Chu, Y. H., & García-Segura, G. 2011, ApJ, 729, 28, doi: 10.1088/0004-637X/729/1/28

Kennicutt, R. C., J., & Hodge, P. W. 1986, ApJ, 306, 130, doi: 10.1086/164326

Klessen, R. S., & Glover, S. C. O. 2016, Saas-Fee Advanced Course, 43, 85, doi: 10.1007/978-3-662-47890-5_2

Klessen, R. S., Heitsch, F., & Mac Low, M.-M. 2000, ApJ, 535, 887, doi: 10.1086/308891

Krist, J. E., Hook, R. N., & Stoehr, F. 2011, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8127, Optical Modeling and Performance Predictions V, 81270J, doi: 10.1117/12.892762

Kroupa, P. 2002, Science, 295, 82, doi: 10.1126/science.1067524

Krumholz, M. R., & Tan, J. C. 2007, ApJ, 654, 304, doi: 10.1086/509101

Ksoll, V. F., Gouliermis, D. A., Klessen, R. S., et al. 2018, MNRAS, 479, 2389, doi: 10.1093/mnras/sty1317

Ksoll, V. F., Gouliermis, D., Sabbi, E., et al. 2020, arXiv e-prints, arXiv:2012.00524. https://arxiv.org/abs/2012.00524

Lucke, P. B., & Hodge, P. W. 1970, AJ, 75, 171, doi: 10.1086/110959

Mac Low, M.-M., & Klessen, R. S. 2004, Rev. Mod. Phys., 76, 125, doi: 10.1103/RevModPhys.76.125

Madau, P., Ferguson, H. C., Dickinson, M. E., et al. 1996, MNRAS, 283, 1388, doi: 10.1093/mnras/283.4.1388

Manara, C. F., Robberto, M., Da Rio, N., et al. 2012, ApJ, 755, 154, doi: 10.1088/0004-637X/755/2/154

McKee, C. F., & Ostriker, E. C. 2007, ARA&A, 45, 565, doi: 10.1146/annurev.astro.45.051806.110602

Meynet, G., & Maeder, A. 2000, A&A, 361, 101. https://arxiv.org/abs/astro-ph/0006404

Nayak, O., Meixner, M., Fukui, Y., et al. 2018, ApJ, 854, 154, doi: 10.3847/1538-4357/aaab5f

Nayak, O., Meixner, M., Indebetouw, R., et al. 2016, ApJ, 831, 32, doi: 10.3847/0004-637X/831/1/32

Nota, A., Sirianni, M., Sabbi, E., et al. 2006, ApJL, 640, L29, doi: 10.1086/503301

Oey, M. S., & Massey, P. 1995, ApJ, 452, 210, doi: 10.1086/176292

Offner, S. S. R., Clark, P. C., Hennebelle, P., et al. 2014, in Protostars and Planets VI, ed. H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning (University of Arizona Press), 53, doi: 10.2458/azu_uapress_9780816531240-ch003

Offner, S. S. R., Hansen, C. E., & Krumholz, M. R. 2009, ApJL, 704, L124, doi: 10.1088/0004-637X/704/2/L124

Padoan, P., Pan, L., Juvela, M., Haugbølle, T., & Nordlund, Å. 2020, ApJ, 900, 82, doi: 10.3847/1538-4357/abaa47

Panagia, N., Gilmozzi, R., Macchetto, F., Adorf, H. M., & Kirshner, R. P. 1991, ApJL, 380, L23, doi: 10.1086/186164

Parker, R. J., & Wright, N. J. 2016, MNRAS, 457, 3430, doi: 10.1093/mnras/stw087

Paxton, B., Bildsten, L., Dotter, A., et al. 2011, ApJS, 192, 3, doi: 10.1088/0067-0049/192/1/3

Paxton, B., Cantiello, M., Arras, P., et al. 2013, ApJS, 208, 4, doi: 10.1088/0067-0049/208/1/4

Paxton, B., Marchant, P., Schwab, J., et al. 2015, ApJS, 220, 15, doi: 10.1088/0067-0049/220/1/15

Robberto, M., Soderblom, D. R., Bergeron, E., et al. 2013, ApJS, 207, 10, doi: 10.1088/0067-0049/207/1/10

Sabbi, E., Sirianni, M., Nota, A., et al. 2008, AJ, 135, 173, doi: 10.1088/0004-6256/135/1/173

Sabbi, E., Anderson, J., Lennon, D. J., et al. 2013, AJ, 146, 53, doi: 10.1088/0004-6256/146/3/53

Sabbi, E., Lennon, D. J., Anderson, J., et al. 2016, ApJS, 222, 11, doi: 10.3847/0067-0049/222/1/11

Shu, F. H., Adams, F. C., & Lizano, S. 1987a, ARA&A, 25, 23, doi: 10.1146/annurev.aa.25.090187.000323

—. 1987b, ARA&A, 25, 23, doi: 10.1146/annurev.aa.25.090187.000323

Stanek, K. Z., Zaritsky, D., & Harris, J. 1998, ApJL, 500, L141, doi: 10.1086/311420

Sun, N.-C., de Grijs, R., Subramanian, S., et al. 2017, ApJ, 835, 171, doi: 10.3847/1538-4357/835/2/171

Tan, J. C., Krumholz, M. R., & McKee, C. F. 2006, ApJL, 641, L121, doi: 10.1086/504150

Udalski, A., Soszynski, I., Szymanski, M., et al. 1999a, AcA, 49, 223. https://arxiv.org/abs/astro-ph/9908317

—. 1999b, AcA, 49, 437. https://arxiv.org/abs/astro-ph/9912096

Vallenari, A., Chiosi, E., & Sordo, R. 2010, A&A, 511, A79, doi: 10.1051/0004-6361/200913053

van der Marel, R. P., & Cioni, M.-R. L. 2001, AJ, 122, 1807, doi: 10.1086/323099

Walch, S., Whitworth, A. P., Bisbas, T. G., Wünsch, R., & Hubber, D. A. 2013, MNRAS, 435, 917, doi: 10.1093/mnras/stt1115

Wong, T., Hughes, A., Ott, J., et al. 2011, ApJS, 197, 16, doi: 10.1088/0067-0049/197/2/16

Zaritsky, D., Harris, J., Thompson, I. B., & Grebel, E. K. 2004, AJ, 128, 1606, doi: 10.1086/423910

Zinnecker, H., & Yorke, H. W. 2007, ARA&A, 45, 481, doi: 10.1146/annurev.astro.44.051905.092549

APPENDIX

A. UMS SELECTION AND FIELD CONTAMINATION

In Section 4.2 we select the brightest UMS sources in the MYSST FOV to estimate their individual stellar extinction and derive a reddening map for N44. While the massive stars in N44's star forming centers are young, the LMC field stars that contaminate the FOV of the survey are generally older evolved populations. Consequently, older stars of high enough mass in their post-MS evolutionary phase can pollute the high brightness regime of the optical CMD. Therefore, we have to ensure that our UMS selection avoids as much field contamination as possible.

To do so, we first limit the candidate UMS stars to objects bluer than 0.5 mag in $m_{F555W} - m_{F814W}$ to avoid the RGB and RC. As the lower brightness limit in m_{F555W} we define a line parallel to the reddening vector. As mentioned in Section 2 the MYSST survey has also observed two LMC fields close to N44 for reference. By comparing the field CMDs to the main one, we can determine the severity of the field contamination in relation to the chosen line and determine a suitable m_{F555W}-axis intercept.

To quantify the contamination of a given UMS selection, we first subsample the main CMD, as its FOV is much larger than the two fields. The latter have approximate surface areas of $1650\,pc^2$ (northern) and $2530\,pc^2$ (southern), while the main field covers about $33,440\,pc^2$. Using the area ratios, we randomly subsample the main CMD once to match each field individually and once for the combination of both fields. For a given UMS selection criterion we then count the selected stars in the subsampled main CMD and the respective field to determine the relative field contamination. The top row in Figure A1 shows examples for the subsampled main CMDs in comparison to the corresponding reference fields (bottom row). To account for randomness, we repeat the sub-sampling procedure 5000 times and average the results.

We determine an m_{F555W}-axis intercept of 18.5 mag for our UMS selection criterion as the best compromise between selecting enough UMS sources to reasonably cover the main FOV and avoiding field contamination. Subsampling to the area and comparing to the CMD of the fields, this selection criterion entails a $21.9\% \pm 2.7\%$ contamination for the northern field, $10.2\% \pm 1.0\%$ for the southern field and $14.8\% \pm 1.1\%$ for both fields combined. Important to note here is that the 14 stars se-lected from the northern field could actually be UMS stars and not just old field contaminants. In the CMD of the northern field (bottom left panel of Figure A1), we actually find a notable population of stars in the PMS region, contrary to the southern field CMD (bottom middle panel), where this area is practically empty. Therefore, it is possible that the northern reference field might have captured a small star forming cluster and its UMS stars.

Out of the 1291 total sources selected by our criterion, only 24 come from the reference fields. Given their CMD positions and the case we have made for the northern field, we cannot easily dismiss these as non-UMS stars. Therefore, we decide to keep them as UMS candidates in our analysis and also derive (low-resolution) extinction maps for the reference fields.

B. UMS EXTINCTION ESTIMATE ERROR

In our extinction estimation approach, presented in Section 4.2, we assume that our selected UMS sources should theoretically lie on the ZAMS. As already mentioned, this ZAMS assumption does not necessarily hold for the rapidly evolving massive stars, and their true position might actually differ from the ZAMS. Consequently, our approach may overestimate the extinction of some UMS stars, at worst providing only an upper extinction limit. We make the ZAMS assumption because we cannot easily date all UMS stars in our selection from the MYSST data alone and want our extinction measure to make the same systematic error everywhere. For some UMS stars in our sample, however, ages have been estimated in previous studies. For example Oey & Massey (1995) find that the massive stars inside N44's bubble are about 10 Myr old, while the ones at the western bubble rim are younger at around 5 Myr. In this appendix we will briefly estimate the error in extinction that our approach entails for these two populations.

To get a first idea of the systematic error of our extinction estimation approach, we estimate how many of our 1291 UMS extinction probes would actually be affected if they were 5 or 10 Myr old instead of falling on the ZAMS. We do so by approximating the point on the ZAMS in the CMD where the 5 or 10 Myr PARSEC isochrone starts to significantly move away from the ZAMS track. We then project this point along the reddening vector to derive a threshold line in the CMD above which the UMS extinction measurement would change if the 5 or 10 Myr isochrone was used instead of the ZAMS track. For the 10 Myr isochrone we identify
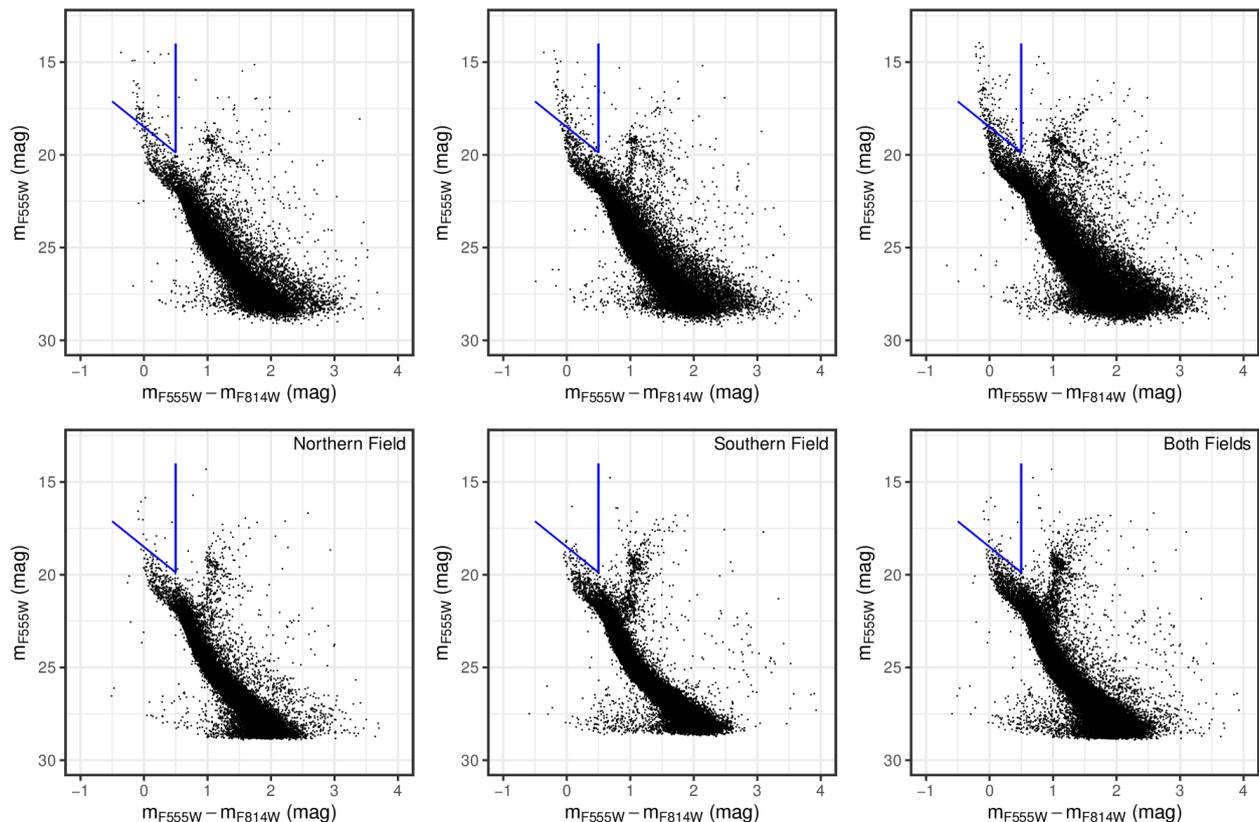
Figure A1. Optical CMDs of the main FOV of the MYSST survey subsampled to match the surface area of the northern (top left), southern (top middle) and both reference fields (top right). For comparison, the bottom row shows the actual CMDs of the northern (left), southern (middle) and both reference fields (right). The blue lines indicate the limits of our UMS selection in all panels, demonstrating that we avoid most of the field contamination.

this point on the ZAMS at about 16 mag in F555W (including the LMC distance modulus and MW foreground extinction) and at 14 mag for the 5 Myr one. Note that the 5 Myr isochrone is fairly irregularly shaped, moving on and off the ZAMS, but appears to finally detach around 14 mag. The left panel of Figure B1 shows these threshold lines in the MYSST CMD along with the corresponding isochrones and the UMS stars for which the extinction measurement would be affected by using the 5 or 10 Myr isochrone instead of the ZAMS. Here we find 41 UMS stars that would be affected if they were 5 Myr old, and 276 in the 10 Myr case in total. Consequently, even if all UMS stars in our selection were 10 Myr old, only 276 out of our 1291 probes would even show a change in the measured extinction. As their spatial distribution in the right panel of Figure B1 in comparison to the UMS extinction map shows, most of the affected UMS stars are located inside N44's bubble and its (western) rim. Particularly interesting in this diagram is that almost none of the affected UMS stars

fall into the high-extinction "filament" extending from the southern bubble rim to the southwest corner of the FOV. As mentioned in the main text, this "filament" has no visible nebulous counterpart (i.e. gas/dust) in long-wavelength observations of N44 (e.g. Spitzer), indicating that our approach is overestimating extinction in this region. If this is indeed the case, then Figure B1 suggests that the UMS stars inside this "filament" must be even older than 10 Myr.

Lastly, to quantify the error of our approach, we measure extinction for the UMS stars in the bubble and rim with the correct 10 and 5 Myr isochrones instead of the ZAMS. In both cases we include all UMS stars that fall into the bubble (solid) and western rim (dashed) outlines in Figure B1 and compute the median absolute error with respect to our ZAMS measurement. Here we find $0.036^{+0.026}_{-0.015}$ mag for the 241 UMS stars in the western rim and $0.045^{+0.020}_{-0.007}$ mag for the 161 stars inside the bubble. Averaged across both populations, the median absolute extinction error is $0.043^{+0.019}_{-0.014}$ mag with a max-
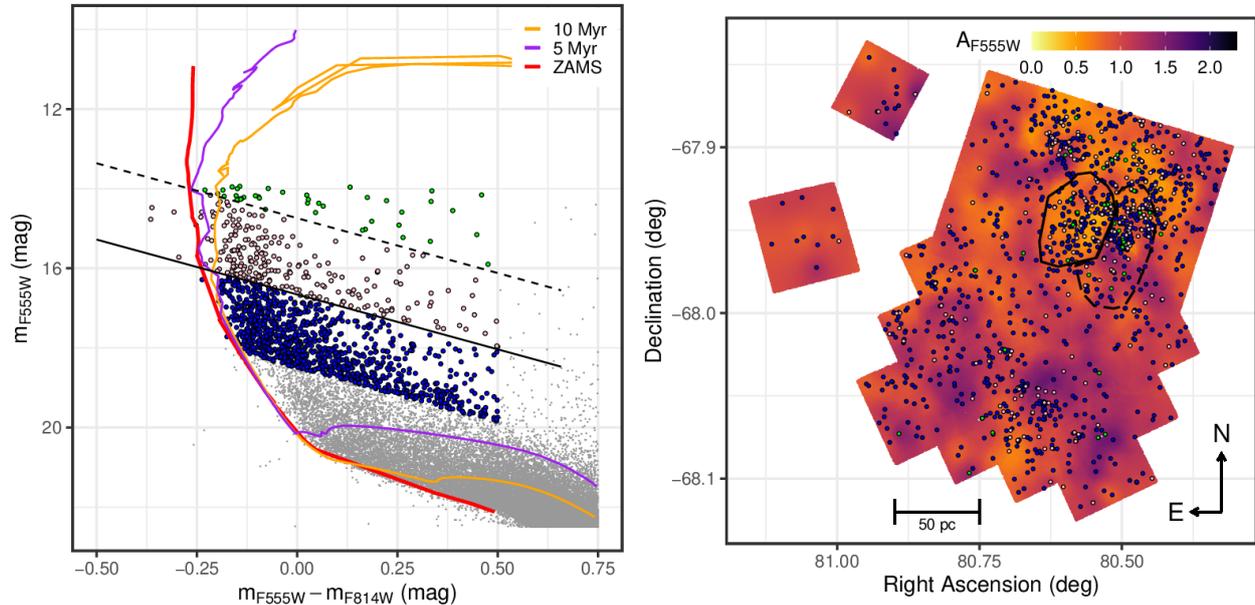
Figure B1. Zoom-in on the bright part of the MYSST CMD (left). The solid lines signify the ZAMS (red), a 5 Myr (purple) and a 10 Myr (orange) PARSEC isochrone, all corrected for the LMC distance modulus and MW foreground extinction. The large colored points mark our UMS selection as in the left panel of Figure 8. The black solid and dashed lines show a projection of the points, where the 10 and 5 Myr isochrones start to move away from the ZAMS, along the reddening vector. The green points mark the UMS stars where the extinction measure would change if a 5 Myr isochrone is used instead of the ZAMS. If a 10 Myr isochrone is used instead of the ZAMS then the extinction measurement of the pink (+green) points is affected. The blue points are the UMS stars that are unaffected even in the 10 Myr case. Right: UMS extinction map of the MYSST survey (as in Figure 10). The large color points mark the position of the three groups of UMS stars identified in the left panel. Additionally, the solid and dashed black lines mark the position of the interior of N44's bubble and its western edge, respectively.

imum of 0.27 mag. The extinction error of our approach is, thus, overall fairly small inside and at the rim of the bubble. This obviously does not extend to UMS sources in our sample that are notably older than 10 Myr, but it at least confirms that our approach estimates the extinction around N44's superbubble fairly accurately and is at worst only an upper limit everywhere else.

C. UMS EXTINCTION AND UMS BROADENING

In Appendix B we have discussed the error entailed in our extinction estimate due to assuming the ZAMS as the true position for the UMS stars. Besides differences in stellar age, there are, however, other physical effects that result in a broadening of the UMS even in the absence of differential extinction. These include e.g. unresolved binarity or metallicity gradients in the observed population. To ascertain the impact of these phenomena on our extinction estimation procedure, we create two synthetic populations that are not affected by *differential* reddening, such that all broadening of the UMS is caused by other effects. For both of the populations we assume an unresolved binarity fraction of 0.4 with a flat mass ratio distribution. We account

for the LMC distance modulus ($(m − M)_0 = 18.55$) and include the constant shift due to MW foreground reddening ($A_V^{mw} = 0.22$ mag). Lastly, we consider a metallicity spread of [Fe/H] = −0.3 to −0.2, corresponding to a range of Z = 0.0076–0.0096 assuming the PARSEC solar metallicity of $Z_\odot = 0.01524$.

Our first synthetic data set represents an approximately single-age population, formed with a constant SF rate between 5 and 5.6 Myr. The second one is a more extreme case, emulating a mixed-age population resulting from a constant SF rate between 3.2 Myr and 12.6 Gyr. The top panels in Figure C1 show these synthetic populations in the bright part of the CMD in comparison to the MYSST data.

We then repeat our UMS selection on these synthetic, intrinsically not reddened populations and estimate "extinction" for the selected stars to quantify the impact of other broadening effects on our estimated extinction distribution. The results are displayed in the bottom panels of Figure C1 in comparison to the outcome on the MYSST data.

In the roughly single-age synthetic population case we find a median "extinction" of $0.269^{+0.032}_{-0.012}$ mag or
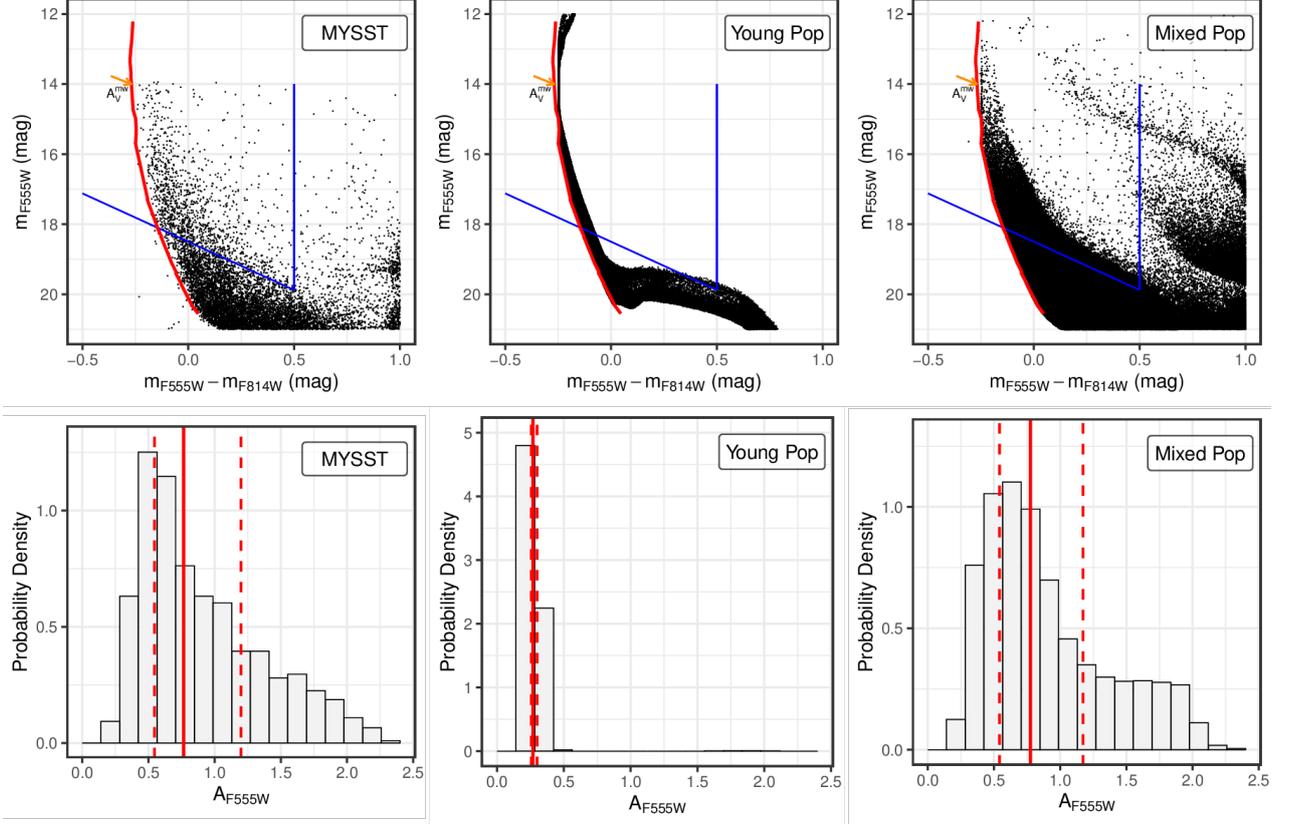
Figure C1. Zoom-in on the bright part of the CMD for the data from the MYSST survey (top left), and two synthetic data sets. One represents a single-age population with a star formation episode between 5 and 5.6 Myr (top middle), while the other emulates a mixed population with a constant star formation rate between 3.2 Myr and 12.6 Gyr (top right). Both synthetic populations are shifted according to the MW foreground reddening but are intrinsically not affected by any extinction. In all three panels the blue lines indicate the limits of our UMS selection for the extinction estimation, while the red line marks the ZAMS, corrected for the LMC distance and MW foreground reddening. Lastly, the orange arrow illustrates the shift of synthetic data and ZAMS due to the MW foreground reddening of $A_V^{mw} = 0.22$ mag (or $A_{F555W}^{mw} = 0.223$ mag). The bottom row shows the extinction distributions (including MW foreground) estimated by our method for the stars falling in the UMS selection of the corresponding data set in the top row. In the bottom panels the solid red line indicates the median estimated extinction, while the dashed red lines mark the 25% and 75% quantiles.

$0.046^{+0.032}_{-0.012}$ mag when subtracting the MW foreground. The error in the estimate introduced by nonextinction broadening effects is, therefore, fairly negligible when dealing with the ideal case of a single-age population. In the case of the mixed-age population, created by SF with a constant rate over several tens of gigayears, the outcome differs significantly. Here we find a median "extinction" of $0.77^{+0.40}_{-0.23}$ mag (or $0.56^{+0.40}_{-0.23}$ mag subtracting A_{F555W}^{mw}) from the broadening of the UMS alone without any real differential extinction. This value is almost identical to our result on the real MYSST data, and as we can see, comparing the bottom left and right panels of Figure C1, the derived extinction distributions are similarly shaped too. While this result at first glance might call our MYSST extinction estimates into ques-

tion, it is important to emphasize here that this synthetic mixed-age population is not set up to match the LMC/N44 but as an extreme case to represent a worst-case scenario. As previous studies (e.g. Oey & Massey 1995) indicate, N44 is a region of multiple recent events of accelerated SF rather than the outcome of a constant star formation process over several gigayears. Additionally, our comparison with the observed LMC reference fields in Appendix A shows clearly that the background contamination in our UMS selection is fairly minimal. Consequently, the synthetic mixed-age population is not likely to be a realistic representation of N44. Therefore, we conclude that the similarity of the MYSST extinction distribution and that of the mixed age synthetic population is a coincidence. Nevertheless, this experi-

ment indicates that our extinction estimation procedure may be susceptible to larger systematic errors if applied to certain populations.

D.

Table D1. MYSST photometric catalog

ID	X	Y	R.A.	Decl.	...	Type	F555W								F814W								$f_{\rm po}$
							m	σ	χ^2	S/R	shrp	rnd	crwd	f	m	σ	χ^2	S/R	shrp	rnd	crwd	f	
	(pixel)	(pixel)	(deg)	(deg)			(mag)	(mag)					(mag)		(mag)	(mag)					(mag)		
MYSST 052124.32-675620.82	27206.8	16934.1	80.35	-67.94		1	15.962	0.003	1.54	426	-0.047	0.029	0.04	0	15.148	0.003	1.47	379	-0.032	0.001	0.03	0	2
MYSST 052140.51-675405.03	24943.0	20342.0	80.42	-67.90		1	16.804	0.005	1.38	220	0.009	0.032	0.10	0	14.762	0.002	0.97	461	-0.007	0.008	0.07	0	2
MYSST 052120.17-675548.08	27795.9	17749.0	80.33	-67.93		1	16.628	0.003	0.84	342	-0.016	0.006	0.03	0	14.964	0.003	0.93	431	-0.007	-0.006	0.03	0	2
MYSST 052207.85-675424.18	21085.5	19879.3	80.53	-67.91		1	16.889	0.004	1.02	291	-0.017	0.014	0.06	0	14.749	0.003	1.09	430	0.010	0.012	0.04	0	2
MYSST 052300.45-675319.41	13664.9	21508.7	80.75	-67.89		1	16.300	0.003	1.15	334	0.001	0.042	0.05	0	15.459	0.003	0.97	333	-0.029	0.013	0.05	0	2
MYSST 052152.65-675612.58	23217.2	17161.2	80.47	-67.94		1	15.735	0.002	0.91	534	-0.019	0.001	0.02	0	15.852	0.004	1.06	269	-0.033	0.012	0.02	0	2
MYSST 052238.97-675509.11	16694.8	18764.8	80.66	-67.92		1	16.468	0.005	1.51	236	-0.022	0.074	0.08	0	15.791	0.004	0.90	281	-0.002	0.024	0.04	0	3
MYSST 052157.45-675612.75	22541.9	171159.5	80.49	-67.94		1	15.927	0.002	1.03	468	-0.030	0.004	0.04	0	15.872	0.005	1.37	206	-0.030	0.021	0.05	0	2
MYSST 052127.44-675437.94	26782.4	19508.9	80.36	-67.91		1	16.982	0.004	0.94	284	0.009	0.015	0.02	0	15.231	0.003	0.96	371	0.009	0.015	0.03	0	2
MYSST 052151.66-675312.27	23376.1	21668.2	80.47	-67.89	...	1	16.090	0.007	2.43	167	0.012	0.034	0.22	0	15.834	0.007	1.78	149	0.030	-0.026	0.18	0	3
MYSST 052120.99-675526.03	27684.7	18301.3	80.34	-67.92		1	16.982	0.004	0.74	287	-0.015	0.035	0.02	0	15.462	0.003	0.79	339	-0.002	-0.021	0.02	0	2
MYSST 052151.10-675303.12	23456.9	21896.3	80.46	-67.88		1	15.832	0.003	1.22	398	-0.003	0.042	0.07	0	15.903	0.004	1.10	242	0.010	0.017	0.07	0	2
MYSST 052147.08-675608.33	24002.2	17263.8	80.45	-67.94		1	16.049	0.003	1.41	327	-0.021	0.039	0.04	0	16.113	0.004	0.77	254	-0.009	0.002	0.04	0	2
MYSST 052214.31-675506.99	20170.8	18811.1	80.56	-67.92		1	17.208	0.003	1.25	406	-0.021	0.020	0.04	2	15.067	0.002	1.16	661	-0.008	0.014	0.02	2	2
MYSST 052156.07-675530.81	22740.2	18207.3	80.48	-67.93		1	15.946	0.002	1.20	521	0.027	-0.010	0.05	2	16.098	0.004	1.33	246	0.013	0.010	0.05	2	2
MYSST 052136.23-675334.68	25551.0	21097.1	80.40	-67.89		1	16.224	0.005	1.72	221	-0.034	0.021	0.05	0	16.112	0.005	1.00	226	-0.021	0.007	0.04	0	2
MYSST 052230.62-675327.57	17875.2	21302.2	80.63	-67.89		1	15.987	0.003	1.17	409	-0.023	0.003	0.01	0	16.144	0.004	0.91	249	-0.018	0.001	0.01	0	3
MYSST 052218.45-675327.24	19593.2	21306.9	80.58	-67.89		1	16.065	0.003	1.13	403	-0.008	0.012	0.03	0	16.119	0.004	0.91	251	0.009	0.015	0.05	0	2
MYSST 052147.82-675521.85	23903.3	18425.9	80.45	-67.92		1	17.120	0.003	1.03	360	-0.028	0.023	0.02	0	15.727	0.003	1.17	357	-0.013	0.013	0.02	0	2
MYSST 052229.28-675354.70	18063.5	20622.9	80.62	-67.90		1	16.152	0.004	1.37	309	-0.020	0.027	0.02	0	16.233	0.005	0.90	234	-0.006	-0.014	0.01	0	2

NOTE—The omitted ($\dots$) columns list combined values (over the two filters) for the PSF fit quality χ^2, signal to noise ratio *SNR*, sharpness *shrp*, roundness *rnd*, crowding *crwd* and direction of major axis (if not round) *mjazdir*.

3.4 Measuring Young Stars in Space and Time - II. The Pre-Main-Sequence Stellar Content of N44

In Ksoll et al. (2021b) the ML classification approach established in Ksoll et al. (2018) is applied to the photometric catalogue of the MYSST survey to identify the young PMS population of N44 and study their spatial distribution. In particular a contour density approach is employed to quantify significant PMS cluster structures, the cluster properties are derived, and the spatial distribution of the identified PMS population is compared to the locations of known H II regions, young massive O stars and YSOs.

Contributions: Dimitrios Gouliermis proposed the application of our ML classification approach to the MYSST survey photometric catalogue to find the PMS population of N44. Dimitrios and myself came up with the idea to use a contour density based approach to study the clustering structure of the identified PMS stars. I conducted the complete analysis for this paper, i.e. constructing a new training set, training and testing the ML models, applying the trained method to the MYSST catalogue, implementing the contour density based clustering approach and quantifying the cluster structure of the PMS stars. I created all figures and wrote all text for the initial manuscript draft. The co-authors provided feedback on the manuscript, additional evaluation ideas and assisted with proofreading.

Measuring Young Stars in Space and Time - II. The Pre-Main-Sequence Stellar Content of N44

Victor F. Ksoll [ID],[1,2] Dimitrios Gouliermis [ID],[1,3] Elena Sabbi [ID],[4] Jenna E. Ryon,[4] Massimo Robberto [ID],[4,5]
Mario Gennaro [ID],[4,5] Ralf S. Klessen [ID],[1,2] Ullrich Koethe [ID],[6] Guido de Marchi [ID],[7] C.-H. Rosie Chen [ID],[8]
Michele Cignoni,[9,10,11] and Andrew E. Dolphin[12,13]

[1]*Universität Heidelberg, Zentrum für Astronomie, Institut für Theoretische Astrophysik,
Albert-Ueberle-Str. 2, 69120 Heidelberg, Germany*
[2]*Universität Heidelberg, Interdisziplinäres Zentrum für Wissenschaftliches Rechnen,
Im Neuenheimer Feld 205, 69120 Heidelberg, Germany*
[3]*Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany*
[4]*Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA*
[5]*Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA*
[6]*Universität Heidelberg, Heidelberg Collaboratory for Image Processing, Visual Learning Lab,
Berliner Str. 43, 69120 Heidelberg, Germany*
[7]*European Space Research and Technology Centre, Keplerlaan 1, 2200 AG Noordwijk, Netherlands*
[8]*Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-53121 Bonn, Germany*
[9]*Department of Physics - University of Pisa, Largo B. Pontecorvo, 3 Pisa, 56127, Italy*
[10]*INFN, Largo B. Pontecorvo 3, 56127, Pisa, Italy*
[11]*INAF-Osservatorio di Astrofisica e Scienza dello Spazio, Via Gobetti 93/3, 40129, Bologna, Italy*
[12]*Raytheon, 1151 E. Hermans Road, Tucson, AZ 85706, USA*
[13]*Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721, USA*

ABSTRACT

The Hubble Space Telescope (HST) survey Measuring Young Stars in Space and Time (MYSST) entails some of the deepest photometric observations of extragalactic star formation, capturing even the lowest-mass stars of the active star-forming complex N44 in the Large Magellanic Cloud. We employ the new MYSST stellar catalog to identify and characterize the content of young pre-main-sequence (PMS) stars across N44 and analyze the PMS clustering structure. To distinguish PMS stars from more evolved line of sight contaminants, a non-trivial task due to several effects that alter photometry, we utilize a machine-learning classification approach. This consists of training a support vector machine (SVM) and a random forest (RF) on a carefully selected subset of the MYSST data and categorize all observed stars as PMS or non-PMS. Combining SVM and RF predictions to retrieve the most robust set of PMS sources, we find $\sim 26,700$ candidates with a PMS probability above 95% across N44. Employing a clustering approach based on a nearest neighbor surface density estimate, we identify 18 prominent PMS structures at 1σ significance above the mean density with sub-clusters persisting up to and beyond 3σ significance. The most active star-forming center, located at the western edge of N44's bubble, is a subcluster with an effective radius of $\sim 5.6\,\mathrm{pc}$ entailing more than 1100 PMS candidates. Furthermore, we confirm that almost all identified clusters coincide with known H II regions and are close to or harbor massive young O stars or YSOs previously discovered by MUSE and Spitzer observations.

1. INTRODUCTION

Star formation is one of the most fundamental processes in our universe, bringing light to the galaxies and ultimately providing the environments for the nucle-

Corresponding author: Victor F. Ksoll
v.ksoll@stud.uni-heidelberg.de

osynthesis of all heavier elements. The primary birth-places of stars in galaxies are giant molecular clouds, enormous reservoirs of atomic and molecular hydrogen, harboring the necessary material to create stars (for a review, see e.g. Klessen & Glover 2016, and references therein). Within these clouds stars tend to form in clusters and, in some instances, create large star-forming complexes with multiple stellar populations of different ages, where the feedback of the massive, but short-lived, constituents can repeatedly trigger new star-forming events (Lee & Chen 2007; Elmegreen 2011). These young and bright objects are the signposts of massive star-forming clusters (Zinnecker & Yorke 2007; Portegies Zwart et al. 2010), but as studies of the stellar initial mass function (IMF) indicate (see Kroupa 2002; Chabrier 2003), intermediate- and low-mass objects actually contribute a significant fraction to a cluster's total stellar mass. Contrary to their massive blue siblings, these low-mass *pre*-main-squence (PMS) stars, still in the Kelvin-Helmholtz contraction phase (Stahler & Palla 2005), require increasingly longer time to reach the main sequence (MS) as their masses get smaller, down to the hydrogen-burning limit (about 0.072 M$_\odot$, Schulz 2012). In the first few megayears PMS stars may still be forming, accreting gas from their immediate surroundings and circumstellar disks (Hartmann et al. 2016). Low-mass PMS objects trace the history of (recent) star formation beyond the few megayears probed by the ephemeral most massive stars. Therefore, our understanding of star formation may greatly benefit from the study and observation of young PMS objects and the stellar clusters within which they are born.

Large photometric surveys of nearby systems are one of the main astronomical methods to perform in-depth studies of remote stellar clusters and identify star-forming regions. For more than three decades one of the most successful tools for such photometric surveys has been the Hubble Space Telescope (HST), providing observations with exceptional spatial resolution and to great depth. In the past the HST has proven especially capable of detecting faint PMS sources in the Magellanic Clouds, the dwarf companion galaxies to our Milky Way (Gouliermis et al. 2006, 2012; Nota et al. 2006; Sabbi et al. 2007; Da Rio et al. 2010, 2012; Sabbi et al. 2016). Aside from harboring the only extragalactic PMS sources we can spatially resolve, the Magellanic Clouds are characterized by a relatively high star-forming activity, observable at lower extinction, since they are not obscured by the dusty Galactic disk. Therefore, the Magellanic Clouds provide very attractive targets for the study and observations of large ensembles of PMS stars (Gouliermis 2012).

One such complex is the active star-forming region N44 (LHα 120–N44; Henize 1956), located in the Large Magellanic Cloud (LMC). It consists of a giant complex of H II regions, one of the most luminous across the entire LMC after 30 Doradus and N11 (Kennicutt & Hodge 1986; Pellegrini et al. 2012), entailing an enormous central superbubble and several compact H II regions along its ridge (Pellegrini et al. 2012; McLeod et al. 2019). The youthfulness of the stars within these ionized gas reservoirs is highlighted by three OB associations (LH47, LH48 and LH49; Lucke & Hodge 1970) and a plethora of more than 30 massive, short-lived O-type stars that have been identified in N44 by spectroscopic studies (McLeod et al. 2019; Will et al. 1997; Oey & Massey 1995; Conti et al. 1986; Rousseau et al. 1978). N44 also exhibits evidence for multiple star-forming events and feedback-triggered star formation, as previous studies have found a $\sim 5\,\mathrm{Myr}$ difference in age between the stellar populations within and at the rim of N44's bubble (Oey & Massey 1995), as well as the presence of a supernova remnant, SNR 0523-679 (Chu et al. 1993), in the vicinity of the bubble (Jaskot et al. 2011). In addition, there is active, ongoing star formation in N44, as Chen et al. (2009) find 59 massive young stellar objects (YSOs) within N44 from observations with the Spitzer Space Telescope. Combining Spitzer data from the SAGE (Surveying the Agents of a Galaxy's Evolution; Meixner et al. 2006) legacy program with optical photometry from the Magellanic Clouds Photometric Survey (MCPS; Zaritsky et al. 1997) and near-infrared photometry from the InfraRed Survey Facility (IRSF; Kato et al. 2007) this list is extended by another 139 YSOs (18 in common with Chen et al. 2009, matched to within 1$''$) by Carlson et al. (2012). In a recent study, Zivkov et al. (2018) have used near infrared observations from the VISTA Survey of the Magellanic Clouds (VMC; Cioni et al. 2011) to estimate the number of PMS sources in N44. Identifying regions containing PMS sources from density excesses in $K_s/(Y - K_s)$ Hess diagrams in comparison to the underlying fields, they find a lower limit to the number of PMS stars in N44 of 1000 ± 38.

N44's complexity is captured by the deep HST imaging of the "Measuring Young Stars in Space and Time" (MYSST) survey, which obtained photometry in two broadband filters for more than 400,000 sources across the extent of N44 (Ksoll et al. 2020a, Paper I). The rich color magnitude diagram (CMD) of the MYSST survey not only has revealed the presence of significant differential reddening within N44 but also entails many populations of different ages in the observed area. Consequently, a significant overlap between the old lower MS (LMS) or red giant branch (RGB), and the PMS popula-
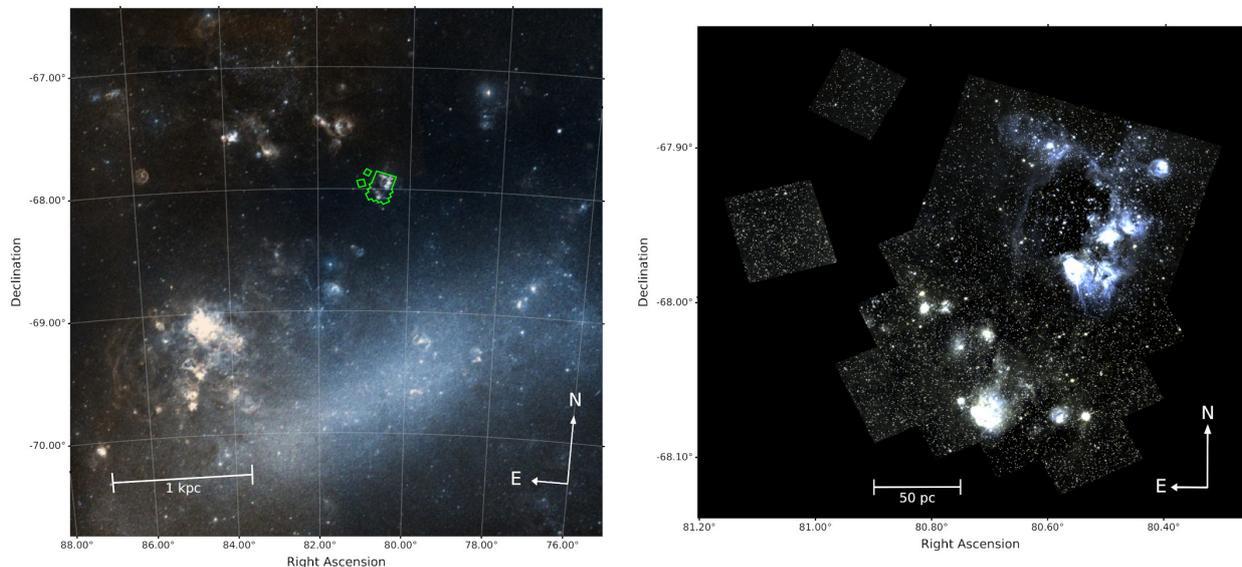
Figure 1. Left: color composite image from the Digitized Sky Survey (Lasker et al. 1996) of the wider LMC neighborhood of N44. The green contour highlights the main FOV and the observed reference fields of the MYSST survey. Right: two-color composite image of N44 from the MYSST survey, presenting the observations in the F555W and F814W filters in blue and green, respectively. Image from Ksoll et al. (2020a) used with permission.

tion occurs in the CMD, making it particularly difficult to distinguish the young N44 cluster constituents from the field contaminants in this large data set without additional information about the excess in emission lines that accompany the PMS phase (e.g. De Marchi et al. 2010).

To disentangle the PMS population from the older stars in a statistically sound manner using only broadband photometry requires sophisticated algorithms, such as the machine-learning (ML) approaches we have demonstrated in a previous study (Ksoll et al. 2018). In the recent years, there have been many examples of established ML approaches successfully applied to astronomical problems involving regression, classification, and clustering tasks (see e.g. Baron 2019; Fluke & Jacobs 2020, for reviews of recent applications).

In this paper we present the identification of the youngest PMS candidates in N44 using the photometric catalog from the HST survey MYSST (Paper I). Our approach, established in Ksoll et al. (2018), consists of an ML-based classification of the PMS and non-PMS constituents of the survey. This study is structured as follows. In Section 2 we provide a brief summary of the MYSST photometric catalog. In Section 3 we begin by describing the construction of the necessary training set for our ML classification approach from a subset of the observational data. This entails the careful selection

of a region within N44 that contains distinct PMS and LMS populations, as well as the addition of examples of field RGB contaminants from suitable areas. Then, we present the training and test performance of our models. In Section 4 we discuss the classification results of our approach, while in Section 5 we analyze the spatial clustering structure of the identified PMS candidate stars. Finally, Section 6 provides a summary and considerations on future developments.

2. DATA

The MYSST program observed the star-forming complex N44, located in the Large Magellanic Cloud, with a deep, high spatial resolution HST survey (Paper I). Its field of view (FOV) of $12.2 \times 14.7\,\mathrm{arcmin}^2$, corresponding to about $180\,\mathrm{pc} \times 215\,\mathrm{pc}$ at the LMC distance $((m-M)_0 = 18.55\pm0.05$; Panagia et al. 1991; De Marchi et al. 2016), entails N44's characteristic superbubble and the region south of it. Figure 1 shows the MYSST FOV in the greater LMC neighborhood of N44 (left) and the MYSST two-color composite image (right). The survey was conducted in two broadband filters, F555W and F814W, with the Advanced Camera for Surveys (ACS) and Wide Field Camera 3 (WFC3) instruments of the HST. Reaching down to about $29\,\mathrm{mag}$ in F555W and $28\,\mathrm{mag}$ in F814W, the MYSST survey is one of the deepest photometric studies of extragalactic stars, probing

even the lowest-mass populations of N44. The F555W detection limit implies the capture of e.g. unreddened 1 Myr PMS stars with masses as low as $0.09\,M_\odot$ (see completeness discussion in Paper I) at the distance of the LMC. In this paper we use the MYSST photometric catalog presented in Paper I, consisting of 461,684 sources across the observed FOV of N44 and two smaller LMC reference fields. This catalog only entails objects up to $14\,\mathrm{mag}$ in F555W and $13\,\mathrm{mag}$ in F814W, as brighter sources were lost owing to saturation. Consequently, the available data are likely missing some of the most massive stars, i.e. early O stars, of the region.

N44 is also subject to a substantial amount of differential reddening. In Paper I we establish reddening properties for the MYSST survey by fitting the slope of the extinction-elongated red clump using the RANSAC algorithm. Furthermore, we derive individual stellar extinctions using upper MS (UMS) stars as extinction probes and assigning a distance-weighted average extinction of the nearest UMS stars to all other sources. This extinction estimate entails some caveats. First, we assume the UMS stars to be on the zero-age-MS (ZAMS) to measure their extinction. For the quickly evolving massive O stars this might not necessarily be the case anymore, even if they are still young. In fact, Oey & Massey (1995) estimate the O-star population in N44's bubble to be about 10 Myr old while O stars in the bubble rim are 5 Myr younger. However, we find that the error for using the ZAMS instead of e.g. a 10 Myr isochrone is only on the order of 0.04 mag for our selection of UMS sources. In any case, this ZAMS assumption for the UMS sources means that the estimated reddening is at worst only an upper limit of the true extinction for older UMS stars. Second, while it has been found that using the reddening of UMS neighbors returns reasonable values for constituents of young star forming regions (De Marchi et al. 2016), such as N44, there is no guarantee that the UMS extinction is representative for field sources, leading to occasional over- or underestimates.

3. TRAINING SET

Ksoll et al. (2018) establish an ML approach for the identification of PMS candidate stars based on HST photometry, which here we apply to the MYSST data. The method entails the careful selection of a training set from the observational data, in which a distinction between examples of PMS and non-PMS stars can be made easily. With this labeled training data the classical ML techniques called support vector machine (SVM) and random forest (RF) are then trained to distinguish these

two classes of stars based on their broadband photometry and estimated extinction.

Due to the different filter passbands between the Hubble Tarantula Treasury Project (HTTP) data of Ksoll et al. (2018) and the MYSST survey, one cannot reuse the HTTP training set. The intrinsic differences between the two star-forming regions would in any case justify the creation of a new training set specific to the MYSST data of N44.

3.1. PMS Training Set

As a base for our training set we select a subset of the MYSST data that is likely to contain a suitable number of PMS stars, as well as LMS contaminants. The latter are likely, for the most part, field constituents, but they could also consist of low-mass remnants of earlier star formation episodes in the N44 region. Given that LMS and PMS stars are located closely together in the low-brightness regime in the CMD, we require examples from both populations in order for our ML models to learn to properly distinguish PMS from non-PMS stars. To find a region within the MYSST data that contains enough examples of PMS stars, we first make a very rough selection of potential candidates in the CMD using the red polygon in the left panel of Figure 2. Performing a kernel density estimate (KDE) on the spatial distribution (using a Gaussian kernel and a fixed bandwidth of $300\,\mathrm{pixels}$, i.e. $\sim 3\,\mathrm{pc}$) of this rough selection, we then determine field areas with high densities of PMS star candidates. Since the majority of these are located in the northern half of the FOV we concentrate on this region. Drawing contours at increasing significant density levels, in units of σ above the mean surface density, we find that a $2\,\sigma$ density contour, located at the western edge of the N44 superbubble, entails a large-enough sample of LMS and PMS stars. This region is enclosed by the black contour in the middle panel of Figure 2. The corresponding Hess diagram (Figure 2, right panel) shows a CMD consisting of a prominent MS as well as a nicely separated young PMS population, which provides an ideal base for the training set of our ML approach. Note that this region is also subject to significant differential reddening, covering the entire range of the extinction estimates, so that this selection already entails the broad extinction range toward N44.

Since our classification scheme distinguishes between two classes, "PMS" and "non-PMS", each star of our training set base requires a label indicating to which of the two categories it belongs. Consequently, we need to quantify which of the stars in our data set are part of the PMS and LMS populations in the low-brightness regime. To achieve this, we have devised a procedure
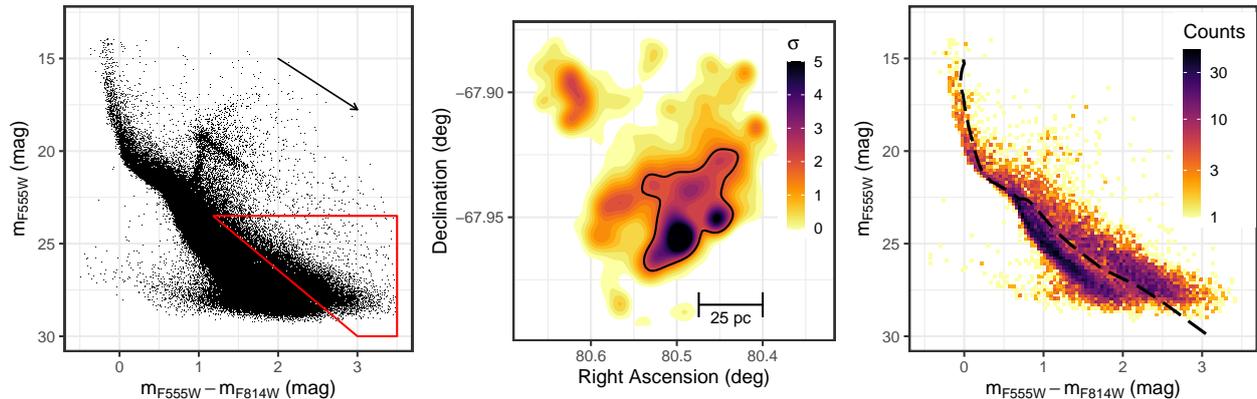
Figure 2. Left: optical CMD of the northern half of the MYSST main FOV, centered on N44's superbubble. The red polygon indicates a rough selection of PMS candidate stars used to identify a training set for our ML approach. The black arrow indicates the direction of the reddening vector of N44, as derived in Paper I. Middle: contour density plot of a kernel density estimate of the rough PMS candidates located in the northern part of the observed field view. The density levels are shown in units of σ above the mean estimated density. The contour highlighted by the solid black line indicates the region selected as a base for the training set. Right: Hess diagram of the black outlined region in the middle panel. This density diagram highlights the presence of two distinct populations of stars in this FOV, namely, a clear MS and PMS. For comparison, the black dashed line indicates a 14 Myr PARSEC isochrone, corrected for the median extinction of the stars in this region and the LMC distance modulus.

in Ksoll et al. (2018), where we fit a Gaussian mixture model to a distance metric in the CMD using the Expectation Maximization (EM) algorithm to determine a probability for every star in the low-brightness regime to be part of the PMS population. Figure A1 in the Appendix shows the selection of the low-brightness stars for this fit. Here we have excluded the UMS and red clump sources, as well as a few objects whose nature we could not identify. While the very red objects among the latter could potentially be PMS stars, which are e.g. variable sources or are undergoing an extreme accretion event, we cannot ascertain this with the MYSST data alone. Therefore, we opt to only find the most secure PMS examples here. Figure A1 also highlights the threshold line derived from PARSEC isochrones (Bressan et al. 2012), which is the basis for the CMD distance measure. Note that this selection and the fit are performed on the extinction-corrected CMD in order to achieve the best possible separation between PMS and LMS objects. We also ignore the uncertainties of photometry and extinction during the Gaussian mixture model fit, because we aim to perform a classification and not a regression, so that the precise probability values are not of great importance.

Once these probabilities are established, we assign our binary labels by selecting a threshold above which we consider a star a true PMS candidate, taking the need for a balanced (ideally 50% positive and 50% negative) training set into account. Due to the overall lower abundance of PMS stars, we cannot reach an

optimal balance, but we find that selecting a threshold probability of $p_{em} \geq 0.85$ achieves a reasonable trade-off between training set balance, strictness in our PMS example choice, and classifier performance. The strictness of the chosen threshold also indirectly accounts for the uncertainties of photometry and extinction, neglected during the fit, as this selection of PMS examples is more conservative than optimistic, already excluding sources in the transition zone that would show the most changes in PMS candidate probability due to measurement uncertainties.

3.2. RGB Training Set

Aside from the field LMS stars, which need to be distinguished from the PMS sources, old stars on the RGB can also fall into the PMS regions of the CMD owing to either distance, extinction, or simply the fact that RGB and PMS tracks can partially overlap in the CMD. Like most of the LMS stars, these RGB contaminants are either foreground or background stars of the LMC that do not belong to the young star-forming clusters we are trying to identify. As the third panel of Figure 2 indicates, our training set basis contains almost no examples of these stars. Consequently, we need to look elsewhere to find additional RGB examples so that our ML models can take these objects into account. To find such examples, we use the KDE of the PMS selection again to now identify regions within the survey that are devoid of PMS stars and entail an RGB population. The top
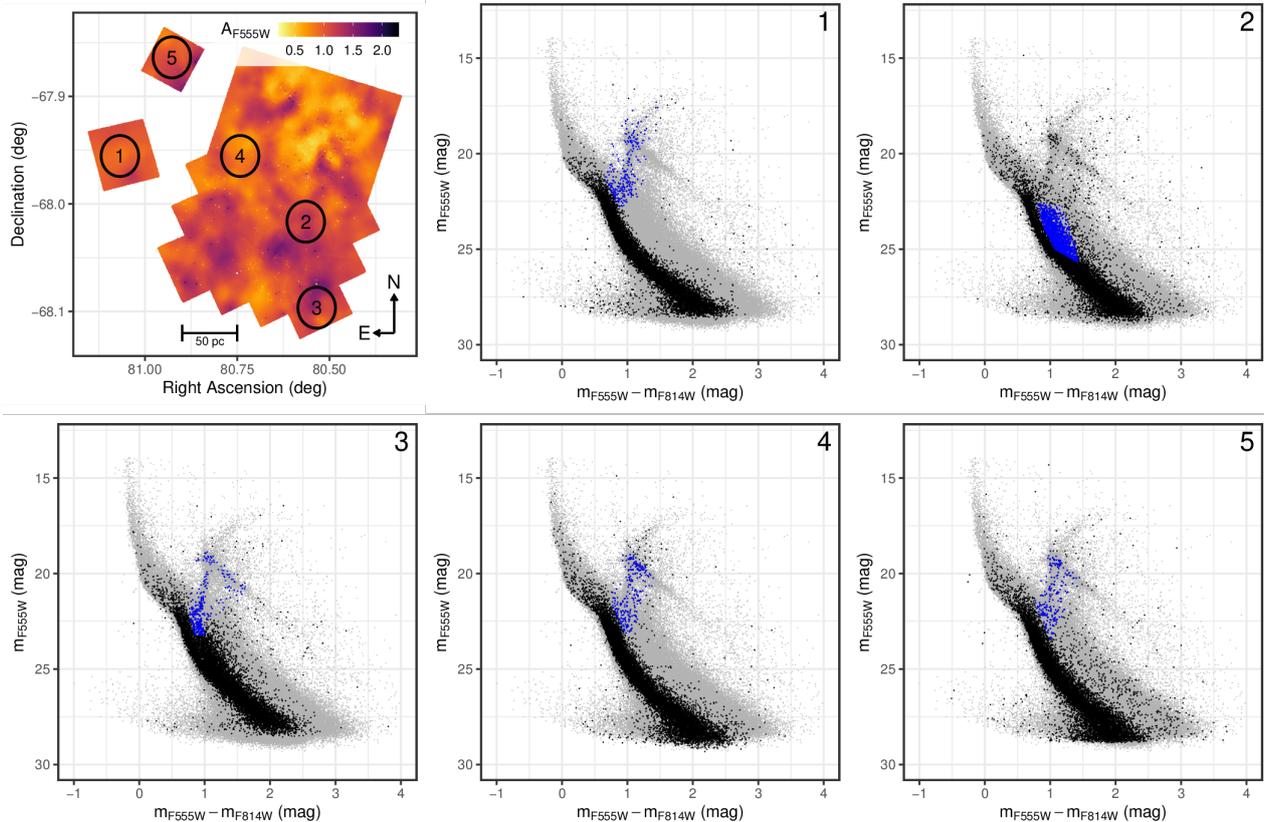
Figure 3. Extinction map of the MYSST photometric catalog (top left). The black numbered circles indicate regions that are identified to be devoid of PMS stars and used to add examples of RGB stars into the training set. The size of the circular selections is chosen to match the surface area of the training set contour in the middle panel of Figure 2. The remaining five panels show the CMDs of the corresponding circles in black in comparison to the total CMD of the MYSST data (gray). Highlighted in blue are the respective non-PMS examples added to the training set. Note that we do not select RGB samples in the top right panel, but rather an emergent feature that resembles a highly extincted MS.

left panel of Figure 3 shows five regions we have identified for this purpose, all encircling the same projected area enclosed by the 2σ irregular contour of our training set basis. We select multiple regions to probe different extinction regimes. The remaining five panels show the corresponding CMDs in comparison to the total CMD of the MYSST survey, the blue points representing the RGB examples to add to the training set. We also include a few example red clump stars along with the RGB selection to avoid potential misclassification on account of the models never having seen any red clump objects during training. Also important to note here is that we do not select RGB examples in region 2, but rather constituents of a feature that looks akin to a heavily reddened MS. This feature does not completely disappear when we correct for extinction. Given that this region appears to be more severely extinguished in the UMS extinction measurements, this feature could potentially be a heavily reddened field population behind N44 for

which we are still underestimating the reddening. Since the nature of these objects is unclear, and because this region is clearly almost devoid of young PMS stars, we decide to include this feature as negative examples so that our ML models can also take it into account.

We add these RGB examples with a fixed PMS probability of $p_{\rm em} = 0$ before applying the previously mentioned labeling threshold to the data.

3.3. Final Training Set

Figure 4 shows our final training set before application of the label threshold. In early training attempts of our ML models we realized that the prediction benefits from including the UMS (examples located at about $m_{\rm F555W} < 21$ and $m_{\rm F555W} - m_{\rm F814W} < 0.5$) as additional negative examples, something that was not necessary in our previous study (Ksoll et al. 2018). Similarly to the RGB stars, we add them with zero probability. Note that this decision will likely exclude the
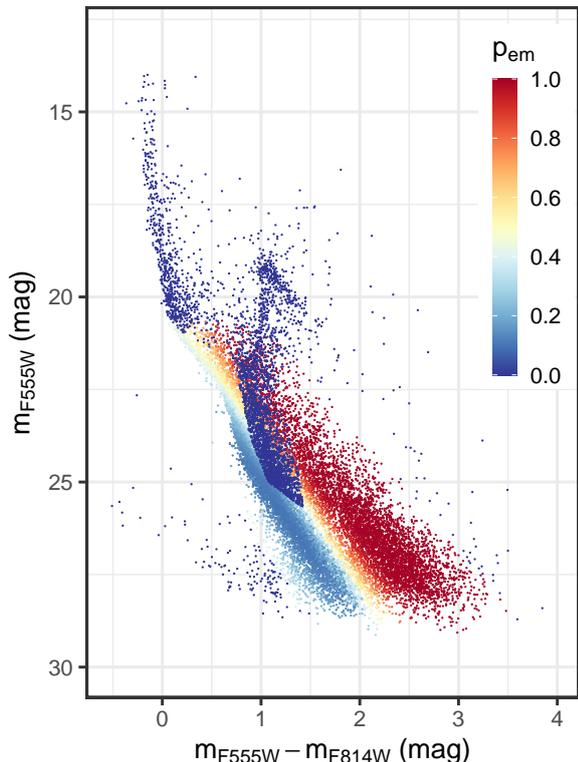
Figure 4. Optical CMD of the final training set selection. Each star is color-coded according to the derived probability that it belongs to the PMS population in the CMD. Note that the UMS and additional RGB examples are included with a fixed probability of 0.

detection of more massive, brighter PMS stars that are close to joining the MS, such as Ae sources. We also re-add the low-brightness objects of unclear nature, which were excluded during the EM fit, as negative examples (i.e. with $p_{em} = 0$). For the most part these are located roughly at $m_{F555W} > 25$ and $m_{F555W} - m_{F814W} < 1$, as well as around $28 > m_{F555W} > 22$ and $1.5 < m_{F555W} - m_{F814W} < 4$.

With that, our training set entails 17,942 stars, of which 5512 ($\sim 31\%$) are PMS candidate stars with $p_{em} \geq 0.85$. Again, the balance between positive and negative examples within the training set is not optimal, but with about a third of the data being positive examples, we believe that our selection is robust enough to not suffer from imbalance issues. At this point it is also important to note that the PMS candidate examples in our training set appear to be mostly younger than ~ 15 Myr when compared to PARSEC isochrones (see Figure 2, right panel). As our ML classification approach will find the siblings of the training PMS candi-

Table 1. Performance Summary for SVM and RF

| Performance | Method | | | |
| | SVM | | RF | |
Measure	Train	Test	Train	Test
Accuracy	0.9851	0.9807	0.9709	0.9680
Balanced accuracy	0.9800	0.9737	0.9628	0.9593
ROC AUC	0.9986	0.9976	0.9957	0.9950
F_1 score	0.9755	0.9683	0.9521	0.9477

NOTE—Both models are trained and tested on the same subsets for comparability.

dates across all of N44, this means in the following that we will recover only the most recent sites of star formation, younger than ~ 15 Myr. Therefore, our method is not sensitive to potential low-mass PMS stars from even earlier star formation events, which are still in the formation process but very close to joining the MS.

Lastly, we also have to note that we do not account for active galactic nuclei (AGNs) or unresolved (background) galaxies in our training data, because we aim for a MYSST survey intrinsic approach and distinguishing these sources with the available data is not straightforward. Consequently, there may be some minor contamination by these types of sources in our training set.

3.4. Training and Test Results

Having established the training set, we follow the approach of Ksoll et al. (2018), training an RF (Breiman 2001) and SVM (Cortes & Vapnik 1995) to distinguish between the "PMS" and "non-PMS" classes based on the photometry in F555W and F814W, as well as the estimated extinction in the F555W filter A_{F555W}. Note that within the method framework established in Ksoll et al. (2018) we do not consider photometric uncertainties. They do not contribute further information when considered as features, and in addition, the implementations of SVM and RF do not have mechanisms to treat uncertainty. For training we split the data set established in the previous section 70:30 into a training and held-out test subset. We use the latter to ascertain training success and performance on unknown data (with known labels) by computing the accuracy, the balanced accuracy, the area under the receiver operating characteristic (ROC AUC) curve (for a detailed description of these performance measures, see e.g. the Appendix in Ksoll et al. 2018) and F_1 score,

$$F_1 = \frac{2TP}{2TP + FN + FP}, \tag{1}$$

where TP, FP, and FN denote the number of true
positives, false positives, and false negatives, respec-
tively. We train both algorithms using a 10-fold cross-
validation, repeated five times, on the training subset us-
ing the ROC AUC as the performance metric for model
selection. For the SVM we employ a Gaussian radial
basis kernel, and we find the best RF solutions em-
ploying 500 trees. As we perform predictions on only
three features, the magnitudes in F555W and F814W
and A_{F555W}, each tree will consider all of these for the
split decisions during tree construction. Aside from a
predicted label, "PMS" or "non-PMS", we setup the
two classifiers such that they also provide a probability
for the "PMS" class. For the RF this probability is es-
timated by the fraction of votes among the 500 trees for
the "PMS" class, while we use Platt's posterior prob-
abilities (Platt 1999) to perform this estimate for the
SVM model.

Table 1 summarizes the training results and perfor-
mance of both algorithms on the held-out test set. Over-
all we find excellent results for both methods. With
accuracies, both regular and balanced, above 96% and
ROC AUC as well as F_1 scores close to the optimal value
of 1, our ML classification approach shows great success
for the given identification task. The almost-equal per-
formance results on the training and test subset across
both methods further indicate that the trained models
do not suffer from overfitting. Comparing the two algo-
rithms we find that the SVM does slightly better than
the RF achieving the highest scores across all measures.
However, given the small differences in the performance
scores, it is safe to say that they exhibit an equal success
rate.

4. IDENTIFICATION OF PMS STARS

Encouraged by our results on the training and test
data, we use the trained models to identify the PMS
stellar content of the entire MYSST survey by classify-
ing all 461,684 objects. The individual prediction re-
sults of the two complementary ML approaches, SVM
and RF, in the form of CMDs color-coded according to
the predicted PMS candidate probabilities, as well as
diagrams of the spatial distribution of the most likely
PMS candidates, can be found in Figures A2, A3 and
A4 in the Appendix. Note that the PMS probabilities
returned by SVM and RF are a measure of the model's
confidence in the prediction of the "PMS" class and *not*
the probabilities derived during our Gaussian mixture
model fit.

In total the SVM identifies 39,818 PMS candidates at
a probability of $p_{\mathrm{SVM}} \geq 0.5$ in the main FOV with a
subset of most likely ($p_{\mathrm{SVM}} \geq 0.95$) candidates consist-
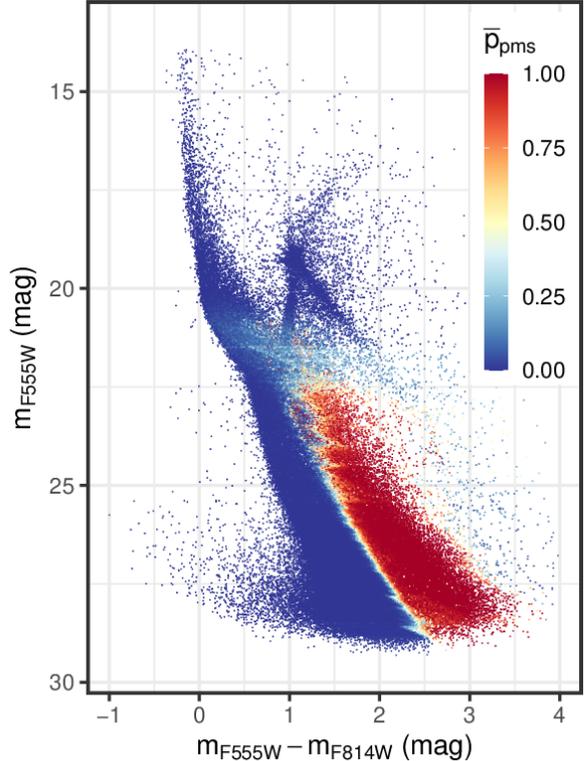
Figure 5. Optical CMD of the MYSST photometric catalog.
Each star is color-coded according to the mean predicted
probability between the SVM and RF model that it belongs
to the PMS.

ing of 29,571 stars, while the RF finds 41,909 and 26,610
candidate objects in these two categories, respectively.
Therefore, it appears that the SVM is slightly more con-
servative in the total predicted number of PMS candi-
dates, while the RF seems to put tighter constraints
on the most probable PMS constituents. With about
39,000 and 25,500 common predictions in the $p \geq 0.5$
and $p \geq 0.95$ regimes, respectively, both methods nicely
agree on the identified PMS population.

Looking at the predictions more in detail, the SVM ex-
hibits a rather smooth decision boundary in the CMD
(Appendix, Figure A2, left panel), while the RF en-
tails a more irregular zig-zag-shaped class separation,
likely an artifact of the underlying partitioning strat-
egy of the RF trees in the low-dimensional feature space
of our problem. We also see that both classifiers return
fairly sharp decision boundaries between the "PMS" and
"non-PMS" classes. From a physical standpoint this
may not seem intuitive, because there is source confu-
sion between the LMS and PMS in the low-brightness
regime and our Gaussian mixture model fit did indeed
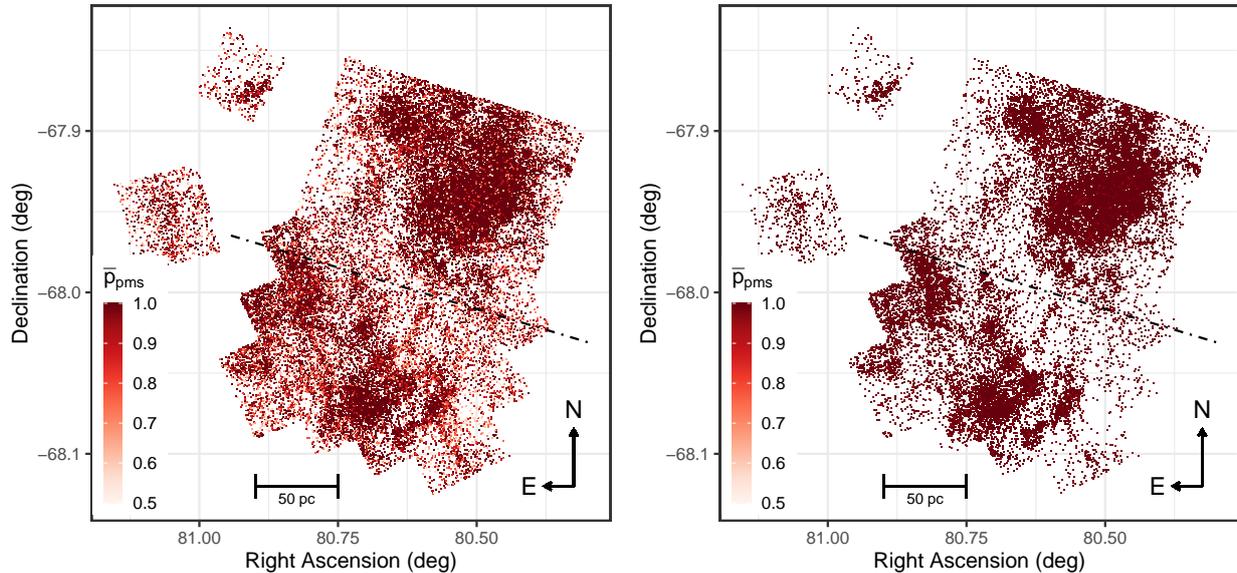
Figure 6. Spatial distribution of the predicted PMS candidate stars. The left panel shows all stars with $\bar{p}_{\mathrm{pms}} \geq 0.5$, while the right panel indicates the positions of the 28,678 most probable candidate PMS stars with $\bar{p}_{\mathrm{pms}} \geq 0.95$. The black dotted-dashed line in both diagrams indicates our north/south division of the FOV for analysis purposes.

show a relatively broad transition from one to the other population (c.f. Figure 4). It is important to emphasize here that this sharp decision boundary is not a physical one, but the one derived by the models to distinguish the two *labels* "PMS" and "non-PMS" based on the examples in the training set. Since we do not perform a regression on PMS probabilities, but a classification in a low-dimensional feature space, the models can, therefore, determine a sharp boundary between our strictly chosen PMS and non-PMS examples in areas where the two classes do not overlap significantly.

A direct star-by-star comparison of the predicted PMS candidate probability (see the right panel of Figure A2 in the Appendix), shows that the RF tends to make more conservative predictions in the CMD area where PMS and RGB overlap. The SVM, on the other hand, exhibits a more conservative decision boundary between the LMS and PMS in the very low-brightness regime. Here we also find that the RF considers several red objects of unclear nature to the right of the PMS as potential candidates, in contrast to the SVM. These very red objects could be young PMS stars that are e.g. undergoing an extreme accretion event or are variable sources during an event of heightened activity. Since we cannot establish the nature of these objects with the MYSST data alone, we consider the latter RF predictions to be debatable, concluding that the SVM returns more robust results here. On the other hand, there are also

some SVM PMS predictions fairly close to and to the left of the RGB, which are likely mispredictions and are not considered as candidates by the RF.

Overall, we come to the same conclusion as in our previous study (Ksoll et al. 2018), that a combination of the two classification outcomes provides the most robust prediction result for the PMS stellar content of N44. Figure 5 exhibits the classification results if we average the predicted PMS probabilities between SVM and RF as the color code of every star in the CMD. Excluding the two reference fields of the survey, this approach returns a total of 40,509 PMS candidates with $\bar{p}_{\mathrm{pms}} \geq 0.5$ within the main FOV and a most probable subset consisting of 26,686 stars with $\bar{p}_{\mathrm{pms}} \geq 0.95$. Figure 6 shows the spatial distributions of these PMS candidates across the area of N44. Notable here is that among the most probable set a majority of 16,976 PMS candidates is located in the northern half of the survey, in and around the massive superbubble of N44, while only 9710 prospective PMS stars are distributed in the southern region. The black dotted-dashed line in Figure 6 indicates our north/south division for the purpose of this discussion. Within the northern part we can see that the PMS stars are mainly concentrated toward the rims of N44's bubble, especially so the western and northwestern edge but excluding the southeastern corner.

We also recover a number of PMS candidates in the two reference fields of the survey, i.e. a total of 646 at $\bar{p}_{\mathrm{pms}} \geq 0.5$ and 346 at $\bar{p}_{\mathrm{pms}} \geq 0.95$ for the northern field, while the southern one hosts 987 and 439 sources in these two confidence regimes, respectively. The candidate stars in the northern field are concentrated almost entirely at the southwestern corner, forming a distinct clump, whereas in the southern field they are more evenly distributed without any apparent structures. Regarding the candidates in the southern field, it should be noted that in Paper I we find only very few UMS candidate stars there for the approximation of individual extinction. Additionally, the UMS status of the selected stars remains unclear, so that we believe the estimated extinction values in the southern field to be the most uncertain. Consequently, we recommend treating the identified PMS candidates in this field with caution.

5. SPATIAL DISTRIBUTION OF PMS STARS

To better understand the star-formation processes in N44, we investigate the spatial distribution of the PMS candidate stars in more detail. We employ a nearest neighbor search to determine the surface density of PMS candidates and characterize their clustering properties. We also look at the correlation of the PMS candidate stars with other star formation indicators, specifically, we compare with the positions of the known O stars, B stars and YSOs in the region, as well as CO, Hα and dust emission observations. Additionally, we evaluate how well our spatial PMS candidate distribution matches the one derived by Zivkov et al. (2018) from VMC observations of N44.

5.1. Location of PMS Stars

To further ascertain the validity of our PMS identification and to study the spatial distribution of these stars, we perform a nearest neighbor density estimation (NNDE). We compute the local source density n_j, first introduced in astronomy by Casertano & Hut (1985), as

$$n_j = \frac{j-1}{\pi r_j^2}, \qquad (2)$$

where r_j denotes the distance to the jth-nearest neighbor, on a regular grid within the MYSST main FOV. Note that we modify the density estimate to a surface number density here, instead of the mass density in Casertano & Hut (1985). Similar to our previous application of a KDE, we then compute surface density contours in terms of significance σ above the mean estimated density. We find that employing the distance to the 20th-nearest neighbor, corresponding to $j = 20$ in Eq. 2, offers a reasonable compromise between reso-

lution and statistical significance and allows us to highlight the structures of the identified PMS clusters. Due to the difference in number and spatial distribution of the identified PMS candidates between the northern and southern half of the survey (see black dotted-dashed line in Figure 6) we perform the NNDE separately on both regions to better quantify the clustering properties of PMS stars. For the same reasons we also treat the two reference fields individually. Both panels of Figure 7 show the corresponding nearest neighbor surface density contours. Due to the individual treatment of the four regions, the nearest neighbor density differs for the same σ-significance level between regions. For instance, at 1σ the nearest neighbor densities are at 2.62 and 1.49 pc^{-2} in the northern and southern halves of the main FOV, while they reach only 0.84 and 0.27 pc^{-2} in the northern and southern reference fields, respectively. As the overall nearest neighbor density in the southern reference field is fairly low, barely reaching 0.5 pc^{-2} even at a 3σ significance, it is obvious that the structures here are not entirely comparable to those found in the main FOV.

For comparison, the right panel of Figure 7 also provides the location of O stars derived from MUSE observations (McLeod et al. 2019, note that this survey only covered the northern half of the MYSST FOV), additional known O-type sources in the SIMBAD database (see Appendix, Table A1), and massive YSOs identified from Spitzer observations (Chen et al. 2009) and Spitzer data combined with optical and near infrared photometry (Carlson et al. 2012). Additionally, the left panel indicates the prominent H II regions of N44, as determined by Pellegrini et al. (2012) and defined in McLeod et al. (2019). This diagram confirms that the PMS stars identified by our ML classification are primarily located within the H II regions of N44. The only notable exception here is the H II region L219, where we do not find a prominent overdensity of PMS candidate stars. Since a large part of this region falls outside of the MYSST FOV, similar to L198 and L194, it is not unlikely that we are simply missing most of the associated PMS clusters. Note also that Pellegrini et al. (2012) does not find HII regions associated with the structures of PMS candidates we identify in the two reference fields.

In Paper I we use a selection of ∼ 1300 UMS stars to derive the extinction toward the region. As previously mentioned, the MYSST survey misses the most massive stars of N44 owing to saturation effects. Therefore, this selection consists primarily of late O- and early B-type UMS stars. Comparing this population of young massive stars to our PMS density maps (see Appendix, Figure A5), we also find evidence that they are preferably lo-
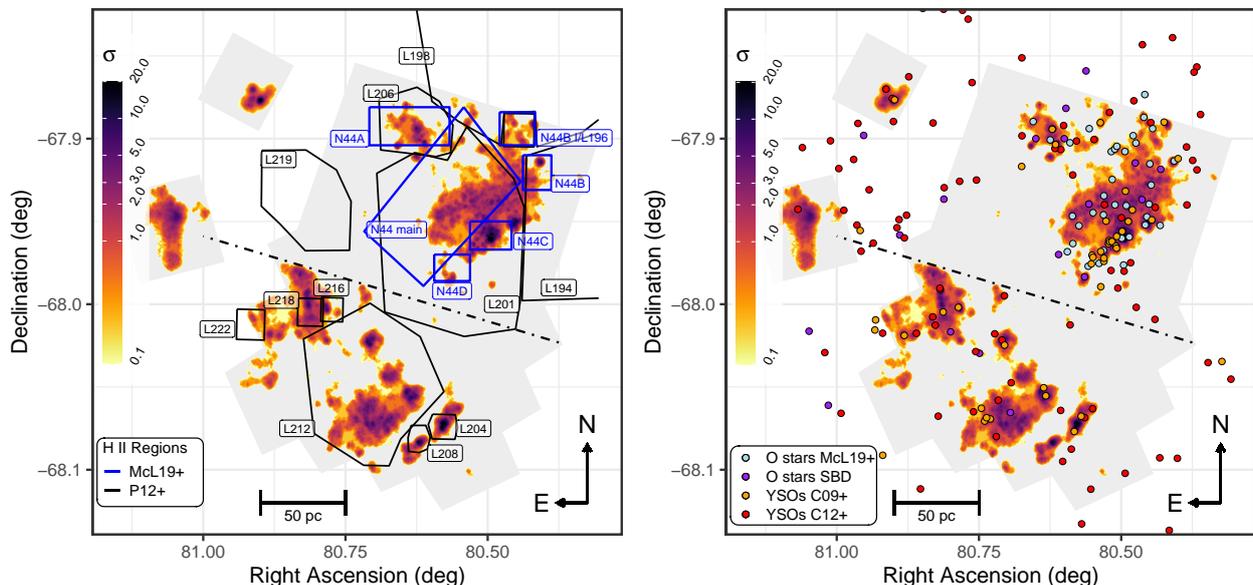
Figure 7. Left: spatial nearest neighbor contour density diagram of the most probable candidate PMS stars of the MYSST survey. The color-coding represents the nearest neighbor density in steps of σ above the mean density. The black dotted-dashed line indicates the north/south separation for the NNDE. The polygons mark known H II regions around N44, in black as determined by Pellegrini et al. (2012) and in blue as defined in McLeod et al. (2019). Note that the boundaries "N44B I" from McLeod et al. (2019) and "L196" from Pellegrini et al. (2012) coincide. The underlying gray shaded regions (in both panels) indicate the MYSST coverage for comparison. Right: same diagram as in the left panel, but overlaid with the positions of O stars (light-blue points) as identified by MUSE observations (McLeod et al. 2019). Note that McLeod et al. (2019) only covered the northern half of the MYSST FOV. The purple points signify other known O stars in the SIMBAD database (Appendix, Table A1) that are not covered by McLeod et al. (2019). The orange points indicate massive YSOs identified from Spitzer observations of N44 as found by Chen et al. (2009). Lastly, the red points mark additional YSOs discovered by Carlson et al. (2012), excluding matches within 1 arcsec with the Chen et al. (2009) list. Note that 14 YSOs from Chen et al. (2009) and 31 from Carlson et al. (2012) fall outside the shown region.

cated in correspondence of the PMS clusters, as more than 35% (62%) of them fall into the 1σ (0σ) PMS density contours. For comparison, in a uniform random distribution (averaged over 100 random realizations) only $9.6\% \pm 0.5\%$ ($30.0\% \pm 0.5\%$) of objects would fall within the same contours. This provides additional confirmation that the PMS candidates we identify tend to be located in the vicinity of more massive young UMS stars.

Using Hess diagrams to identify PMS regions as density excesses over local field populations, Zivkov et al. (2018) recently provided a PMS surface density map of N44 based on data by the VMC survey. In the left panel of Figure 8 we show their PMS surface density contours in comparison to our PMS nearest neighbor density map. Aside from two of their distinct density peaks that fall outside of the MYSST coverage, we find a good match to our nearest neighbor density map within the southern half of the main MYSST FOV. In the northern half we also have a decent agreement along the western edge of the main bubble. However, we identify three

density peaks from Zivkov et al. (2018) as well that do not have a significant counterpart in our PMS nearest neighbor density map. These Zivkov et al. (2018) density peaks are located at the northern bubble rim (R.A. = 80.53°; Dec. = −67.90°), the eastern bubble edge (R.A. = 80.73°; Dec. = −67.94°) and just south of the bubble (R.A. = 80.67°; Dec. = −67.98°), respectively. The discrepancy in these three regions could be an effect of both the angular resolution and completeness differences between the VMC and MYSST surveys. Employing the VISTA telescope, the VMC project achieves an angular resolution on the order of 0.34″, a value that is almost 10 times larger than the 0.04″ resolution obtained with the HST in the MYSST observations. Additionally, Zivkov et al. (2018) state that the 5σ magnitude limit of their photometry catalog corresponds to the brightness of 1 Myr old PMS stars with 0.7 $M_\odot$ (reddening corrected), while the MYSST survey reaches down to 0.09 $M_\odot$ (albeit unreddened) for stars of that age (Paper I).
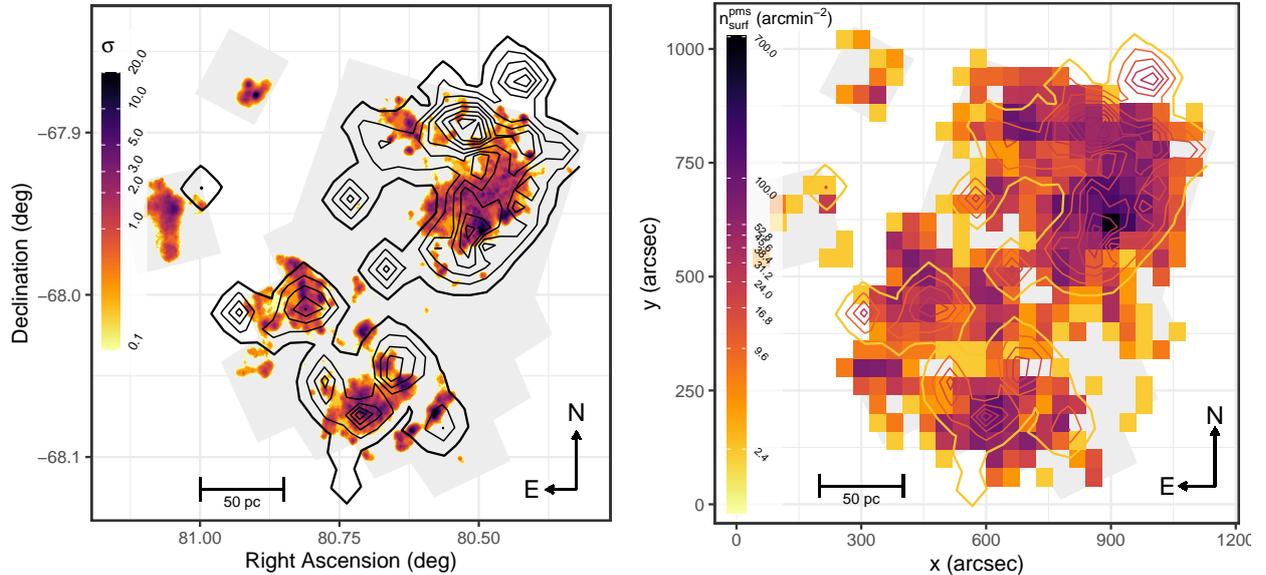
Figure 8. Left: same PMS nearest neighbor density diagram as in Figure 7. Overlaid in black are the PMS density contours derived from VMC observations of N44 by Zivkov et al. (2018). The outermost black contour indicates a number density of 2.4 arcmin^{-2}, while each consecutive inner contour indicates an increase by 3×2.4 arcmin^{-2}. For comparison, the gray shaded region marks the coverage of the MYSST survey in both panels. Right: 2D binned surface density diagram with $40'' \times 40''$ bins of a subset of our PMS catalog matching the PMS mass completeness limit of the VMC survey used in Zivkov et al. (2018). Overlaid with the same color scheme as the 2D density map are the PMS surface density contours from Zivkov et al. (2018) to allow for quick comparison.

To test whether the completeness (and resolution) differences between the MYSST and VMC survey can indeed explain the missing density peaks in our PMS distribution in the three identified regions, we select a subset of our PMS candidate catalog that matches the VMC PMS mass limit of 0.7 $M_\odot$. Using the 1 Myr PAR-SEC isochrone and accounting for the average extinction measured in Paper I, the 0.7 $M_\odot$ cutoff translates to a limiting magnitude of 24.86 mag in F555W. Selecting only PMS candidates brighter than this limit reduces our catalog of most likely PMS sources from 27,471 to only 4002 across the entire MYSST FOV, including the two reference fields. Missing more than 85% of our identified PMS candidates from this limit alone, it is not unlikely that the PMS density map derived from the VMC data overestimates the significance of these three regions compared to the rest. In fact, Zivkov et al. (2018) only find about 1000 ± 38 PMS stars (as a lower limit) in N44 based on the VMC data.

To approximate the spatial resolution of the Zivkov et al. (2018) approach for identifying PMS regions – they use a grid of overlapping circular elements with a radius of $40''$– we compute a 2D binned surface density map with $40'' \times 40''$ bins from the reduced PMS candidate catalog. The right panel of Figure 8 shows this map in

comparison to the Zivkov et al. (2018) PMS density contours, where bins and contours share the color scheme to easily highlight matching number density levels (in arcmin^{-2}). While our low-resolution 2D number density map generally tends to larger values, in particular at the western edge of the bubble, we find that the surface densities in the three regions in question actually match up reasonably well. It is also interesting to note that the small density peak found by Zivkov et al. (2018) close to our southern reference field is matched fairly well in our low resolution density map, even though half of it is actually outside the MYSST FOV. Given these results, the missing density peaks in our full resolution nearest neighbor density map appear to be well explained as a result of the lower completeness and spatial resolution of the VMC data. Therefore, we conclude that our results agree well with the Zivkov et al. (2018) study and provide a significant extension toward very low mass PMS stars at a higher spatial resolution.

Lastly, we also compare our spatial PMS distribution with other star formation tracers, such as gas and dust emission. In Figure 9 our PMS nearest neighbor density distribution is shown in comparison to contours of CO emission derived from the Magellanic Mopra Assessment (MAGMA; Wong et al. 2011, 2017) survey. Here
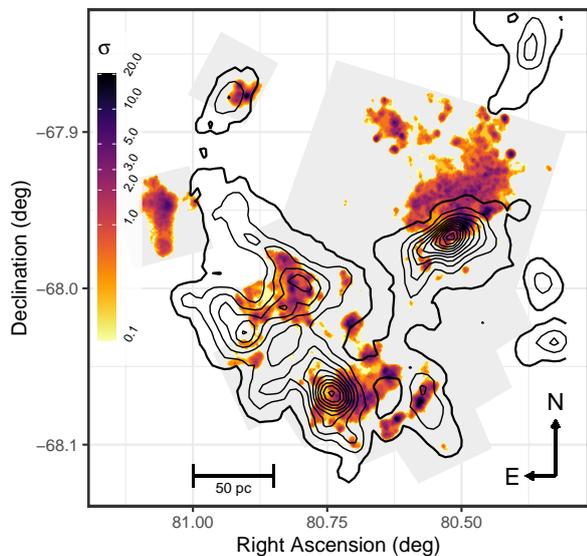
Figure 9. Same PMS nearest neighbor density diagram as in Figure 7. Overlaid in black are intensity contours of CO emission as observed by the MAGMA survey (Wong et al. 2011, 2017). The outermost contour marks the mean CO intensity in the FOV at 1.7 K km s^{-1}, and each consecutive inner contour marks an increase in intensity by $1\sigma = 3.5$ K km s^{-1} up to the 10σ level for the innermost line. For comparison, the gray shaded region marks the coverage of the MYSST survey.
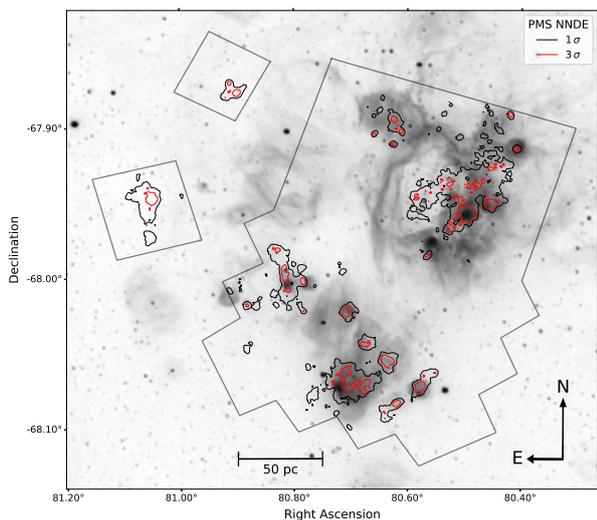
Figure 10. Inverted grayscale image of Hα emission in N44 as captured by the MCELS survey (Pellegrini et al. 2012). Overlaid in black and red are our 1 and 3σ PMS nearest neighbor density contours (see Figure 7) for comparison. The gray outlines indicate the coverage of the MYSST survey.

the outermost contour signifies the mean CO intensity in the FOV of 1.7 K km s^{-1}, while each subsequent contour marks an increase by $1\sigma = 3.5$ K km s^{-1} up to a maximum of 10σ. We find a clear correlation of enhanced CO emission to regions of high PMS density in the southern half of the main FOV. In the northern half there is also a very prominent peak in the CO emission that partially coincides with the highest PMS nearest neighbor density at the western edge of the bubble. Quite notable is the absence of CO emission along the northern bubble rim and inside of the bubble, where we still find notable structures of PMS sources. As the very massive stars have cleared out the gas and dust in the bubble, the absence of CO emission there is not surprising. Interesting as well is a small peak of CO emission in the northern reference field, coinciding with the PMS overdensity we have identified there. In contrast, the southern field does not exhibit any CO emission.

Figure 10 shows an inverted grayscale image of Hα emission in N44 as captured by the Magellanic Cloud Emission-Line Survey (MCELS; Pellegrini et al. 2012) in comparison to the 1σ (black) and 3σ (red) contours derived from our PMS nearest neighbor density. Overall this figure demonstrates that the most significant structures of our PMS candidate stars appear correlated with enhanced Hα emission. As Hα traces regions of ionized hydrogen, this falls in line with our previous assessment that our PMS candidates correlate with the identified H II regions in N44 (which is not entirely surprising, as the MCELS Hα images contributed to the definition of the H II region boundaries in Pellegrini et al. (2012) to begin with). Notable exceptions to this correlation with enhanced Hα emission are the part of the large 1σ contour at the western bubble edge (N4, c.f. Fig. 12, Section 5.2) that extends into the bubble interior and both structures found in the two reference fields. For the bubble interior this is, again, consistent with the fact that the very massive stars located here have driven out most of the gas and dust of their natal environment.

In the left panel of Figure 11 we overlay our 1σ (black) and 3σ (red) PMS density contours on an inverted grayscale image of dust emission at 8 μm, i.e. emission from polycyclic aromatic hydrocarbon (PAH), as observed by the SAGE survey (Meixner et al. 2006) with the Spitzer Space Telescope. In the right panel we show the same comparison with a color composite image of dust emission, combining 70 μm emission Spitzer observations from SAGE with 160 μm and 350 μm Herschel images from the HERschel Inventory of The Agents of Galaxy Evolution project (HERITAGE; Meixner et al. 2013). Visual inspection of both dust maps reveals that many of the structures at 1 and 3σ of our PMS near-
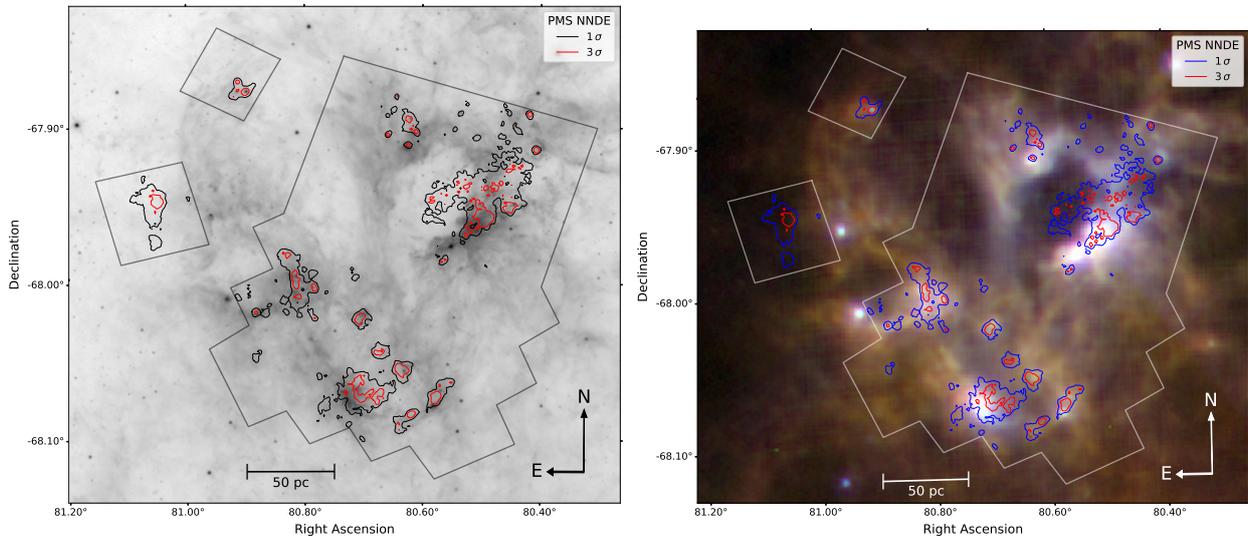
Figure 11. Inverted grayscale image of dust emission at 8 μm in N44 from the SAGE survey (Meixner et al. 2006) of the LMC (left). Overlaid in black and red are the 1σ and 3σ nearest neighbor density contours of the most likely MYSST PMS candidates (as in Figure 7). They gray outline indicates the coverage of the MYSST survey for comparison. Right: color composite dust emission image of N44 combining Spitzer 70 μm in blue, Herschel 160 μm in green and Herschel 350 μm in red. Overlaid in blue and red are the 1σ and 3σ PMS density contours, and the MYSST coverage is indicated in gray. The Herschel observations were taken as part of the HERITAGE survey (Meixner et al. 2013).

est neighbor density distribution coincide with areas of increased dust emission, although the 3σ density peaks are often slightly offset from the maxima of dust surface brightness (e.g. in Region S4, see Figure 12, Section 5.2). This finding is consistent with the hypothesis that in large concentrations of young stars the irradiation of the dusty remnants of the stellar birth environments leads to bright dust emission in the far-IR as the dust reemits the incoming stellar radiation at longer wavelengths. The large 1σ structure (N4, c.f. Fig. 12, Section 5.2) that partially extends into the bubble is, as for the Hα emission, again one of the notable exceptions here, explained of course by the feedback of the very massive stars in the bubble interior having cleared out gas and dust. There are also three more prominent structures in the southern half of the main FOV (S6, S7, S8; see Fig. 12, Section 5.2) that do not appear particularly bright in the dust emission. In the reference fields we find again slightly enhanced emission for the structure found in the northern one, but almost no dust emission in the southern field.

5.2. *Identifying PMS Clusters*

The spatial distribution of PMS candidates in N44, e.g. as indicated in Figure 6, clearly shows that these stars are distributed in a hierarchical and highly clustered fashion. To identify the PMS clusters, we utilize the nearest neighbor density map (Figure 7) to first find all density contours at a 1σ significance level. These

contours define our preliminary PMS cluster candidates. We then down-select the most prominent PMS clusters if they fulfill a *persistence* criterion of exhibiting substructures at 3σ density significance. Preliminary, we remove all candidate contours that contain less than 100 stars in total (PMS and non-PMS), as they are likely an outcome of noise fluctuations at the 1σ level and would therefore never fulfill the persistence criterion in the first place. The limit of 100 corresponds to approximately the square root of the number of all PMS sources located in the 1σ contours.

Applying this contour-density-based clustering approach, we identify seven prominent PMS structures at 1σ significance in the northern half of the FOV, nine distinct clusters of PMS candidates in the south, and one each in the two reference fields. Again, for this step we use the individual NNDEs of the northern and southern half of the main FOV and the reference fields to be more sensitive to local density structures by avoiding the large difference in stellar numbers between the individual regions. Figure 12 indicates the spatial positions of these 18 prominent PMS structures. Note that we only show the 1σ contours (black) of the structures that pass our persistence criteria here. In this figure we also highlight the subclusters at 3σ significance (white) of the prominent 1σ structures, excluding, however, those that do not contain at least 50 stars in total. Again, this serves to decrease statistical noise, this time at the 3σ level,
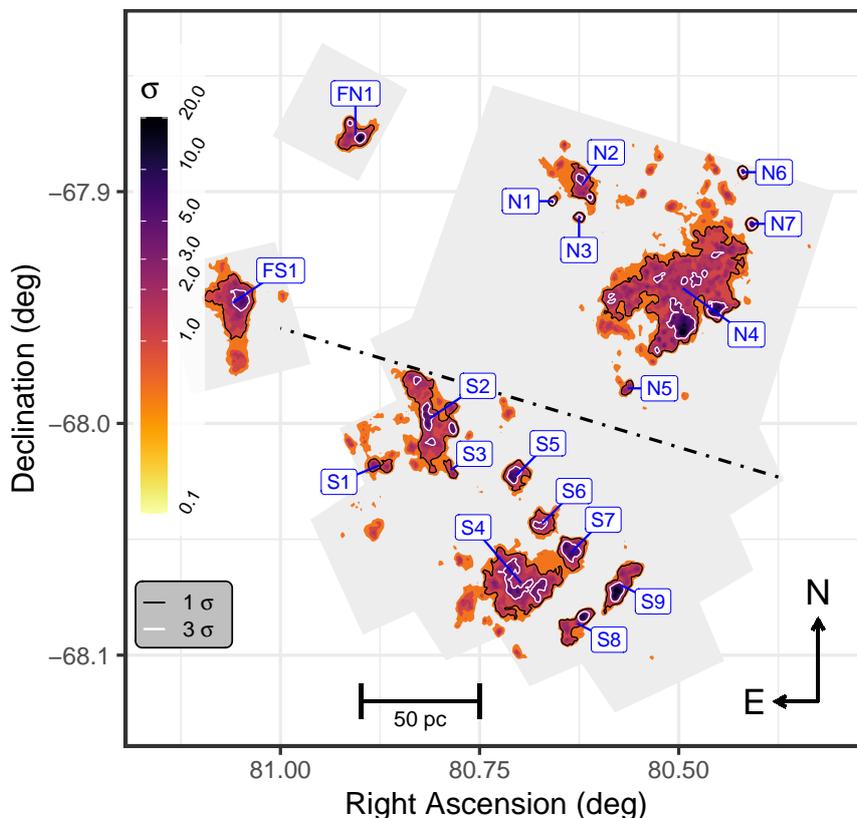
Figure 12. NNDE contour density diagram of the most likely PMS population of N44. The black dotted-dashed line indicates the north/south separation chosen for the NNDE. The gray shaded regions mark the MYSST coverage for comparison. The black contours indicate density structures at 1σ significance that harbor a minimum of 100 stars and have substructures that persist up to 3σ significance in density. The white contours mark the 3σ subclusters that entail at least 50 stars. The labels give an ID of the 1σ structures for easier reference. Properties of these 1σ structures and 3σ subclusters are summarized in Tables 2 and A2.

with 50 being approximately the root of the number of all PMS sources inside the 3σ contours.

Comparison with Figure 7 indicates that almost all of the structures that we identify as PMS clusters are close to or harbor one or more massive O stars/YSOs. Star formation theory predicts, and recent studies confirm (Cignoni et al. 2015; Stephens et al. 2017), that PMS clusters are located primarily in the vicinity of young massive stars. Therefore, our approach identifies clusters of PMS stars where one would expect to encounter them. Combined with the fact that we also find them within the H II regions, which are the remnants of recent star formation events, the comparison with the MUSE (McLeod et al. 2019), SIMBAD, and YSO (Chen et al. 2009; Carlson et al. 2012) data provides an independent confirmation of the validity of our ML classification approach. There is one possible exception, namely, the H

II region N44D, where we only find a small amount of PMS stars that do not immediately coincide with the two MUSE O stars and three YSOs located there, but only with one O-type source in the SIMBAD database. This particular region suffers from a large amount of incompleteness in the MYSST survey due to saturation effects likely caused by the massive O stars/YSOs at its center (see Paper I). We note that the small offset between our identified PMS grouping and the other O star/YSOs within this H II region is consistent with the hypothesis that we are simply missing most of the PMS stars around these massive objects.

5.3. Properties of the PMS Clusters in N44

To further characterize the properties of the PMS clusters, we first determine their center-of-mass position on the sky. This is simply obtained as the average of the position of all cluster members, because we do

Table 2. Properties of the 1σ PMS Density Structures

ID	R.A.$_{\rm cent}$	Decl.$_{\rm cent}$	$A_{\rm surf}$	$R_{\rm eff}$	N_*	$n_{\rm surf}^{total}$	$N_{\rm PMS}$	$n_{\rm surf}^{pms}$	$N_{\rm O}^{M19}$	$N_{\rm O}^{SD}$	$N_{\rm YSO}^{C09}$	$N_{\rm YSO}^{C12}$	$N_{\rm sub}^{2\sigma}$	$N_{\rm sub}^{3\sigma}$	Q	σ_Q
	(deg)	(deg)	(pc^2)	(pc)		(pc^{-2})		(pc^{-2})								
N1	80.6585	-67.9041	10.3	1.8	164	15.9	50	4.9	0	0	0	0	1	1(1)	0.76	0.5
N2	80.6212	-67.8971	95.9	5.5	1516	15.8	434	4.5	1	0	2	1	1	4(2)	0.68	0.47
N3	80.6251	-67.9112	12.8	2	197	15.4	66	5.2	0	0	0	0	1	1(1)	0.77	0.52
N4	80.4950	-67.9416	1365.3	20.8	23164	17	6397	4.7	16	2	15	5 (2)	28	29(10)	0.64	0.43
N5	80.5641	-67.9849	18.2	2.4	252	13.9	75	4.1	0	1	0	0	1	2(0)	0.72	0.5
N6	80.4202	-67.8916	15.6	2.2	239	15.3	85	5.4	0	0	1	0	1	1(1)	0.78	0.5
N7	80.4087	-67.9139	19.6	2.5	370	18.9	126	6.4	1	0	0	0	1	1(1)	0.8	0.54
S1	80.8765	-68.0182	37.1	3.4	367	9.9	86	2.3	0	0	1	0	2	1(0)	0.60	0.43
S2	80.8122	-67.9978	367.9	10.8	4312	11.7	902	2.5	0	0	2	2	8	5(4)	0.55	0.39
S3	80.7864	-68.0201	19.5	2.5	257	13.2	39	2.0	0	0	0	0	1	1(0)	0.58	0.43
S4	80.6981	-68.0688	504.4	12.7	6500	12.9	1308	2.6	0	1	3	3 (1)	10	4(2)	0.71	0.46
S5	80.7043	-68.0226	75.8	4.9	1045	13.8	233	3.1	0	0	1	1	1	1(1)	0.73	0.47
S6	80.6714	-68.0432	68.8	4.7	1112	16.2	181	2.6	0	0	0	0	1	1(1)	0.75	0.46
S7	80.6347	-68.0556	103.8	5.7	1524	14.7	336	3.2	0	0	2	1 (1)	1	1(1)	0.76	0.48
S8	80.6275	-68.0863	88.3	5.3	1019	11.5	255	2.9	0	0	0	0	3	2(1)	0.54	0.41
S9	80.5722	-68.0698	131.4	6.5	1644	12.5	454	3.5	0	0	2	3 (1)	1	3(1)	0.63	0.48
FN1	80.9058	-67.8756	102.7	5.7	1286	12.5	179	1.7	0	0	1	2	2	3(2)	0.65	0.44
FS1	81.0598	-67.9480	272.5	9.3	3051	11.2	116	0.4	0	0	0	1	3	3(1)	0.68	0.45

NOTE— Properties of the 1σ PMS density structures that persist with substructures up to 3σ in density and consist of at least 100 stars. Listed are the structure ID as in Figure 12, the right ascension R.A.$_{\rm cent}$ and declination decl.$_{\rm cent}$ of the structure center, the surface area $A_{\rm surf}$ enclosed by the given density contour, an effective radius $R_{\rm eff}$ derived from the surface area, the total number N_* of MYSST catalog stars within the structure, the total surface stellar number density $n_{\rm surf}^{total}$, the number of identified most likely PMS stars $N_{\rm PMS}$ inside the contour, the corresponding surface number density of PMS sources $n_{\rm surf}^{pms}$, the number of enclosed McLeod et al. (2019) O stars $N_{\rm O}^{M19}$, SIMBAD O stars (see Table A1) $N_{\rm O}^{SD}$, Chen et al. (2009) YSOs $N_{\rm YSO}^{C09}$, Carlson et al. (2012) YSOs $N_{\rm YSO}^{C12}$ (the number in parentheses indicates matches in $N_{\rm YSO}^{C09}$), and the number of substructures $N_{\rm sub}^{2\sigma}$, $N_{\rm sub}^{3\sigma}$ at a density significance of 2 and 3σ, respectively. The value in parentheses in the $N_{\rm sub}^{3\sigma}$ column indicates the number of subclusters at 3σ with at least 50 stars, corresponding to the white contours in Figure 12. Lastly, we also provide the Cartwright & Whitworth (2004) Q parameter and its uncertainty as an indicator of cluster "clumpiness."

not have reliable estimates of the physical masses of the PMS candidate stars at this moment[1]. We also compute the surface area $A_{\rm surf}$ encompassed by the corresponding density contour, an effective radius derived as $R_{\rm eff} = \sqrt{A_{\rm surf}/\pi}$, the total number of MYSST stars N_* inside the structure, as well as the number of most likely PMS candidate stars $N_{\rm PMS}$, and corresponding surface number densities for total $n_{\rm surf}^{total}$ and PMS candidates $n_{\rm surf}^{PMS}$. Additionally, we count the enclosed O stars and YSOs from McLeod et al. (2019), the SIMBAD database (Table A1), Chen et al. (2009) and Carlson et al. (2012). For the prominent 1σ structures we determine the number of substructures at 2 and 3σ significance in density, $N_{\rm sub}^{2\sigma}$ and $N_{\rm sub}^{3\sigma}$, based on the dendrogram decomposition

(Rosolowsky et al. 2008) of the spatial distribution of the PMS candidate stars. Furthermore, we also compute the subclustering parameter Q as defined by Cartwright & Whitworth (2004) and its uncertainty σ_Q (more details on the Q-parameter follow at the end of this section).

A summary of these properties can be found in Table 2 for the 1σ structures and in Table A2 in the Appendix for their 3σ subclusters. Note that the IDs of the 3σ substructures in Table A2 indicate the 1σ structures that they belong to, i.e. N1.1 is inside N1, N2.1 in N2, etc. (see also Figure A8 in the Appendix for indicators of their spatial position). Additionally, Figures A6 and A7 in the Appendix provide dendrograms of the NNDE density structures in the main FOV up to the 5σ significance level for a more in-depth visualization of the hierarchical clustering structure that we encounter here.

We find that the PMS clusters in N44 cover a wide range of mass and size, with clearly the most prominent

---

[1] This situation will be improved with the application of more advanced ML techniques (Ksoll et al. 2020b, INN) in a future study.

structure being the one denoted as N4. With a surface area of more than $1300\,\mathrm{pc}^2$ and effective radius of over 20 pc, it is a very large structure of PMS candidates that traces the western ridge of N44's superbubble and extends into the bubble itself. It stretches across two of the H II regions, namely, 'N44 main' and 'N44C', and contains almost 6400 candidate PMS stars. Given the size of this structure, it is unlikely to be a single massive cluster (we are, after all, working only on a 2D projection).

The radial velocities of the 16 O stars located within this contour, as measured in McLeod et al. (2019), do not exhibit a noticeable trend in comparison to the remaining O stars. It appears that this cluster formed "in situ" in a region of higher gas density as the shell of the expanding H II bubble expands into the ambient medium. Additionally, N4 encloses a total of 18 YSOs, ranging from 6.5 to 22.1 $M_\odot$ (Chen et al. 2009; Carlson et al. 2012), and contains ample amounts of substructure (see Figure A6). At the 3σ density significance level this structure still contains 10 subclusters with at least 50 constituents, one of which, the subcluster N4.5 (see Appendix, Figure A8), entails more than 1100 PMS candidates. With a PMS surface number density of about $11.1\,\mathrm{pc}^{-2}$, N4.5 is the most prominent star-forming center that we identify in N44. It also harbors three O stars (O5 III, O8 V, and O9.5 V; McLeod et al. 2019) and one 9.2 $M_\odot$ YSO (Chen et al. 2009). The second-largest 3σ structure, hosting about 400 PMS candidates, is N4.9, which is likely another active star-forming cluster given its PMS surface number density above $11.6\,\mathrm{pc}^{-2}$. It comprises one O5 V star (McLeod et al. 2019) and a massive 17.4 $M_\odot$ YSO (Chen et al. 2009) as well.

In the south we do not find any structures with 1σ density significance as large as N4. The most prominent ones are S4 and S2, hosting 1308 and 902 PMS candidate stars, respectively. Additionally, S4 entails five YSOs (4.8–15.6 $M_\odot$, Chen et al. 2009; Carlson et al. 2012) and the O9 II giant Sk-67 82a (see Table A1), whereas S2 hosts four YSOs (6.9–16.5 $M_\odot$, Chen et al. 2009; Carlson et al. 2012). Both clusters exhibit notable substructuring with two and four clusters at 3σ density significance (see also Figure A7). Overall, the southern PMS structures appear to be less dense in their PMS stellar content, as the PMS surface number density lies on average around $4.9\,\mathrm{pc}^{-2}$ in the subclusters at 3σ significance, which is only about half the average density of the corresponding structures in the northern part. This lower average surface density of PMS candidate sources could indicate less star-forming activity in the regions south of the main bubble, due to e.g. less available gas, resulting in fewer present PMS sources. Alternatively,

most of the potential PMS sources could be older than 15 Myr, which is the maximum age our classification approach is sensitive to. The most 'active' star-forming subclusters here are S4.1 and S9.1 (see Appendix, Figure A8) with 312 and 245 PMS candidates, respectively.

The one structure in the northern reference field, FN1, appears similar in spatial extent to S5 – S9, but exhibits a notably lower surface density of PMS candidate stars at only 1.7 pc^{-2}, which is closer to but still below the smallest 1σ structures S1 and S3 in the southern main FOV. FN1 also exhibits substructure, with two subclusters at 3σ, and hosts three YSOs (3.8–16 $M_\odot$; Chen et al. 2009; Carlson et al. 2012). FN1's two 3σ substructures share PMS candidate surface densities comparable with the 3σ subclusters in the southern main FOV, with values of 3.7 and 4.5 pc^{-2} for FN1.1 and FN1.2 (see Appendix, Figure A8), respectively. This structure appears as a valid cluster candidate along with those identified in the main FOV, although it is located in one of the reference fields, which were supposed to only capture the LMC field population.

The 1σ structure identified in the southern reference field is among the largest (in area), comparable to S2 and S4. Hosting only 116 PMS candidates, however, it has by far the lowest PMS candidate surface density with 0.4 pc^{-2}. This value is by factors of 12.6 and 6.8 smaller than the average PMS surface density of the 1σ structures in the northern and southern main FOV. Together with the very low PMS nearest neighbor density that defines this 1σ structure and the uncertainty of the extinction estimate for this field, we believe that it is unclear whether FS1 actually traces a star-forming center. There is, however, one YSO (6.2 $M_\odot$; Carlson et al. 2012) in FS1, providing some evidence for recent star formation in this structure.

Instead of using the number of substructures identified in the dendrogram analysis as an indication of the 'clumpiness' or hierarchical nature of the PMS clusters, we can also look at the Q parameter introduced by Cartwright & Whitworth (2004). It is defined as the ratio of the mean edge length $\bar{m}$ in a minimum spanning tree (Prim 1957) constructed from the cluster stars and the mean stellar separation $\bar{s}$, both normalized to the effective cluster radius R_{eff}. Values of $Q < 0.8$ are indicative of a high degree of substructure, whereas larger values of Q are found in clusters that have a well-defined power-law radial density profile (Cartwright & Whitworth 2004; Schmeja & Klessen 2006; Allison et al. 2009). For an application to the structure of young stars in other clusters, see e.g. Schmeja et al. (2009) and Gennaro et al. (2017). The numbers in Table 2 indeed indicate that the Q values are lowest in clusters with well-

defined subclusters ($N_{\mathrm{sub}}^{2\sigma}$ and $N_{\mathrm{sub}}^{3\sigma}$ above 1) with the one exception possibly being cluster S4. We note that all clusters identified in N44 have $Q < 0.8$ as expected for hierarchically structured or fractal systems. This is confirmed by visual inspection of Figure A9 which shows the spatial distribution of the PMS candidates in the 18 clusters N1 – N7 in the north, S1 – S9 in the south, and FN1 and FS1 in the reference fields, none of which exhibit a clear power-law density falloff.

6. SUMMARY

In this study we present the identification of the PMS stellar population of the star-forming complex N44 in the Large Magellanic Cloud based on the photometric catalog of the deep HST survey MYSST. For this purpose we apply an ML classification approach, which we have previously established (Ksoll et al. 2018), to distinguish the observed sources into the two classes 'PMS' and 'non-PMS' based on their photometry in the F555W and F814W filters, as well as an estimate of individual stellar extinction.

To apply our classification scheme to the observations of N44, we first construct a suitable training set by selecting a region of N44 that exhibits a high density in PMS sources (as determined by a kernel density estimate on a rough selection of PMS candidate stars). This region provides both a distinct PMS and LMS population, which we distinguish using a Gaussian mixture model approach described in Ksoll et al. (2018). As stars on the RGB can also contaminate the CMD region usually occupied by PMS stars, we extend our training set through the addition of RGB 'non-PMS' examples selected from a series of LMC field regions within the observed FOV. Our final training set consists of 17,942 stars of which 5512 are PMS examples.

In the following, we train an SVM and an RF classifier to distinguish the two classes 'PMS' and 'non-PMS' using the magnitudes in F555W and F814W, as well as the estimated stellar extinction as the feature space. To evaluate training success, we hold out a randomly selected subset (30% of the total training data) as a test set and compute a series of standard performance measures, i.e. the normal and balanced accuracy, the area under the receiver operating characteristic curve (ROC AUC), and the F_1 score. We find that both models achieve excellent results on both the training and test subsets with accuracies exceeding 96%, as well as ROC AUCs above 0.99 and F_1 scores beyond 0.94.

Classifying the remaining data of the MYSST survey, we determine that an average of the predicted probability for the 'PMS' class between the SVM and RF methods provides the most robust outcome. *With that we find*

40,509 potential PMS candidates satisfying $\bar{p}_{\mathrm{pms}} \geq 0.5$ and a most likely subset with $\bar{p}_{\mathrm{pms}} \geq 0.95$ consisting of 26,686 sources across N44. Adopting the latter criterion, a majority of 16,976 PMS candidate stars are identified in and around N44's massive superbubble, located in the northern half of the MYSST FOV, while only 9710 candidate PMS sources are found in the region south of the bubble.

We then perform a nearest neighbor density estimate (NNDE; Casertano & Hut 1985) on the set of most likely PMS candidates to characterize their spatial distribution and clustering structures. Comparing with previous studies of the H II regions of N44 (McLeod et al. 2019; Pellegrini et al. 2012), we confirm that the majority of the dominant groupings of PMS candidate stars revealed by our ML classification approach coincide with N44's known H II regions. Further comparison with MUSE observations (McLeod et al. 2019) of the most massive young O star population of N44's bubble reveals that, at least within the FOV overlap of the two studies, almost all of our PMS clusters harbor one or more of the young high-mass stars. We find a similar result comparing with the positions of massive YSOs identified in N44 (Carlson et al. 2012; Chen et al. 2009). Therefore, we conclude that our classification approach identifies PMS sources exactly where one would expect to find them, i.e. within N44's gas reservoirs and in the vicinity of its massive young population. *This supports the hypothesis that stars tend to form in clusters* (see also Lada & Lada 2003; Klessen et al. 1998; Bonnell et al. 1998).

Additionally, we perform a comparison of our spatial PMS candidate distribution with the Zivkov et al. (2018) study, which has previously established a lower limit of 1000 ± 38 for the number of PMS stars in N44 and derived a PMS surface density map for the region, based on the VMC survey. We find an overall decent agreement with their results, in particular, when we account for the completeness and resolution limits of the VMC survey, and conclude that our study provides an excellent extension of their results to much lower brightness and higher spatial resolution.

We also compare the spatial distribution of our PMS candidate stars to other tracers of star formation, i.e. images of CO, Hα and dust (at 8 μm, 70 μm, 160 μm and 350 μm) emission. Here we find that *most of the prominent structures of PMS candidates appear correlated with areas of enhanced gas and dust emission*, with the most prominent exception being the interior of N44's superbubble, where massive stellar feedback has cleared out most of the material.

To assess the prominent PMS structures across N44, we use the NNDE to identify dominant groupings as density contours at 1σ significance (above the mean estimated density), which entail at least 100 stars in total and have substructures that persist up to the 3σ level. Here we perform separate NNDEs for the northern and southern half of the main FOV, as well as the two reference fields, to account for the difference in number of PMS candidates between the four regions and be more sensitive to the local clustering structures. *This procedure reveals 18 dominant PMS structures at 1σ in total, seven located in the north, nine in the south, and one each in the two fields.* For all of these we derive several properties, i.e., the center coordinates, surface area, effective radius, numbers and surface number densities of total/PMS stars, as well as the Cartwright & Whitworth (2004) Q-parameter for cluster "clumpiness". In the north the most dominant structure we find is a very large grouping of more than 6500 PMS stars that stretches along the western edge of the superbubble and extends into the bubble itself. While this structure is too large in size to be a single PMS cluster, it appears as a common envelope connecting the numerous star-forming centers at 3σ significance that fall within it. In the south we find more but slightly smaller PMS groupings that appear overall less densely populated in terms of PMS sources, i.e. they exhibit PMS surface number densities that are on average only half as large as in the north. We suspect that this *hints at a reduced star-forming activity in the south compared to the north.* On top of that, the identified dominant PMS groupings in both the north and south exhibit ample hierarchical substructures.

Following the outcomes of this study, there are a few open questions, which we plan to address in a future investigation. First and foremost is the physical characterization of the identified PMS candidates by estimating their most fundamental properties, age, and mass. We aim to achieve this through further development of an invertible-neural-network-based regression approach that we have recently presented in a pilot study (Ksoll et al. 2020b) with very promising results on the test cases of Westerlund 2 and NGC6397. Establishing these physical properties of the PMS stars of N44 will allow us to quantify the star formation history of this complex and investigate whether there is, e.g. an age difference between the clustering structures we have identified in the northern and southern part of the MYSST FOV. Furthermore, we plan to reevaluate our clustering analysis with regard to the predicted physical properties of the PMS stars to establish a comprehensive picture of the spatial distribution of star formation in this star-forming complex.

REFERENCES

Allison, R. J., Goodwin, S. P., Parker, R. J., et al. 2009,
    MNRAS, 395, 1449,
    doi: 10.1111/j.1365-2966.2009.14508.x

Baron, D. 2019, arXiv e-prints, arXiv:1904.07248.
    https://arxiv.org/abs/1904.07248

Bohannan, B., & Walborn, N. R. 1989, PASP, 101, 520,
    doi: 10.1086/132463

Bonnell, I. A., Bate, M. R., & Zinnecker, H. 1998, MNRAS,
    298, 93, doi: 10.1046/j.1365-8711.1998.01590.x

Breiman, L. 2001, Machine Learning, 45, 5,
    doi: 10.1023/A:1010933404324

Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS,
    427, 127, doi: 10.1111/j.1365-2966.2012.21948.x

Brunet, J. P., Imbert, M., Martin, N., et al. 1975, A&AS,
    21, 109

Cannon, A. J., & Pickering, E. C. 1993, VizieR Online
    Data Catalog, III/135A

Carlson, L. R., Sewiło, M., Meixner, M., Romita, K. A., &
    Lawton, B. 2012, A&A, 542, A66,
    doi: 10.1051/0004-6361/201118627

Cartwright, A., & Whitworth, A. P. 2004, MNRAS, 348,
    589, doi: 10.1111/j.1365-2966.2004.07360.x

Casertano, S., & Hut, P. 1985, ApJ, 298, 80,
    doi: 10.1086/163589

Chabrier, G. 2003, PASP, 115, 763, doi: 10.1086/376392

Chen, C. H. R., Chu, Y.-H., Gruendl, R. A., Gordon, K. D.,
    & Heitsch, F. 2009, ApJ, 695, 511,
    doi: 10.1088/0004-637X/695/1/511

Chu, Y.-H., Mac Low, M.-M., Garcia-Segura, G., Wakker,
    B., & Kennicutt, Robert C., J. 1993, ApJ, 414, 213,
    doi: 10.1086/173069

Cignoni, M., Sabbi, E., van der Marel, R. P., et al. 2015,
    ApJ, 811, 76, doi: 10.1088/0004-637X/811/2/76

Cioni, M. R. L., Clementini, G., Girardi, L., et al. 2011,
    A&A, 527, A116, doi: 10.1051/0004-6361/201016137

Conti, P. S., Garmany, C. D., & Massey, P. 1986, AJ, 92,
    48, doi: 10.1086/114133

Cortes, C., & Vapnik, V. 1995, Machine Learning, 20, 273,
    doi: 10.1007/BF00994018

Da Rio, N., Gouliermis, D. A., & Gennaro, M. 2010, ApJ,
    723, 166, doi: 10.1088/0004-637X/723/1/166

Da Rio, N., Gouliermis, D. A., Rochau, B., et al. 2012,
    MNRAS, 422, 3356,
    doi: 10.1111/j.1365-2966.2012.20851.x

De Marchi, G., Panagia, N., & Romaniello, M. 2010, ApJ,
    715, 1, doi: 10.1088/0004-637X/715/1/1

De Marchi, G., Panagia, N., Sabbi, E., et al. 2016,
    MNRAS, 455, 4373, doi: 10.1093/mnras/stv2528

Elmegreen, B. G. 2011, in EAS Publications Series, Vol. 51,
    EAS Publications Series, ed. C. Charbonnel &
    T. Montmerle (EDP Sciences, France), 45–58,
    doi: 10.1051/eas/1151004

Fluke, C. J., & Jacobs, C. 2020, WIREs Data Mining and
    Knowledge Discovery, 10, e1349, doi: 10.1002/widm.1349

Gennaro, M., Goodwin, S. P., Parker, R. J., Allison, R. J.,
    & Brandner, W. 2017, MNRAS, 472, 1760,
    doi: 10.1093/mnras/stx2098

Gouliermis, D., Brandner, W., & Henning, T. 2006, ApJ,
    641, 838, doi: 10.1086/500500

Gouliermis, D. A. 2012, SSRv, 169, 1,
    doi: 10.1007/s11214-012-9868-2

Gouliermis, D. A., Schmeja, S., Dolphin, A. E., et al. 2012,
    ApJ, 748, 64, doi: 10.1088/0004-637X/748/1/64

Hartmann, L., Herczeg, G., & Calvet, N. 2016, ARA&A,
    54, 135, doi: 10.1146/annurev-astro-081915-023347

Henize, K. G. 1956, ApJS, 2, 315, doi: 10.1086/190025

Jaskot, A. E., Strickland, D. K., Oey, M. S., Chu, Y. H., &
    García-Segura, G. 2011, ApJ, 729, 28,
    doi: 10.1088/0004-637X/729/1/28

Kato, D., Nagashima, C., Nagayama, T., et al. 2007, PASJ,
    59, 615, doi: 10.1093/pasj/59.3.615

Kennicutt, R. C., J., & Hodge, P. W. 1986, ApJ, 306, 130,
    doi: 10.1086/164326

Klessen, R. S., Burkert, A., & Bate, M. R. 1998, ApJL,
    501, L205, doi: 10.1086/311471

Klessen, R. S., & Glover, S. C. O. 2016, Star Formation in
    Galaxy Evolution: Connecting Numerical Models to
    Reality, Saas-Fee Advanced Course, Volume 43. ISBN
    978-3-662-47889-9. Springer-Verlag Berlin Heidelberg,
    2016, p. 85, 43, 85, doi: 10.1007/978-3-662-47890-5_2

Kroupa, P. 2002, Science, 295, 82,
    doi: 10.1126/science.1067524

Ksoll, V. F., Gouliermis, D. A., Klessen, R. S., et al. 2018,
    MNRAS, 479, 2389, doi: 10.1093/mnras/sty1317

Ksoll, V. F., Gouliermis, D., Sabbi, E., et al. 2020a, arXiv
    e-prints, arXiv:2012.00521.
    https://arxiv.org/abs/2012.00521

Ksoll, V. F., Ardizzone, L., Klessen, R., et al. 2020b,
    MNRAS, 499, 5447, doi: 10.1093/mnras/staa2931

Lada, C. J., & Lada, E. A. 2003, ARA&A, 41, 57,
    doi: 10.1146/annurev.astro.41.011802.094844

Lasker, B. M., Doggett, J., McLean, B., et al. 1996, in
    Astronomical Society of the Pacific Conference Series,
    Vol. 101, Astronomical Data Analysis Software and
    Systems V, ed. G. H. Jacoby & J. Barnes, 88

Lee, H.-T., & Chen, W. P. 2007, ApJ, 657, 884,
    doi: 10.1086/510893

Lucke, P. B., & Hodge, P. W. 1970, AJ, 75, 171, doi: 10.1086/110959

McLeod, A. F., Dale, J. E., Evans, C. J., et al. 2019, MNRAS, 486, 5263, doi: 10.1093/mnras/sty2696

Meixner, M., Gordon, K. D., Indebetouw, R., et al. 2006, AJ, 132, 2268, doi: 10.1086/508185

Meixner, M., Panuzzo, P., Roman-Duval, J., et al. 2013, AJ, 146, 62, doi: 10.1088/0004-6256/146/3/62

Nota, A., Sirianni, M., Sabbi, E., et al. 2006, ApJL, 640, L29, doi: 10.1086/503301

Oey, M. S., & Massey, P. 1995, ApJ, 452, 210, doi: 10.1086/176292

Panagia, N., Gilmozzi, R., Macchetto, F., Adorf, H. M., & Kirshner, R. P. 1991, ApJL, 380, L23, doi: 10.1086/186164

Pellegrini, E. W., Oey, M. S., Winkler, P. F., et al. 2012, ApJ, 755, 40, doi: 10.1088/0004-637X/755/1/40

Platt, J. C. 1999, in ADVANCES IN LARGE MARGIN CLASSIFIERS (MIT Press), 61–74

Portegies Zwart, S. F., McMillan, S. L., & Gieles, M. 2010, ARA&A, 48, 431, doi: 10.1146/annurev-astro-081309-130834

Prim, R. C. 1957, The Bell System Technical Journal, 36, 1389, doi: 10.1002/j.1538-7305.1957.tb01515.x

Rosolowsky, E. W., Pineda, J. E., Kauffmann, J., & Goodman, A. A. 2008, ApJ, 679, 1338, doi: 10.1086/587685

Rousseau, J., Martin, N., Prévot, L., et al. 1978, A&AS, 31, 243

Sabbi, E., Sirianni, M., Nota, A., et al. 2007, AJ, 133, 44, doi: 10.1086/509257

Sabbi, E., Lennon, D. J., Anderson, J., et al. 2016, ApJS, 222, 11, doi: 10.3847/0067-0049/222/1/11

Sanduleak, N. 1970, Contributions from the Cerro Tololo Inter-American Observatory, 89

Schmeja, S., Gouliermis, D. A., & Klessen, R. S. 2009, ApJ, 694, 367, doi: 10.1088/0004-637X/694/1/367

Schmeja, S., & Klessen, R. S. 2006, A&A, 449, 151, doi: 10.1051/0004-6361:20054464

Schulz, N. S. 2012, The Formation and Early Evolution of Stars (Springer, Berlin), doi: 10.1007/978-3-642-23926-7

Smith Neubig, M. M., & Bruhweiler, F. C. 1999, AJ, 117, 2856, doi: 10.1086/300867

Stahler, S. W., & Palla, F. 2005, The Formation of Stars (Wiley-VCH, Weinheim)

Stasińska, G., Testor, G., & Heydari-Malayeri, M. 1986, A&A, 170, L4

Stephens, I. W., Gouliermis, D., Looney, L. W., et al. 2017, ApJ, 834, 94, doi: 10.3847/1538-4357/834/1/94

Will, J. M., Bomans, D. J., & Dieball, A. 1997, A&AS, 123, 455, doi: 10.1051/aas:1997169

Wong, T., Hughes, A., Ott, J., et al. 2011, ApJS, 197, 16, doi: 10.1088/0067-0049/197/2/16

Wong, T., Hughes, A., Tokuda, K., et al. 2017, ApJ, 850, 139, doi: 10.3847/1538-4357/aa9333

Zaritsky, D., Harris, J., & Thompson, I. 1997, AJ, 114, 1002, doi: 10.1086/118531

Zinnecker, H., & Yorke, H. W. 2007, ARA&A, 45, 481, doi: 10.1146/annurev.astro.44.051905.092549

Zivkov, V., Oliveira, J. M., Petr-Gotzens, M. G., et al. 2018, A&A, 620, A143, doi: 10.1051/0004-6361/201833951
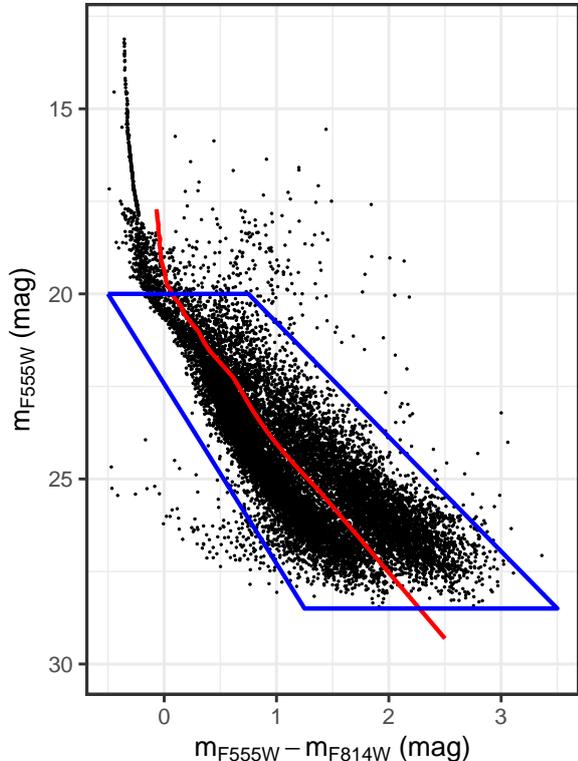
Figure A1. Optical CMD of the region selected as the training set base, corrected for extinction according to the A_{F555W} measurements. The red line indicates a threshold line, derived from PARSEC isochrones in the age range of 1–14 Myr. This threshold is used to generate slopes for our EM approach to quantify the constituents of the two populations. The blue polygon indicates the stars considered for the EM fit, excluding the UMS, red clump, and a few objects of unclear nature.

A. ADDITIONAL MATERIAL

This appendix provides complimentary figures and discussion to the main paper. Figure A1 shows the CMD of the 2σ surface density region used as a basis for our training set in Section 3. It highlights in particular the data selection we make for the Gaussian mixture model fit (see Ksoll et al. 2018) that we perform in order to distinguish the PMS and LMS populations, as well as the threshold curve that provides the basis for the underlying distance measure of this fit. The threshold curve corresponds to the 14 Myr PARSEC isochrone between 22.3 and 25.3 mag in F555W, extended by isochrones down to 1 Myr above 22.3 mag, and follows a combina-

tion of isochrones up to 50 Myr below 25.3 mag. Please note that this threshold serves not as a hard cut between the PMS and the LMS but as a guide for the Gaussian mixture model fit that determines the final population assignments. We refer to Ksoll et al. (2018) for more details on this mixture model approach.

In Figure A2 we show the individual prediction outcomes on the entire MYSST data set of the trained SVM and RF (left and middle panel) as the color-coding of the CMD, as well as a direct star-by-star comparison of the predicted PMS probabilities (right panel). As previously described in Section 4 these diagrams demonstrate how well the SVM and RF predictions agree overall and the few regions where they disagree, which lead us to the conclusion that a combination of the outcomes provides the most robust solution.

Complementary to Figure 6, which presents the spatial distribution of identified PMS stars from the combination of the two ML approaches, Figures A3 and A4 provide the corresponding distributions derived from SVM and RF individually, respectively.

Complementing the catalog from McLeod et al. (2019), Table A1 provides a list of other known O-type stars from the literature in and close to N44. For each star the table lists an identifier, coordinates in R.A. and decl., the spectral type and the literature reference of the studies that derive the latter.

Analogous to Figure 7 we show the spatial contour density diagram of our NNDE on the most probable candidate PMS stars in Figure A5, here in comparison to the positions of the 1291 UMS sources (light-blue points), likely late O to early B type, that we select in Paper I to derive extinction estimates for N44. As in Figure 7 this diagram demonstrates that the prominent PMS groupings we identify tend to be located in the vicinity of the massive young population of N44.

Figures A6 (north) and A7 (south) show the corresponding dendrograms of the clustering structures we have identified in Figure 12 in the main paper. These dendrograms are based on the NNDE we perform in the northern and southern half of the FOV and are iteratively constructed by considering each significance density contour as the root/parent structure of the contours/subclusters located inside of it. Both of these dendrograms highlight the intricate hierarchical substructure of the identified PMS groupings.

Analogous to Table 2 we present the characteristic properties of the 34 subclusters at 3σ surface density significance in Table A2. The properties include center posi-
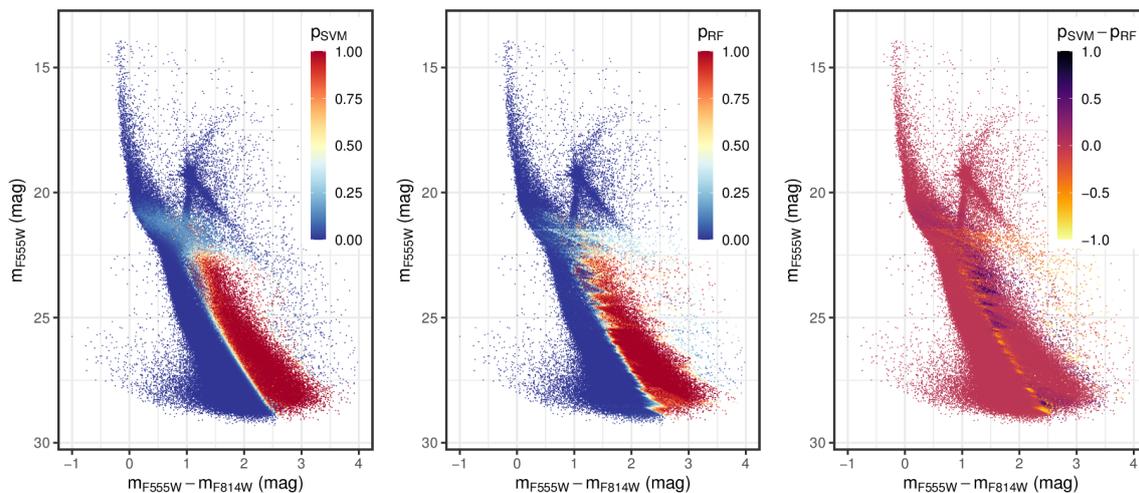
Figure A2. Optical CMD of the entire MYSST photometric catalog. Each star is color-coded according to the predicted probability that it belongs to the PMS as given by the SVM model (left), by the RF model (middle), and the difference $p_{\mathrm{svm}} - p_{\mathrm{rf}}$ of the predicted PMS candidate probabilities (right), respectively.
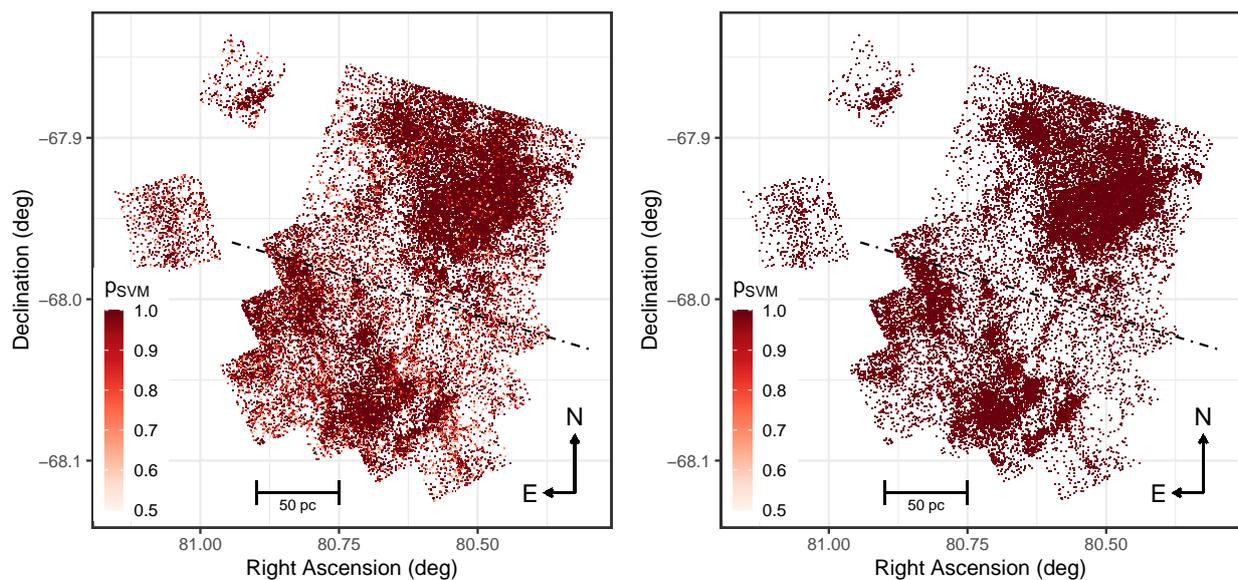
Figure A3. Spatial distribution diagrams of the candidate PMS stars as predicted by the SVM model. The left panel shows all candidate stars with $p_{\mathrm{SVM}} \geq 0.5$, while the right panel indicates the positions of the most likely PMS stars with $p_{\mathrm{SVM}} \geq 0.95$.

tion, surface area, effective radius, numbers, and number densities of the total/PMS stars of each 3σ substructure, as well as the Q-parameter (Cartwright & Whitworth 2004) as a measure of cluster clumpiness. Figure A8 indicates the positions of the 3σ subclusters analogous to Figure 12.

To complement the analysis of the cluster substructures in Section 5.3 Figure A9 provides the spatial distribution

diagrams of the PMS stars within the 18 prominent PMS structures presented in Table 2. These diagrams provide a visual confirmation of the Q-parameter (Cartwright & Whitworth 2004) analysis, indicating the overall clumpiness and hierarchical structure of all the prominent PMS clusterings we have identified.
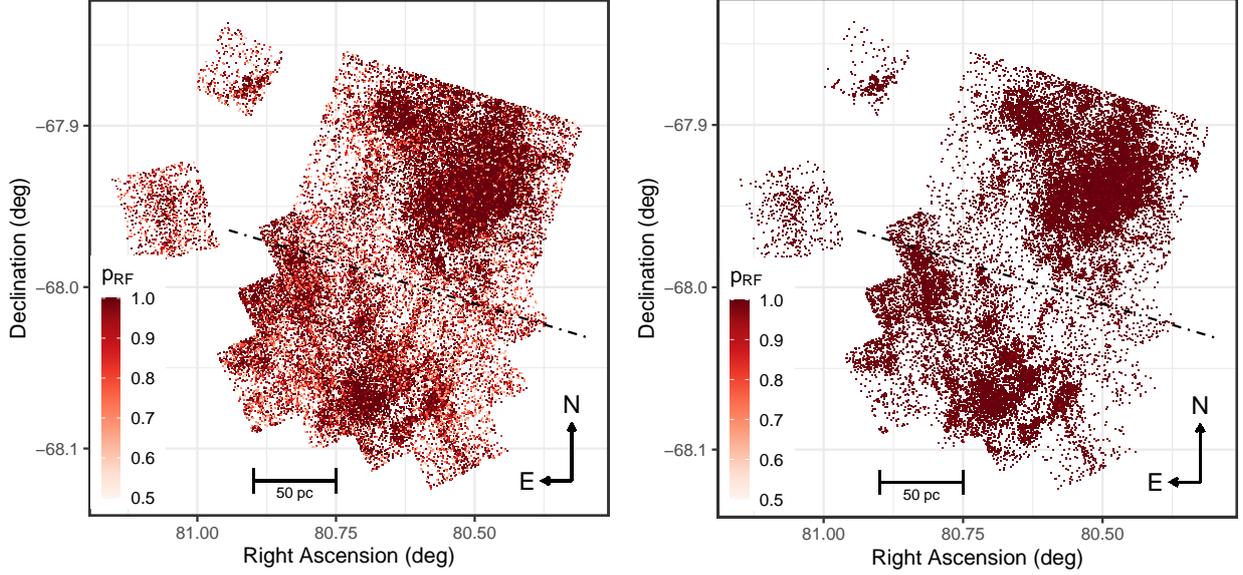
Figure A4. Spatial distribution diagrams of the candidate PMS stars as predicted by the RF model. In the left panel all stars with $p_{RF} \geq 0.5$ are shown, while the right panel marks the locations of the most likely candidate PMS stars with $p_{RF} \geq 0.95$.

Table A1. List of known O stars within or close to the MYSST FOV

Identifier	R.A.	Decl.	Spectral Type	Reference
	(deg)	(deg)		
SK -67 86	80.56181	-67.85908	OB	Sanduleak (1970)
HD 269412	80.47527	-67.91469	OB	Sanduleak (1970)
SK -67 94	80.88919	-67.95822	OB	Sanduleak (1970)
SK -68 76	81.01482	-68.06106	OB	Sanduleak (1970)
SK -68 72a	80.69472	-68.06542	O9II	Conti et al. (1986)
HD 269445	80.74911	-68.02962	Ofpe/WN9	Bohannan & Walborn (1989)
LH 47-355	80.56437	-67.98347	O9.5V	Oey & Massey (1995)
LH 47-335	80.55683	-67.93951	O9.5V	Oey & Massey (1995)
LH 47-14	80.43046	-67.9185	O9.5V	Oey & Massey (1995)
LH 48-122	80.59713	-67.88168	O9.5V	Oey & Massey (1995)
LH 47-84	80.46338	-67.93679	O9.5V	Oey & Massey (1995)
BI 155	80.95265	-67.89803	O7V	Smith Neubig & Bruhweiler (1999)
BI 159	81.04867	-68.0163	O/B0	Brunet et al. (1975)
[L72] LH 48-9	80.65	-67.9	O7III	Conti et al. (1986)
[L72] LH 48-21	80.6	-67.9	O5III	Conti et al. (1986)
HD 269449	80.8	-68.01667	O	Cannon & Pickering (1993)
[STH86] Star 2	80.61	-67.97	O	Stasińska et al. (1986)
SK -67 92	80.81166	-67.93655	OB	Sanduleak (1970)

NOTE— All O stars found in the SIMBAD database that are not captured by the McLeod et al. (2019) MUSE observations. Listed are each stars identifier, right ascension, declination, spectral type and the literature reference for the studies that derive the latter.
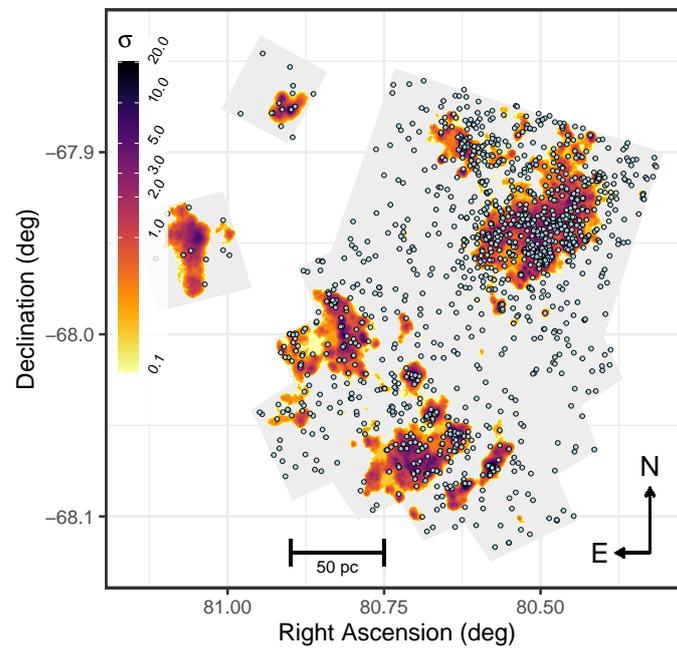
Figure A5. Spatial nearest neighbor contour density diagram of the most probable candidate PMS stars of the MYSST survey (as in Figure 7). The color-coding represents the nearest neighbor density in steps of σ above the mean density. The black dotted-dashed line indicates the north/south separation for the NNDE. The light blue points mark the positions of the 1,291 UMS stars selected in Paper I to derive extinction measures for the survey. For comparison, the gray shaded regions mark the MYSST coverage.

Table A2. Properties of the PMS Subclusters at 3σ Density

ID	R.A._cent	Decl._cent	A_{surf}	R_{eff}	N_*	$n_{\mathrm{surf}}^{total}$	N_{PMS}	n_{surf}^{pms}	N_{O}^{M19}	N_{YSO}^{C09}	N_{YSO}^{C12}	Q	σ_Q
	(deg)	(deg)	(pc^2)	(pc)		(pc^{-2})		(pc^{-2})					
N1.1	80.6590	-67.9046	2.6	0.9	50	19.3	28	10.8	0	0	0	0.88	0.61
N2.1	80.6236	-67.8946	10.3	1.8	210	20.5	82	8	0	1	1	0.73	0.48
N2.2	80.6116	-67.9024	4.3	1.2	94	22.1	36	8.5	0	0	0	0.8	0.53
N3.1	80.6247	-67.9110	2.8	0.9	57	20.5	25	9	0	0	0	0.76	0.53
N4.1	80.5843	-67.9474	3.3	1	64	19.4	29	8.8	0	0	0	0.64	0.46
N4.2	80.5852	-67.9450	2.6	0.9	62	23.7	17	6.5	0	0	0	0.73	0.47
N4.3	80.5250	-67.9372	8.2	1.6	163	19.8	55	6.7	0	0	0	0.61	0.45
N4.4	80.5271	-67.9684	2.9	1	55	19	29	10	0	0	0	0.77	0.49
N4.5	80.4973	-67.9561	100	5.6	2147	21.5	1109	11.1	3	1	0	0.69	0.48
N4.6	80.4933	-67.9384	5.7	1.4	120	20.9	53	9.2	0	0	0	0.74	0.47
N4.7	80.4788	-67.9380	15.3	2.2	311	20.3	124	8.1	0	0	0	0.7	0.44
N4.8	80.4682	-67.9352	4.5	1.2	88	19.6	30	6.7	0	0	0	0.75	0.47
N4.9	80.4532	-67.9506	32.8	3.2	750	22.9	381	11.6	1	1	0	0.75	0.49
N4.10	80.4534	-67.9271	6.2	1.4	143	23.1	52	8.4	0	0	0	0.68	0.5
N6.1	80.4203	-67.8912	4.9	1.3	92	18.6	45	9.1	0	0	0	0.8	0.5
N7.1	80.4082	-67.9140	7.9	1.6	200	25.3	90	11.4	1	0	0	0.76	0.5
S2.1	80.8289	-67.9815	6.6	1.5	101	15.3	27	4.1	0	0	0	0.81	0.5
S2.2	80.8161	-67.9974	31.3	3.2	441	14.1	151	4.8	0	0	0	0.6	0.44
S2.3	80.8113	-68.0080	6.9	1.5	102	14.8	27	3.9	0	0	0	0.74	0.49
S2.4	80.7842	-68.0023	9.9	1.8	139	14.1	47	4.8	0	1	0	0.74	0.5
S4.1	80.7049	-68.0678	73.5	4.8	1075	14.6	312	4.2	0	0	0	0.51	0.36
S4.2	80.6758	-68.0716	31.1	3.1	544	17.5	127	4.1	0	0	0	0.62	0.42
S5.1	80.7065	-68.0223	25	2.8	409	16.4	138	5.5	0	1	1	0.74	0.48
S6.1	80.6747	-68.0440	12.8	2	236	18.5	43	3.4	0	0	0	0.69	0.49
S7.1	80.6359	-68.0550	40	3.6	677	16.9	214	5.3	0	1	0	0.75	0.48
S8.1	80.6185	-68.0834	13.6	2.1	257	18.9	87	6.4	0	0	0	0.8	0.5
S9.1	80.5774	-68.0728	31.8	3.2	592	18.6	245	7.7	0	0	0	0.79	0.51
FN1.1	80.9122	-67.8704	3.8	1.1	55	14.5	14	3.7	0	0	1	0.75	0.50
FN1.2	80.8991	-67.8768	14.1	2.1	214	15.2	64	4.5	0	1	1	0.77	0.51
FS1.1	81.0500	-67.9466	41.7	3.6	491	11.8	32	0.8	0	0	0	0.72	0.48

NOTE— Properties of the PMS subclusters at 3σ density within the prominent 1σ density structures. This list only contains subclusters which entail at least 50 stars, so it corresponds to the solid light-blue contours depicted in Figure 12. Each object's ID indicates the 1σ structure it belongs to, e.g. N1.1 is within N1, N2.1 in N2, etc. As in Table 2 listed are the right ascension R.A._cent and declination Decl._cent of the subcluster center, the surface area A_{surf} enclosed by the given density contour, an effective radius R_{eff} derived from the surface area, the total number N_* of MYSST catalog stars within the structure, the total surface stellar number density $n_{\mathrm{surf}}^{total}$, the number of identified most likely PMS candidates N_{PMS} inside the contour, the corresponding surface number density of PMS sources n_{surf}^{pms}, the number of enclosed McLeod et al. (2019) O stars N_{O}^{M19}, Chen et al. (2009) YSOs N_{YSO}^{C09}, Carlson et al. (2012) YSOs N_{YSO}^{C12} and the Cartwright & Whitworth (2004) Q parameter along with its uncertainty σ_Q as an indicator of cluster "clumpiness". Note that none of the SIMBAD O stars (Table A1) fall into any of the σ contours, so they are not listed here.
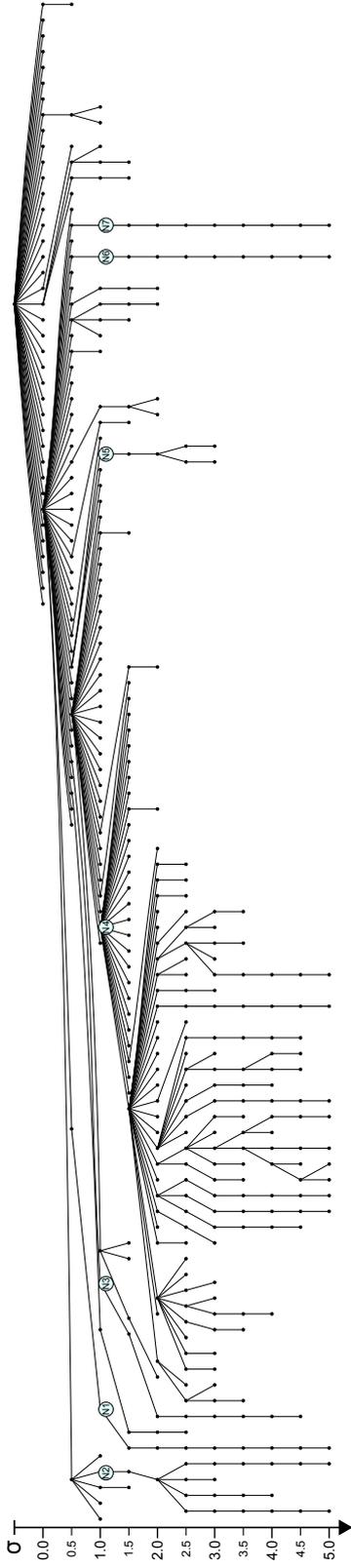
Figure A6. Dendrogram of the NNDE density structures of the most likely PMS candidates located in the northern half of the FOV. The large labeled circles indicate the 1σ density structures with at least 100 stars that persist up to 3σ density significance. The labels correspond to the IDs of the regions marked in Figure 12 and summarized in Table 2.
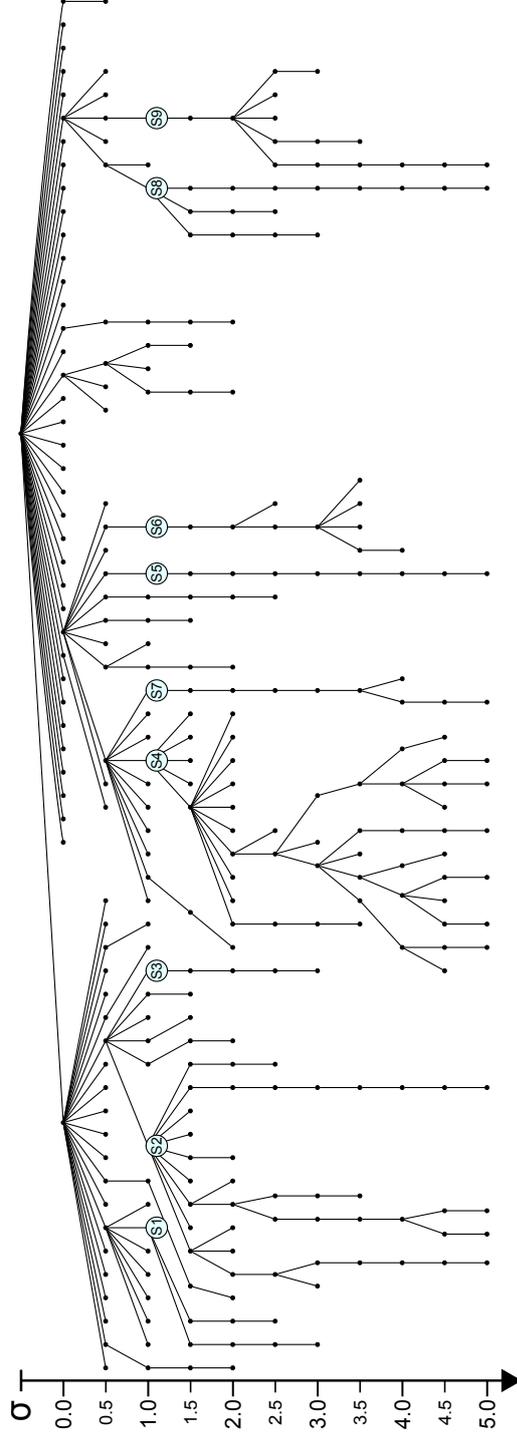
Figure A7. Dendrogram of the NNDE density structures of the most likely PMS candidates located in the southern half of the FOV. The large labeled circles indicate the 1σ density structures with at least 100 stars that persist up to 3σ density significance. The labels correspond to the IDs of the regions marked in Figure 12 and summarized in Table 2.
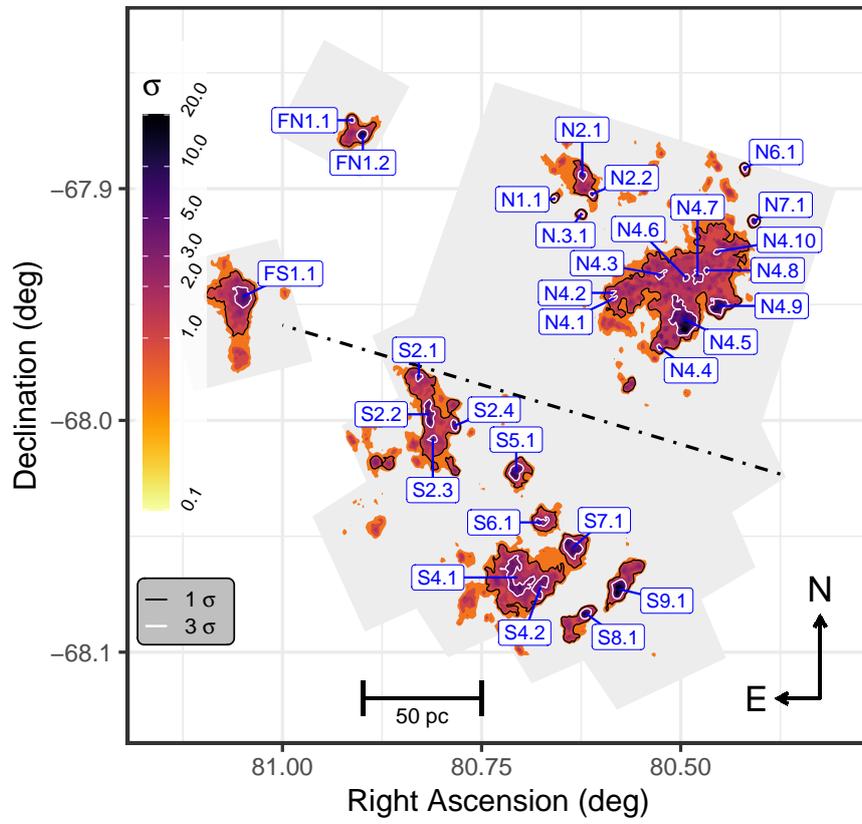
Figure A8. Same as Figure 12, but now the ID labels for the 3σ structures (see Table A2) are provided instead of the 1σ ones.
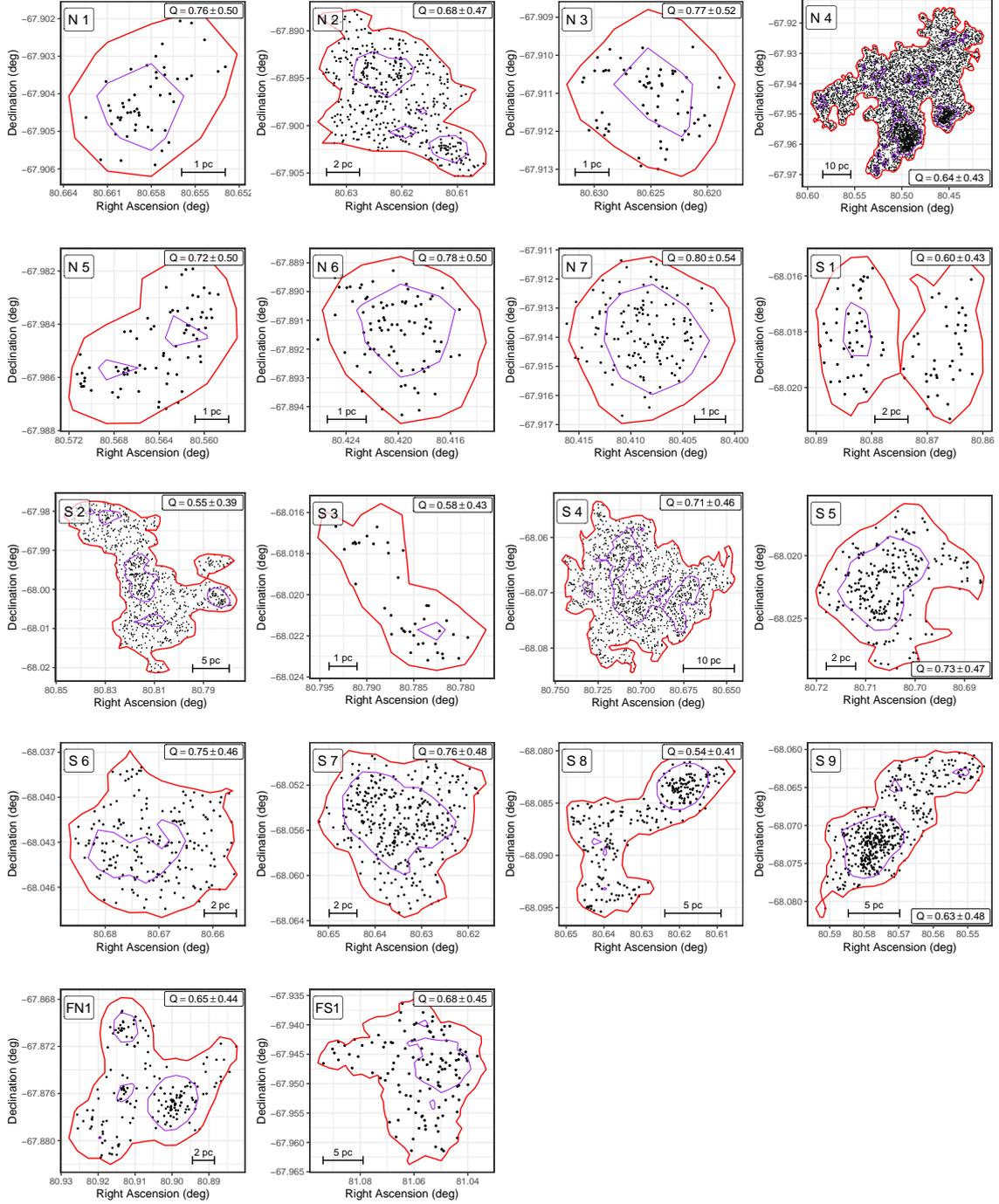
Figure A9. Spatial distribution diagrams of the PMS candidates in the 18 prominent PMS structures identified across the main MYSST FOV and the two reference fields. The identifiers correspond to the list presented in Table 2. In each diagram the red line indicates the 1σ contour associated with each cluster, while the purple lines mark the substructures at the 3σ density significance level. Note that in a few of these diagrams a small number of stars may fall outside the 1σ density contours of their assigned cluster. This is caused by minor inaccuracies in the transformation of the density contours from pixel space, in which they are defined, to the R.A.-decl. coordinate system presented in this diagram. In pixel space all stars are strictly interior to their respective cluster contours.

4

Summary, Discussion and Outlook

4.1 Summary

In this thesis we have set out to *explore* and test the effectiveness of machine and deep learning approaches for the purposes of *identifying* and *characterising* young pre-main-sequence stars from photometric observations, as well as to study star formation in the Large Magellanic Cloud. Towards these goals we have completed the work summarised below.

Identifying PMS stars

In Ksoll et al. (2018), presented in Section 3.1, we establish an ML classification procedure to tackle the disentanglement of PMS stars from more evolved populations in large star-forming complexes based on HST photometry, in this case from the "Hubble Tarantula Treasury Project" of the Tarantula Nebula. This publication is based on my master thesis, but comprises a refinement and completion of the original analysis, which was conducted as part of the dissertation. It is included in this thesis mainly for completeness and as a means to introduce the ML procedure that is later applied and further tested on the MYSST data. This approach consists of a two-class classification scheme, distinguishing between "PMS" and "NonPMS" stars, and is set up as follows. In a preparatory step we construct an extinction map of the observed region using massive upper main-sequence (UMS) stars as reddening probes under the assumption that their true CMD position should be the ZAMS. Subsequently, we estimate individual extinction for the remaining objects in the HTTP stellar catalogue, assigning a distance-weighted

average reddening value of the 20 nearest UMS neighbours to every star. We then identify tentative regions within the observed FoV that exhibit high densities of PMS sources using 2D kernel density estimation on the total photometric catalogue (or a rough CMD cut of PMS sources, which we do in Section 3.4 instead). This step follows from the assumption that PMS objects are most likely found within regions of high stellar density as stars preferably form in clustered environments. Among these regions we then select one that exhibits both a prominent PMS and contaminating LMS population in the low-brightness regime as the basis for the training set of the ML classification procedure. To label the training data in the low-brightness regime, i.e. quantify which stars belong to the LMS and PMS, we fit a Gaussian mixture model to the distribution of distances from the apparent gap between the two populations in the CMD using the EM algorithm. Afterwards, we further augment this training set with additional examples of RGB sources, extracted from regions devoid of PMS stars, to account for the potential overlap of pre- and post-main-sequence populations in the CMD.

We then train and test different classical ML classification techniques, i.e. decision tree, random forest and support vector machine classifiers (see Section 1.4), and determine the best performing approach. Here we find the RF and SVM models to be the most robust. We also investigate which combination of input observations provides the best compromise between algorithm performance and overall coverage of the HTTP FoV, determining the set of $\{m_{\mathrm{F555W}}, m_{\mathrm{F775W}}, A_{\mathrm{F555W}}\}$ to be the most optimal. Finally, we apply the trained classifiers to the entire HTTP photometric catalogue and conclude that an average of the RF and SVM predictions provides the most robust PMS classification outcome. With that we identify a total of 19,831 stars with a PMS classification probability above 95% across the entire Tarantula Nebula. Lastly, we verify that the densest clusterings of the PMS candidate stars correspond to the known star-forming centres NGC 2070 with R136 at its core and NGC 2060.

In Ksoll et al. (2021a), Section 3.3, we introduce the "Measuring Young Stars in Space and Time" (MYSST) survey, which targets the star-forming complex LHα 120–N44 in the Large Magellanic Cloud in two HST broad-band filters, F555W and F814W. We present the survey's observing strategy and data

reduction procedure, and construct the photometric catalogue, returning a rich stellar sample of 461,684 sources with a detection limit of $\sim 0.09\,M_\odot$ for unreddened PMS stars at an age of 1 Myr. We also perform a first qualitative analysis of the stellar populations captured by the survey, identifying numerous old LMS and RGB sources that are almost uniformly distributed across the MYSST FoV – likely field constituents of the LMC – as well as UMS and PMS stars exhibiting a clustered distribution, in particular around N44's main superbubble. Additionally, we analyse extinction properties of N44, introducing the RANSAC algorithm to constrain the slope of the reddening vector from the extinction-elongated RC feature (cf. Section 1.3) in the CMD. We demonstrate that RANSAC is a robust tool to jointly determine the constituents of the RC and measure its slope, which we find to be slightly steeper than the one of the diffuse Milky Way ISM. Afterwards, we construct extinction maps for the observed region, one following the UMS-based method as in Ksoll et al. (2018) and one using the RC stars as extinction probes. Lastly, we quantify uncertainties of our UMS-based extinction estimation approach with regards to the assumed true UMS position in the CMD, stellar rotation effects and potential metallicity gradients.

In our subsequent MYSST study Ksoll et al. (2021b), Section 3.4, we apply and further test the ML classification approach prototyped in my master thesis and Ksoll et al. (2018) on the new MYSST photometric catalogue and recover a census of the young PMS stellar population of N44. In total we identify 26,686 candidate PMS stars across the entire star-forming complex. Subsequently, we analyse the spatial distribution and clustering behaviour of the identified PMS candidates employing a nearest neighbour density estimation (NNDE) approach. Based on this NNDE we devise a contour-based clustering procedure, in which we identify density contours at the 1σ significance level (above the mean nearest neighbour density) that persist up to 3σ in density as prominent clusterings of PMS stars. Doing so we find a total of 18 PMS clusters across N44. Additionally, we verify our ML classification outcome by comparison of the PMS spatial distribution and clustering structure with other tracers of star formation, including $H\alpha$, CO and dust emission maps, the location of known H II regions, as well as the distribution of previously identified massive YSOs and O stars. Here we find that the identified PMS candidates and clusters are located primarily in the

vicinity of young massive stars within N44's H II regions, and, for the most part, close to regions of enhanced emission in all considered tracers. Lastly, we compare our PMS density distribution with a previously derived PMS density map by Zivkov et al. (2018) and find a satisfying agreement with our results (when accounting for resolution and completeness differences between ours and their survey). This provides an independent confirmation of the validity of our classification approach from a different methodology for the identification of PMS regions.

Characterising PMS stars

In Ksoll et al. (2020), Section 3.2, we develop a deep learning approach for predicting stellar physical parameters (on a star-by-star basis) from photometric observations alone, using a conditional invertible neural network (see Section 1.5.2). For this pilot implementation of the cINN approach we make the following simplifying assumptions: a) the metallicity of the target population is fixed and known, b) individual extinction measures exist for each query star, and c) we can neglect measurement uncertainties and physical effects such as stellar variability or unresolved binarity. We begin by constructing a suitable training set for the cINN approach from the PARSEC stellar evolution models. Here we augment the available isochrone tables (including synthetic HST photometry) via (per-isochrone) spline interpolation in order to evenly sample the age and initial mass parameter space for the training set. Afterwards, we add further examples of each synthetic observation at different levels of extinction. We then train cINN models to predict stellar age, initial/current mass, luminosity, temperature and surface gravity from photometric magnitudes in several filter bands and the measure of individual extinction. Here, the number and combination of filters depends on the target survey to be analysed. In this study we perform a real data benchmark on HST observations of the well-studied, supposedly single-age Milky Way clusters Westerlund 2 (2 filters) and NGC 6397 (5 filters).

We proceed by training and testing the cINN on the synthetic PARSEC data. Here, we find that the cINN does an overall excellent job in recovering the target physical parameters, although stellar age appears as the hardest

parameter to constrain, showing on average the broadest predicted posterior distributions. Additionally, the cINN manages to correctly highlight degeneracies in the inverse problem, e.g. the inherent overlap between post- and pre-main-sequence evolution in the HRD. To further test how well the trained cINN models generalise to new, unseen data (i.e. beyond the held-out PARSEC test set) we perform experiments on synthetic data from the MIST and Dartmouth evolutionary models, finding overall a satisfactory performance. Lastly, we apply the trained cINN to the observational data of Westerlund 2 and NGC 6397. On Westerlund 2 we successfully recover the cluster's age of about 1 Myr, and find results on both the IMF slope and mass segregation that are in good agreement with previous studies. On the NGC 6397 data on the other hand, the cINN performance is not as satisfactory. Although the predictions for certain properties of the cluster stars are within reasonable ranges, the age (~ 13.5 Gyr) is heavily underestimated for almost all query sources. We identify notable discrepancies between the underlying PARSEC isochrone models and the observed photometry as the main culprit for these severe mispredictions. With the success on the Westerlund 2 data, we conclude that the cINN is an efficient and powerful tool to solve the degenerate task of stellar parameter prediction from photometry, but requires a careful selection of the underlying synthetic model data.

4.2 Discussion

Over the four studies presented in this thesis we have demonstrated the successful application of a variety of ML approaches, ranging from classification methods, such as RFs and SVMs, over density estimation and fitting techniques (RANSAC, EM-algorithm) to deep learning regression with invertible neural networks.

Our first main accomplishment is an effective PMS identification procedure for surveys of large star-forming regions, which is trained on (a subset of) the query survey data itself and, thus, does not require complex modelling. Although initially prototyped in my master thesis, we further refined this approach in Ksoll et al. (2018) and extensively tested the procedure in our

subsequent application on new data, Ksoll et al. (2021b), by comparing with other tracers of star formation and a completely different PMS analysis approach, proving the method's overall robustness in this dissertation. We have to acknowledge three primary limitations of our methodology, though. First, as individual stellar extinction serves as an input, our method is, to a degree, dependent on the precision of the reddening measurement. However, in our applications so far we find that the set of predicted PMS candidates does not significantly change when extinction is not considered as an input. Second, our approach may only recover PMS sources that are similar to the chosen examples in the training set. In particular, in both our studies our method is not sensitive to low-mass PMS stars that are older than $\gtrsim 14$ Myr, i.e. those very close to joining the ZAMS. Lastly, our method is specifically designed for surveys of large star-forming complexes, i.e. it requires the survey data to contain at least one sufficiently numerous subpopulation of PMS stars that can serve as the training data.

The second central accomplishment is the successful application of our PMS identification procedure to a different large photometric survey, the MYSST project. Combined with our previous results on the HTTP data, we present a census of more than $46,000$ prospective PMS stars (with $\bar{p}_{\mathrm{pms}} \geq 0.95$) for two of the largest and most active star-forming regions in the LMC. These two PMS candidate catalogues provide an excellent sample to further investigate and constrain the star-formation histories of these two star-forming complexes through analysis of their spatial distribution and characterisation of their physical properties. For N44 we already lay the foundation for this spatial analysis and determine the positions, stellar number densities and structure for 18 prominent PMS clusters.

The third main outcome of this thesis is the successful proof-of-concept of a cINN approach for stellar parameter prediction. We demonstrate that the cINN is a very efficient tool, being able to predict full posterior distributions for the physical parameters for thousands of stars in a matter of minutes once trained, and that it is well-suited to recover and highlight the degeneracies of the problem. In addition, we show that our approach can correctly recover physical parameters on real observations of PMS populations with our

Westerlund 2 application. We conclude that the cINN demonstrates great potential as an evaluation tool for large (photometric) surveys.

The proof-of-concept implementation of our method is, however, subject to a few limiting factors. First are the obvious constraints imposed by the simplifying assumptions in the pilot study, i.e. that metallicity and extinction are an input requirement, and that measurement uncertainties are not taken into account. Second, there is no effective way yet to incorporate prior knowledge into model the itself. In our study we experiment with age cuts in the training sets to account for the known ages of the test clusters, but do not find them to have a strong influence on the predictions for e.g. Westerlund 2. Third is the central limitation that the cINN predictions can only ever be as good as the underlying physical models it is trained on. As the NGC 6397 experiment shows, a notable mismatch between observations and models can severely impact the predictive capability of our model. Besides the points already raised in our discussion in Section 3.2 regarding the suboptimal fit of the PARSEC models to globular cluster CMDs, we have recently learned about a potential property of NGC 6397 that provides additional insights into our prediction issues for the cluster. Whereas the HUGS HST photometry of NGC 6397 led us to believe that the cluster consists of a single-population, Correnti et al. (2018) and Mészáros et al. (2020) find some evidence that NGC 6397 is host to (at least) two populations with slightly differing chemical composition. If that is indeed the case, this is another source for discrepancy between our underlying models, which assume a single metallicity, and the observations. Lastly, the cINN is subject to the general limitation of most neural network architectures, namely, being unable to make predictions if one or more observational features are missing. More specifically, our approach has to be retrained for every new combination of photometric filters, tailored to the coverage of the survey it is to be applied on.

Because of these limiting factors of our prototype cINN, we have not yet made the connection between our PMS identification and characterisation procedures. The Tarantula Nebula and N44 are much more complex environments than Westerlund 2 or NGC 6397, so that further development of the cINN approach to e.g. lift the limitations set by our simplifying assumptions

seem warranted before an application to the MYSST and HTTP survey data is reasonable.

Regarding the central goals of this thesis (see Section 1.0.1), i.e. *exploring* ML approaches to *identify* and *characterise* PMS stars, and studying star-forming environments and their formation histories in the LMC, we have accomplished most but not all we have set out to do. We have *explored* a wide variety of ML approaches and demonstrated their successful application to different astronomical problems. We have extensively tested an ML classification procedure to *identify* PMS stars based on photometry and have used it to recover candidate PMS populations for both of our main-target star-forming complexes. Lastly, we have provided the proof-of-concept of an efficient cINN-based tool for *characterising* stars by predicting their physical properties. We have, however, not yet managed to combine the two aspects of our analysis and actually characterise the identified PMS stars within our main targets, N44 and the Tarantula Nebula. Consequently, we are yet to achieve our goal of further constraining the star formation histories of these two complexes. Even though we have not realised the full set of our goals within this thesis and doctoral project, we provide a considerable foundation for subsequent studies of star formation in the LMC and further ML-related methodological advances in astronomy.

4.3 Outlook

In this final section we shall outline some of the possible directions in which our results can be taken in.

The most obvious route is to finally combine our two approaches for identification and characterisation of PMS sources for the observational data of HTTP and MYSST. Recovering the physical parameters of the identified PMS populations and combining the results with the spatial distribution analysis that we have already conducted for e.g. N44 will allow to spatially map the star formation history across each of the two regions, recover their IMFs

and provide further insights into the cluster structures (e.g. mass segregation). Although the cINN approach (once fully developed, see below) may potentially supersede our ML-classification-based identification procedure – an algorithm that can directly predict stellar physical parameters essentially classifies objects at the same time – we believe that a combination of both will provide the most robust analysis path. Here the identification step, which is purely based on the observation data itself, will serve as an important, independent verification mechanism for the characterisation algorithm that is entirely based on synthetic models.

The PMS identification we have presented for N44 can also potentially be improved further. As we discuss in Section 4.1 our PMS classification outcome is, to a minor degree, dependent on the input extinction estimate. Likewise we have detailed in Section 3.3 that our UMS-based extinction estimate is subject to some uncertainty due to the assumptions we make on the true CMD position of the UMS probes. Consequently, recovering a more precise estimate for the reddening of the UMS population of N44 will not only improve the extinction maps of the region, but may also allow for some refinement of the PMS identification outcome. However, we have to emphasise again that the classification outcomes in our tests on HTTP and MYSST do not appear to be overly sensitive to extinction as an input. A precision improvement of the extinction estimate may be achieved through the exploration of synergies of the MYSST survey with other programs that have covered N44, such as e.g. Gaia or HAWK-I. Additional photometric information beyond the two filters of MYSST for the very massive, young UMS stars will allow for a more precise determination of their true CMD position and subsequently extinction. A cINN that predicts extinction rather than taking it as an input (see discussion further below) may even be used for this purpose. As stated in Section 3.3 this refinement of the MYSST extinction map is a planned follow-up to the study presented in this thesis. Note that the HTTP analysis can, of course, be refined in a similar manner.

In this thesis we show that the cINN is a very efficient approach to predict full posterior distributions for stellar parameters from photometry and particularly well suited to capture and highlight the inherent degeneracies of this task. Beyond these two properties the cINN is also a very *versatile* tool and may

solve other inverse problems in astronomy as well. In a soon-to-be-published collaborative study, Haldemann et al. (in prep.), with colleagues from Bern, we already successfully adapt our cINN approach to predict physical parameters of exoplanets from observations with the recently launched CHEOPS (Characterising Exoplanet Satellite) satellite. There is also another team at the Institute for Theoretical Astrophysics in Heidelberg that is currently developing a cINN approach for the prediction of the bulk physical properties of star-forming clouds. These different application examples show that the cINN has the potential to become a general analysis tool not only for photometric surveys but across many disciplines in astronomy.

Both of the potential follow-ups to our studies suggested so far include the continued development of the cINN approach for stellar parameter prediction. The central next steps in this development consist of addressing the limitations of the proof-of-concept implementation discussed in Section 4.1. First is an extension of the predictive capabilities of our approach to metallicity and extinction. Both of these properties introduce additional degeneracy in this inverse problem and are not always available a priori for the analysis of observational data. Although we do have our fairly effective UMS-based extinction estimation approach for a future cINN application to the data of HTTP and MYSST, this method may likely over- or underestimate reddening for field sources as they are not spatially colocated with the UMS probes (see discussions in Sections 3.1 and 3.3). Extending the cINN to jointly predict extinction and stellar parameters would, thus, constitute a major improvement. In Section 3.2 we already show in a first test on NGC 6397 that the prediction of extinction is well within the realm of possibility. Using the photometry in the five HUGS survey filters, we demonstrate that even an unmodified cINN can already get close ($\bar{A}_V^{\mathrm{MAP}} = 0.417 \pm 0.468$ mag) to recovering the previously measured average extinction ($\bar{A}_V^{\mathrm{Brown}} = 0.574 \pm 0.006$ mag; Brown et al., 2018) of NGC 6397 (c.f. Figure 25 in Section 3.2). Even though we are not able to achieve the same from only two photometric filters for Westerlund 2 in Section 3.2 – Westerlund 2 is also a much more extensively reddened region – this preliminary result indicates that (at least) robust extinction predictions will likely be possible with our approach once the cINN architecture has been adapted to account for the increased complexity of this extension of the inverse problem.

Secondly, there is the matter of addressing measurement uncertainties in the input observations. For this extension, too, we already test the most straightforward approach as part of our collaboration on the Haldemann et al. (in prep.) study. This method consists of resampling a given observation within its measurement uncertainties, predicting parameter posterior distributions for each of these samples and subsequently combining the predictions to derive the full posterior of the uncertain observation. In Haldemann et al. (in prep.) we show that this procedure successfully recovers the correct posterior distribution by comparison with a standard MCMC approach. Although this experiment shows that this auxiliary procedure can correctly account for uncertainty, it may also be desirable to develop a cINN intrinsic mechanism. This could potentially be achieved by augmentation of the training data with additional examples generated according to the expected measurement error of a given observational instrument.

Next is the open question of incorporating prior knowledge efficiently into the cINN approach. As we have discussed before, the straightforward idea to modify the training data according to prior information does not prove particularly effective in our tests. A potential way to improve this may lie in the latent variables of the cINN or more specifically their priors. As described in Section 1.5.2, these priors are set to normal Gaussian distributions in the standard cINN, primarily for simplicity. A possible modification with regards to prior information is, therefore, to match the distribution of one of the latent variables to a known prior of one of the target physical parameters. Although we cannot say for certain if this will incorporate prior knowledge more effectively at this point, it appears as a promising experiment for a follow-up investigation. With regards to the application on observational data of stellar clusters, another path worth investigating is related to spatial correlations. The latter constitute another form of constraint on some of the prediction target parameters. Neighbouring stars in a cluster, for instance, should have similar extinction in most cases. To address this matter one could for instance image a post-processing procedure that breaks degeneracies, i.e. multi-modalities, in the predicted posterior distributions following a spatial consistency criterion.

With the future improvements proposed so far and its excellent efficiency, our cINN approach has the potential to become a comprehensive and powerful general tool for the analysis of large (photometric) surveys. This is not just limited to observations with the HST, but may also include e.g. Pan-STARRS or the wealth of data expected from upcoming new observational facilities such as the Vera Rubin Observatory, the Roman Space Telescope or the James Webb Space Telescope. Towards such a general-purpose analysis tool for the astronomical community the missing-feature limitation needs to be addressed. As this restriction is inherent to the method, it has to be mitigated through the development of a mechanism that automatically generates the required synthetic training data and trains the cINN model for new query surveys. This can be supplemented by the compilation of a suite of fine-tuned, pre-trained cINN models for common observational features (e.g. broad-band filter combinations) from popular observing facilities (e.g. HST) or surveys. Along this path it may also be worthwhile to investigate an own implementation of the underlying stellar evolution models, using e.g. the publicly available Modules for Experiments in Stellar Astrophysics (MESA; Paxton et al., 2011, 2013, 2015, 2018, 2019), to gain better control over the physics involved in the generated training data. Specifically, this would also allow for a direct investigation of the resimulation error of the cINN, which so far we have to approximate in Section 3.2 via a nearest neighbour search owing to the PARSEC evolutionary model code not being openly available.

In summary, with the work we have presented in this thesis we lay the foundation for multiple worthwhile follow-up investigations. These include the continued study and quantification of the star formation histories of both the Tarantula Nebula and N44 with the ML methods we have introduced, as well as the development of an efficient, powerful cINN-based tool for the analysis of large surveys for the astronomical community.

Acknowledgements

Additionally, many thanks go to Simon Heinze and Philipp Austermühl, who have taken the time to proofread this manuscript and have greatly helped me with bringing this thesis to completion.

Finally, I want to thank my parents and family for their immense support through good and bad times, and for allowing me to pursue this path. It is without a doubt that this thesis would not exist if not for you. Thank you for everything.

Bibliography

B. J. Anthony-Twarog, B. A. Twarog, and N. B. Suntzeff. CCD Stromgren Studies in NGC 6397. *Astronomical Journal*, 103:1264, Apr 1992. doi: 10.1086/116140.

L. Ardizzone, J. Kruse, S. Wirkert, D. Rahner, E. W. Pellegrini, R. S. Klessen, L. Maier-Hein, C. Rother, and U. Köthe. Analyzing Inverse Problems with Invertible Neural Networks. *arXiv e-prints*, art. arXiv:1808.04730, Aug. 2018.

L. Ardizzone, C. Lüth, J. Kruse, C. Rother, and U. Köthe. Guided Image Generation with Conditional Invertible Neural Networks. *arXiv e-prints*, art. arXiv:1907.02392, July 2019.

D. J. Armstrong et. al. Automatic vetting of planet candidates from ground-based surveys: machine learning with NGTS. *Monthly Notices of the Royal Astronomical Society*, 478: 4225–4237, 2018. doi: 10.1093/mnras/sty1313.

J. Ascenso, J. Alves, Y. Beletsky, and M. T. V. T. Lago. Near-IR imaging of Galactic massive clusters: Westerlund 2. *Astronomy and Astrophysics*, 466(1):137–149, Apr 2007. doi: 10.1051/0004-6361:20066433.

J. Ballesteros-Paredes, L. Hartmann, and E. Vázquez-Semadeni. Turbulent Flow-driven Molecular Cloud Formation: A Solution to the Post-T Tauri Problem? *Astrophysical Journal*, 527(1):285–297, Dec. 1999. doi: 10.1086/308076.

D. Baron. Machine Learning in Astronomy: a practical overview. *arXiv e-prints*, art. arXiv:1904.07248, Apr. 2019.

M. Beech. *Introducing the stars*. Springer eBooks : Physics and Astronomy. Springer, Cham, 2019. ISBN 978-3-030-11704-7. doi: 10.1007/978-3-030-11704-7. URL `https://doi.org/10.1007/978-3-030-11704-7`.

E. P. Bellinger, G. C. Angelou, S. Hekker, S. Basu, W. H. Ball, and E. Guggenberger. Fundamental Parameters of Main-Sequence Stars in an Instant with Machine Learning. *Astrophysical Journal*, 830(1):31, Oct. 2016. doi: 10.3847/0004-637X/830/1/31.

G. Besla, N. Kallivayalil, L. Hernquist, R. P. van der Marel, T. J. Cox, and D. Kereš. The role of dwarf galaxy interactions in shaping the Magellanic System and implications for Magellanic Irregulars. *Monthly Notices of the Royal Astronomical Society*, 421(3): 2109–2138, Apr. 2012. doi: 10.1111/j.1365-2966.2012.20466.x.

M. S. Bessell. Standard Photometric Systems. *Annual Review of Astronomy and Astrophysics*, 43(1):293–336, Sept. 2005. doi: 10.1146/annurev.astro.41.082801.100251.

J. M. Bestenlehner, J. S. Vink, G. Gräfener, F. Najarro, C. J. Evans, N. Bastian, A. Z. Bonanos, E. Bressert, P. A. Crowther, E. Doran, K. Friedrich, V. Hénault-Brunet, A. Herrero, A. de Koter, N. Langer, D. J. Lennon, J. Maíz Apellániz, H. Sana, I. Soszynski, and W. D. Taylor. The VLT-FLAMES Tarantula Survey. III. A very massive star in apparent isolation from the massive cluster R136. *Astronomy and Astrophysics*, 530:L14, June 2011. doi: 10.1051/0004-6361/201117043.

F. Bigiel, A. Leroy, F. Walter, E. Brinks, W. J. G. de Blok, B. Madore, and M. D. Thornley. The Star Formation Law in Nearby Galaxies on Sub-Kpc Scales. *Astronomical Journal*, 136(6):2846–2871, Dec. 2008. doi: 10.1088/0004-6256/136/6/2846.

C. M. Bishop. *Pattern recognition and machine learning*. Information science and statistics. Springer, New York [u.a.], 8. (corr. printing) edition, 2009. ISBN 0-387-31073-8 and 978-0-387-31073-2. Hier auch später erschienene, unveränderte Nachdrucke.

P. H. Bodenheimer. *Principles of Star Formation*. Astronomy and astrophysics library. Springer, Berlin ; Heidelberg [u.a.], 2011. ISBN 978-3-642-15062-3. doi: 10.1007/978-3-642-15063-0. Literaturverz. S. 325 - 338.

W. Brandner, E. K. Grebel, R. H. Barbá, N. R. Walborn, and A. Moneti. Hubble Space Telescope NICMOS Detection of a Partially Embedded, Intermediate-Mass, Pre-Main-Sequence Population in the 30 Doradus Nebula. *Astronomical Journal*, 122(2):858–865, Aug. 2001. doi: 10.1086/322065.

L. Breiman. Random Forests. *Machine Learning*, 45:5–32, Jan. 2001. doi: 10.1023/A:1010933404324.

A. Bressan, P. Marigo, L. Girardi, B. Salasnich, C. Dal Cero, S. Rubele, and A. Nanni. PARSEC: stellar tracks and isochrones with the PAdova and TRieste Stellar Evolution Code. *Monthly Notices of the Royal Astronomical Society*, 427(1):127–145, Nov. 2012. doi: 10.1111/j.1365-2966.2012.21948.x.

T. M. Brown, S. Casertano, J. Strader, A. Riess, D. A. VandenBerg, D. R. Soderblom, J. Kalirai, and R. Salinas. A High-precision Trigonometric Parallax to an Ancient Metal-poor Globular Cluster. *Astrophysical Journal, Letters*, 856(1):L6, Mar. 2018. doi: 10.3847/2041-8213/aab55a.

T. Cantat-Gaudin, F. Anders, A. Castro-Ginard, C. Jordi, M. Romero-Gómez, C. Soubiran, L. Casamiquela, Y. Tarricq, A. Moitinho, A. Vallenari, A. Bragaglia, A. Krone-Martins, and M. Kounkel. Painting a portrait of the Galactic disc with its stellar clusters. *Astronomy and Astrophysics*, 640:A1, Aug. 2020. doi: 10.1051/0004-6361/202038192.

J. A. Cardelli, G. C. Clayton, and J. S. Mathis. The Relationship between Infrared, Optical, and Ultraviolet Extinction. *Astrophysical Journal*, 345:245, Oct. 1989. doi: 10.1086/167900.

L. R. Carlson, M. Sewiło, M. Meixner, K. A. Romita, and B. Lawton. Identifying young stellar objects in nine Large Magellanic Cloud star-forming regions. *Astronomy and Astrophysics*, 542:A66, June 2012. doi: 10.1051/0004-6361/201118627.

C. M. Casey, D. Narayanan, and A. Cooray. Dusty star-forming galaxies at high redshift. *Physics Reports*, 541(2):45–161, Aug. 2014. doi: 10.1016/j.physrep.2014.02.009.

G. Chabrier. Galactic Stellar and Substellar Initial Mass Function. *Publications of the ASP*, 115(809):763–795, July 2003. doi: 10.1086/376392.

C. H. R. Chen, Y.-H. Chu, R. A. Gruendl, K. D. Gordon, and F. Heitsch. Spitzer View of Young Massive Stars in the Large Magellanic Cloud H II Complex N 44. *Astrophysical Journal*, 695(1):511–541, Apr. 2009. doi: 10.1088/0004-637X/695/1/511.

Y. Chen, Q. D. Wang, E. V. Gotthelf, B. Jiang, Y.-H. Chu, and R. Gruendl. Chandra ACIS Spectroscopy of N157B: A Young Composite Supernova Remnant in a Superbubble. *Astrophysical Journal*, 651(1):237–249, Nov. 2006. doi: 10.1086/507017.

Y. Chen, L. Girardi, A. Bressan, P. Marigo, M. Barbieri, and X. Kong. Improving PARSEC models for very low mass stars. *Monthly Notices of the Royal Astronomical Society*, 444 (3):2525–2543, Nov 2014. doi: 10.1093/mnras/stu1605.

Y. Chen, A. Bressan, L. Girardi, P. Marigo, X. Kong, and A. Lanza. PARSEC evolutionary tracks of massive stars up to 350 $M_\odot$ at metallicities $0.0001 \leq Z \leq 0.04$. *Monthly Notices of the Royal Astronomical Society*, 452(1):1068–1080, Sep 2015. doi: 10.1093/mnras/stv1281.

M. Chevance, J. M. D. Kruijssen, A. P. S. Hygate, A. Schruba, S. N. Longmore, B. Groves, J. D. Henshaw, C. N. Herrera, A. Hughes, S. M. R. Jeffreson, P. Lang, A. K. Leroy, S. E. Meidt, J. Pety, A. Razza, E. Rosolowsky, E. Schinnerer, F. Bigiel, G. A. Blanc, E. Emsellem, C. M. Faesi, S. C. O. Glover, D. T. Haydon, I. T. Ho, K. Kreckel, J. C. Lee, D. Liu, M. Querejeta, T. Saito, J. Sun, A. Usero, and D. Utomo. The lifecycle of molecular clouds in nearby star-forming disc galaxies. *Monthly Notices of the Royal Astronomical Society*, 493(2):2872–2909, Apr. 2020a. doi: 10.1093/mnras/stz3525.

M. Chevance, J. M. D. Kruijssen, E. Vazquez-Semadeni, F. Nakamura, R. Klessen, J. Ballesteros-Paredes, S.-i. Inutsuka, A. Adamo, and P. Hennebelle. The Molecular Cloud Lifecycle. *Space Science Reviews*, 216(4):50, Apr. 2020b. doi: 10.1007/s11214-020-00674-x.

J. Choi, A. Dotter, C. Conroy, M. Cantiello, B. Paxton, and B. D. Johnson. Mesa Isochrones and Stellar Tracks (MIST). I. Solar-scaled Models. *Astrophysical Journal*, 823(2):102, June 2016. doi: 10.3847/0004-637X/823/2/102.

F. R. Chromey. *To Measure the Sky: An Introduction to Observational Astronomy*. Cambridge University Press, 2010. doi: 10.1017/CBO9780511794810.

Y.-H. Chu, M.-M. Mac Low, G. Garcia-Segura, B. Wakker, and J. Kennicutt, Robert C. Hidden Supernova Remnants in the Large Magellanic Cloud H II Complex N44. *Astrophysical Journal*, 414:213, Sept. 1993. doi: 10.1086/173069.

M. Cignoni, E. Sabbi, R. P. van der Marel, M. Tosi, D. Zaritsky, J. Anderson, D. J. Lennon, A. Aloisi, G. de Marchi, D. A. Gouliermis, E. K. Grebel, L. J. Smith, and P. Zeidler. Hubble Tarantula Treasury Project. II. The Star-formation History of the Starburst Region NGC 2070 in 30 Doradus. *Astrophysical Journal*, 811(2):76, Oct. 2015. doi: 10.1088/0004-637X/811/2/76.

M. Cignoni, E. Sabbi, R. P. van der Marel, D. J. Lennon, M. Tosi, E. K. Grebel, I. Gallagher, J. S., A. Aloisi, G. de Marchi, D. A. Gouliermis, S. Larsen, N. Panagia, and L. J. Smith. Hubble Tarantula Treasury Project V. The Star Cluster Hodge 301: The Old Face of 30 Doradus. *Astrophysical Journal*, 833(2):154, Dec. 2016. doi: $10.3847/1538\text{-}4357/833/2/154$.

M. R. L. Cioni, G. Clementini, L. Girardi, R. Guandalini, M. Gullieuszik, B. Miszalski, M. I. Moretti, V. Ripepi, S. Rubele, G. Bagheri, K. Bekki, N. Cross, W. J. G. de Blok, R. de Grijs, J. P. Emerson, C. J. Evans, B. Gibson, E. Gonzales-Solares, M. A. T. Groenewegen, M. Irwin, V. D. Ivanov, J. Lewis, M. Marconi, J. B. Marquette, C. Mastropietro, B. Moore, R. Napiwotzki, T. Naylor, J. M. Oliveira, M. Read, E. Sutorius, J. T. van Loon, M. I. Wilkinson, and P. R. Wood. The VMC survey. I. Strategy and first data. *Astronomy and Astrophysics*, 527:A116, Mar. 2011. doi: $10.1051/0004\text{-}6361/201016137$.

P. S. Conti, C. D. Garmany, and P. Massey. Studies of massive stars in the Magellanic Clouds. I. New spectral classifications of OB types in the LMC. *Astronomical Journal*, 92: 48–59, July 1986. doi: $10.1086/114133$.

M. Correnti, M. Gennaro, J. S. Kalirai, R. E. Cohen, and T. M. Brown. The Age of the Old Metal-poor Globular Cluster NGC 6397 Using WFC3/IR Photometry. *Astrophysical Journal*, 864(2):147, Sept. 2018. doi: $10.3847/1538\text{-}4357/aad805$.

C. Cortes and V. Vapnik. Support-vector networks. *Mach. Learn.*, 20(3):273–297, Sept. 1995. ISSN 0885-6125. doi: $10.1023/A{:}1022627411411$. URL https://doi.org/10.1023/A:1022627411411.

P. A. Crowther, O. Schnurr, R. Hirschi, N. Yusof, R. J. Parker, S. P. Goodwin, and H. A. Kassim. The R136 star cluster hosts several stars whose individual masses greatly exceed the accepted 150M$_{solar}$ stellar mass limit. *Monthly Notices of the Royal Astronomical Society*, 408(2):731–751, Oct. 2010. doi: $10.1111/j.1365\text{-}2966.2010.17167.x$.

P. A. Crowther, N. Castro, C. J. Evans, J. S. Vink, J. Melnick, and F. Selman. Dissecting the Core of the Tarantula Nebula with MUSE. *The Messenger*, 170:40–44, Dec. 2017. doi: $10.18727/0722\text{-}6691/5053$.

R. de Grijs, J. E. Wicker, and G. Bono. Clustering of Local Group Distances: Publication Bias or Correlated Measurements? I. The Large Magellanic Cloud. *Astronomical Journal*, 147(5):122, May 2014. doi: $10.1088/0004\text{-}6256/147/5/122$.

G. De Marchi and N. Panagia. The extinction law inside the 30 Doradus nebula. *Monthly Notices of the Royal Astronomical Society*, 445(1):93–106, Nov. 2014. doi: $10.1093/mnras/stu1694$.

G. De Marchi, F. Paresce, N. Panagia, G. Beccari, L. Spezzi, M. Sirianni, M. Andersen, M. Mutchler, B. Balick, M. A. Dopita, J. A. Frogel, B. C. Whitmore, H. Bond, D. Calzetti, C. M. Carollo, M. J. Disney, D. N. B. Hall, J. A. Holtzman, R. A. Kimble, P. J. McCarthy, R. W. O'Connell, A. Saha, J. I. Silk, J. T. Trauger, A. R. Walker, R. A. Windhorst, and E. T. Young. Star Formation in 30 Doradus. *Astrophysical Journal*, 739(1):27, Sept. 2011. doi: $10.1088/0004\text{-}637X/739/1/27$.

G. De Marchi, N. Panagia, E. Sabbi, D. Lennon, J. Anderson, R. van der Marel, M. Cignoni, E. K. Grebel, S. Larsen, D. Zaritsky, P. Zeidler, D. Gouliermis, and A. Aloisi. Hubble Tarantula Treasury Project - IV. The extinction law. *Monthly Notices of the Royal Astronomical Society*, 455(4):4373–4387, Feb. 2016. doi: 10.1093/mnras/stv2528.

G. Deco and W. Brauer. Nonlinear higher-order statistical decorrelation by volume-conserving neural architectures. *Neural Networks*, 8(4):525–535, 1995. ISSN 0893-6080. doi: https://doi.org/10.1016/0893-6080(94)00108-X. URL https://www.sciencedirect.com/science/article/pii/089360809400108X.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *J. Royal Stat. Soc., Series B*, 39(1):1–38, 1977.

L. Dinh, J. Sohl-Dickstein, and S. Bengio. Density estimation using Real NVP. *arXiv e-prints*, art. arXiv:1605.08803, May 2016.

C. L. Dobbs, M. R. Krumholz, J. Ballesteros-Paredes, A. D. Bolatto, Y. Fukui, M. Heyer, M. M. M. Low, E. C. Ostriker, and E. Vázquez-Semadeni. Formation of Molecular Clouds and Global Conditions for Star Formation. In H. Beuther, R. S. Klessen, C. P. Dullemond, and T. Henning, editors, *Protostars and Planets VI*, page 3, Jan. 2014. doi: 10.2458/azu_uapress_9780816531240-ch001.

A. Dotter. MESA Isochrones and Stellar Tracks (MIST) 0: Methods for the Construction of Stellar Isochrones. *Astrophysical Journal, Supplement*, 222(1):8, Jan. 2016. doi: 10.3847/0067-0049/222/1/8.

P. L. Dufton, P. R. Dunstall, C. J. Evans, I. Brott, M. Cantiello, A. de Koter, S. E. de Mink, M. Fraser, V. Hénault-Brunet, I. D. Howarth, N. Langer, D. J. Lennon, N. Markova, H. Sana, and W. D. Taylor. The VLT-FLAMES Tarantula Survey: The Fastest Rotating O-type Star and Shortest Period LMC Pulsar—Remnants of a Supernova Disrupted Binary? *Astrophysical Journal, Letters*, 743(1):L22, Dec. 2011. doi: 10.1088/2041-8205/743/1/L22.

G. B. Field and W. C. Saslaw. A Statistical Model of the Formation of Stars and Interstellar Clouds. *Astrophysical Journal*, 142:568, Aug. 1965. doi: 10.1086/148318.

M. Fischler and R. Bolles. *Random sample consensus*. ACM, 1981.

C. J. Fluke and C. Jacobs. Surveying the reach and maturity of machine learning and artificial intelligence in astronomy. *WIREs Data Mining and Knowledge Discovery*, 10(2):e1349, Jan. 2020. doi: 10.1002/widm.1349.

D. Forsyth and J. Ponce. *Computer vision*. Prentice Hall series in artificial intelligence. Prentice Hall, Upper Saddle River, NJ, 2003. ISBN 0-13-085198-1 and 978-0-13-085198-7. An Alan R. Apt Book.

Y. Fukui, N. Mizuno, R. Yamaguchi, A. Mizuno, and T. Onishi. On the Mass Spectrum of Giant Molecular Clouds in the Large Magellanic Cloud. *Publications of the ASJ*, 53(6):L41–L44, Dec. 2001. doi: 10.1093/pasj/53.6.L41.

Gaia Collaboration et al. The Gaia mission. *Astronomy and Astrophysics*, 595:A1, Nov. 2016. doi: 10.1051/0004-6361/201629272.

Gaia Collaboration et al. Gaia Data Release 2. Summary of the contents and survey properties. *Astronomy and Astrophysics*, 616:A1, Aug. 2018. doi: 10.1051/0004-6361/201833051.

Y. Gal and Z. Ghahramani. Bayesian Convolutional Neural Networks with Bernoulli Approximate Variational Inference. *arXiv e-prints*, art. arXiv:1506.02158, June 2015.

D. Gamerman and H. F. Lopes. *Markov chain Monte Carlo*. Number AR-RAY(0x5594b3a7f028) in Texts in statistical science series. Chapman & Hall/CRC, Boca Raton ; London ; New York, second edition edition, 2006. ISBN 1-58488-587-4 and 978-1-58488-587-0. Literaturverzeichnis: Seite [289]-310.

L. Girardi. Red Clump Stars. *Annual Review of Astronomy and Astrophysics*, 54:95–133, Sept. 2016. doi: 10.1146/annurev-astro-081915-023354.

P. Girichidis, C. Federrath, R. Banerjee, and R. S. Klessen. Importance of the initial conditions for star formation - II. Fragmentation-induced starvation and accretion shielding. *Monthly Notices of the Royal Astronomical Society*, 420(1):613–626, Feb. 2012. doi: 10.1111/j.1365-2966.2011.20073.x.

P. Girichidis, S. S. R. Offner, A. G. Kritsuk, R. S. Klessen, P. Hennebelle, J. M. D. Kruijssen, M. G. H. Krause, S. C. O. Glover, and M. Padovani. Physical Processes in Star Formation. *Space Science Reviews*, 216(4):68, June 2020. doi: 10.1007/s11214-020-00693-8.

G. C. Gómez and E. Vázquez-Semadeni. Filaments in Simulations of Molecular Cloud Formation. *Astrophysical Journal*, 791(2):124, Aug. 2014. doi: 10.1088/0004-637X/791/2/124.

I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

K. D. Gordon, G. C. Clayton, K. A. Misselt, A. U. Land olt, and M. J. Wolff. A Quantitative Comparison of the Small Magellanic Cloud, Large Magellanic Cloud, and Milky Way Ultraviolet to Near-Infrared Extinction Curves. *Astrophysical Journal*, 594(1):279–293, Sept. 2003. doi: 10.1086/376774.

K. D. Gordon, M. Fouesneau, H. Arab, K. Tchernyshyov, D. R. Weisz, J. J. Dalcanton, B. F. Williams, E. F. Bell, L. Bianchi, M. Boyer, Y. Choi, A. Dolphin, L. Girardi, D. W. Hogg, J. S. Kalirai, M. Kapala, A. R. Lewis, H.-W. Rix, K. Sandstrom, and E. D. Skillman. The Panchromatic Hubble Andromeda Treasury. XV. The BEAST: Bayesian Extinction and Stellar Tool. *Astrophysical Journal*, 826(2):104, Aug. 2016. doi: 10.3847/0004-637X/826/2/104.

D. A. Gouliermis. Low-Mass Pre-Main-Sequence Stars in the Magellanic Clouds. *Space Science Reviews*, 169(1-4):1–25, Sept. 2012. doi: 10.1007/s11214-012-9868-2.

D. A. Gouliermis, T. Henning, W. Brandner, A. E. Dolphin, M. Rosa, and B. Brandl. Discovery of the Pre-Main-Sequence Population of the Stellar Association LH 95 in the Large Magellanic Cloud with Hubble Space Telescope Advanced Camera for Surveys Observations. *Astrophysical Journal, Letters*, 665(1):L27–L30, Aug. 2007. doi: 10.1086/521224.

R. G. Gratton, A. Bragaglia, E. Carretta, G. Clementini, S. Desidera, F. Grundahl, and S. Lucatello. Distances and ages of NGC 6397, NGC 6752 and 47 Tuc. *Astronomy and Astrophysics*, 408:529–543, Sep 2003. doi: 10.1051/0004-6361:20031003.

A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(25):723–773, 2012. URL http://jmlr.org/papers/v13/gretton12a.html.

R. Hainich, U. Rühling, H. Todt, L. M. Oskinova, A. Liermann, G. Gräfener, C. Foellmi, O. Schnurr, and W. R. Hamann. The Wolf-Rayet stars in the Large Magellanic Cloud. A comprehensive analysis of the WN class. *Astronomy and Astrophysics*, 565:A27, May 2014. doi: 10.1051/0004-6361/201322696.

J. Haldemann, V. F. Ksoll, D. Walter, Y. Alibert, R. S. Klessen, and W. Benz. Planet characterization using invertible neural networks. in prep.

L. Hartmann, J. Ballesteros-Paredes, and E. A. Bergin. Rapid Formation of Molecular Clouds and Stars in the Solar Neighborhood. *Astrophysical Journal*, 562(2):852–868, Dec. 2001. doi: 10.1086/323863.

T. Hastie, R. Tibshirani, and J. H. Friedman. *The elements of statistical learning*. Springer series in statistics. Springer, New York, NY, 2. ed. edition, 2009. ISBN 978-0-387-84857-0.

T. M. Heckman, G. Kauffmann, J. Brinchmann, S. Charlot, C. Tremonti, and S. D. M. White. Present-Day Growth of Black Holes and Bulges: The Sloan Digital Sky Survey Perspective. *Astrophysical Journal*, 613(1):109–118, Sept. 2004. doi: 10.1086/422872.

J. S. Heiner, E. Vázquez-Semadeni, and J. Ballesteros-Paredes. Molecular cloud formation as seen in synthetic H I and molecular gas observations. *Monthly Notices of the Royal Astronomical Society*, 452(2):1353–1374, Sept. 2015. doi: 10.1093/mnras/stv1153.

F. Heitsch and L. Hartmann. Rapid Molecular Cloud and Star Formation: Mechanisms and Movies. *Astrophysical Journal*, 689(1):290–301, Dec. 2008. doi: 10.1086/592491.

K. G. Henize. Catalogues of Hα-emission Stars and Nebulae in the Magellanic Clouds. *Astrophysical Journal, Supplement*, 2:315, Sept. 1956. doi: 10.1086/190025.

P. Hennebelle and E. Falgarone. Turbulent molecular clouds. *Astronomy and Astrophysics Reviews*, 20:55, Nov. 2012. doi: 10.1007/s00159-012-0055-y.

P. Hennebelle and M. Pérault. Dynamical condensation in a thermally bistable flow. Application to interstellar cirrus. *Astronomy and Astrophysics*, 351:309–322, Nov. 1999.

V. Hill, S. Andrievsky, and M. Spite. Chemical evolution of the Magellanic Clouds. VI. Chemical composition of nine F supergiants from different regions of the large Magellanic Cloud. *Astronomy and Astrophysics*, 293:347–359, Jan. 1995.

S. Hony, F. Galliano, S. C. Madden, P. Panuzzo, M. Meixner, C. Engelbracht, K. Misselt, M. Galametz, M. Sauvage, J. Roman-Duval, K. Gordon, B. Lawton, J. P. Bernard, A. Bolatto, K. Okumura, C. H. R. Chen, R. Indebetouw, F. P. Israel, E. Kwon, A. Li, F. Kemper, M. S. Oey, M. Rubio, and H. E. Triou. The Herschel revolution: Unveiling the morphology of the high-mass star-formation sites N44 and N63 in the LMC. *Astronomy and Astrophysics*, 518:L76, July 2010. doi: 10.1051/0004-6361/201014628.

C.-W. Huang, D. Krueger, A. Lacoste, and A. Courville. Neural Autoregressive Flows. *arXiv e-prints*, art. arXiv:1804.00779, Apr. 2018.

M. Huertas-Company et al. A Catalog of Visual-like Morphologies in the 5 CANDELS Fields Using Deep Learning. *Astrophysical Journal, Supplement*, 221:8, 2015. doi: 10.1088/0067-0049/221/1/8.

L. K. Hunt and H. Hirashita. The size-density relation of extragalactic H II regions. *Astronomy and Astrophysics*, 507(3):1327–1343, Dec. 2009. doi: 10.1051/0004-6361/200912020.

A. Hyvärinen and P. Pajunen. Nonlinear independent component analysis: Existence and uniqueness results. *Neural Networks*, 12(3):429–439, 1999. ISSN 0893-6080. doi: https://doi.org/10.1016/S0893-6080(98)00140-3. URL https://www.sciencedirect.com/science/article/pii/S0893608098001403.

P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *CVPR*, 2017.

G. James, D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning*. Springer texts in statistics. Springer, New York ; Heidelberg ; Dordrecht ; London, corrected at 8th printing edition, 2017. ISBN 978-1-4614-7137-0 and 1-4614-7137-0. doi: 10.1007/978-1-4614-7138-7. ©Springer ... 2013 (Corrected at 8th printing 2017).

A. E. Jaskot, D. K. Strickland, M. S. Oey, Y. H. Chu, and G. García-Segura. Observational Constraints on Superbubble X-ray Energy Budgets. *Astrophysical Journal*, 729(1):28, Mar. 2011. doi: 10.1088/0004-637X/729/1/28.

A. Javadi, J. T. van Loon, H. G. Khosroshahi, F. Tabatabaei, R. Hamedani Golshan, and M. Rashidi. The UK Infrared Telescope M 33 monitoring project - V. The star formation history across the galactic disc. *Monthly Notices of the Royal Astronomical Society*, 464 (2):2103–2119, Jan. 2017. doi: 10.1093/mnras/stw2463.

P. M. W. Kalberla and J. Kerp. The Hi Distribution of the Milky Way. *Annual Review of Astronomy and Astrophysics*, 47(1):27–61, Sept. 2009. doi: 10.1146/annurev-astro-082708-101823.

J. Kennicutt, R. C. and P. W. Hodge. H II Regions and Star Formation in the Magellanic Clouds. *Astrophysical Journal*, 306:130, July 1986. doi: 10.1086/164326.

R. C. Kennicutt and N. J. Evans. Star Formation in the Milky Way and Nearby Galaxies. *Annual Review of Astronomy and Astrophysics*, 50:531–608, Sept. 2012. doi: 10.1146/annurev-astro-081811-125610.

D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. *arXiv e-prints*, art. arXiv:1412.6980, Dec. 2014.

D. P. Kingma and P. Dhariwal. Glow: Generative Flow with Invertible 1x1 Convolutions. *arXiv e-prints*, art. arXiv:1807.03039, Jul 2018.

D. P. Kingma, T. Salimans, and M. Welling. Variational Dropout and the Local Reparameterization Trick. *arXiv e-prints*, art. arXiv:1506.02557, June 2015.

D. P. Kingma, T. Salimans, R. Jozefowicz, X. Chen, I. Sutskever, and M. Welling. Improved variational inference with inverse autoregressive flow. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper/2016/file/ddeebdeefdb7e7e7a697e1c3e3d8ef54-Paper.pdf.

R. Kippenhahn, A. Weigert, and A. Weiss. *Stellar structure and evolution*. Astronomy and astrophysics library. Springer, Berlin ; Heidelberg, second edition edition, 2012. ISBN 3-642-30255-6 and 978-3-642-44524-8 and 978-3-642-30255-8. doi: 10.1007/978-3-642-30304-3. Hier auch später erschienene, unveränderte Nachdrucke.

R. S. Klessen and S. C. O. Glover. Physical Processes in the Interstellar Medium. *Saas-Fee Advanced Course*, 43:85, Jan. 2016. doi: 10.1007/978-3-662-47890-5_2.

R. S. Klessen and P. Hennebelle. Accretion-driven turbulence as universal process: galaxies, molecular clouds, and protostellar disks. *Astronomy and Astrophysics*, 520:A17, Sept. 2010. doi: 10.1051/0004-6361/200913780.

R. S. Klessen, F. Heitsch, and M.-M. Mac Low. Gravitational Collapse in Turbulent Molecular Clouds. I. Gasdynamical Turbulence. *Astrophysical Journal*, 535(2):887–906, June 2000. doi: 10.1086/308891.

M. Kounkel, K. Covey, and K. G. Stassun. Untangling the Galaxy. II. Structure within 3 kpc. *Astronomical Journal*, 160(6):279, Dec. 2020. doi: 10.3847/1538-3881/abc0e6.

H. Koyama and S.-i. Inutsuka. An Origin of Supersonic Motions in Interstellar Clouds. *Astrophysical Journal, Letters*, 564(2):L97–L100, Jan. 2002. doi: 10.1086/338978.

R. P. Kraft and I. I. Ivans. A Globular Cluster Metallicity Scale Based on the Abundance of Fe II. *Publications of the ASP*, 115(804):143–169, Feb 2003. doi: 10.1086/345914.

M. G. H. Krause, S. S. R. Offner, C. Charbonnel, M. Gieles, R. S. Klessen, E. Vázquez-Semadeni, J. Ballesteros-Paredes, P. Girichidis, J. M. D. Kruijssen, J. L. Ward, and H. Zinnecker. The Physics of Star Cluster Formation and Evolution. *Space Science Reviews*, 216(4):64, June 2020. doi: 10.1007/s11214-020-00689-4.

P. Kroupa. On the variation of the initial mass function. *Monthly Notices of the Royal Astronomical Society*, 322(2):231–246, Apr. 2001. doi: 10.1046/j.1365-8711.2001.04022.x.

P. Kroupa, C. Weidner, J. Pflamm-Altenburg, I. Thies, J. Dabringhausen, M. Marks, and T. Maschberger. *The Stellar and Sub-Stellar Initial Mass Function of Simple and Composite Populations*, volume 5, page 115. 2013. doi: 10.1007/978-94-007-5612-0_4.

J. M. D. Kruijssen, A. Schruba, M. Chevance, S. N. Longmore, A. P. S. Hygate, D. T. Haydon, A. F. McLeod, J. J. Dalcanton, L. J. Tacconi, and E. F. van Dishoeck. Fast and inefficient star formation due to short-lived molecular clouds and rapid feedback. *Nature*, 569(7757): 519–522, May 2019. doi: 10.1038/s41586-019-1194-3.

M. R. Krumholz. *Notes on Star Formation*. Open Astrophysics Bookshelf, 2020. URL https://github.com/Open-Astrophysics-Bookshelf/star_formation_notes.

M. R. Krumholz, C. F. McKee, and J. Bland-Hawthorn. Star Clusters Across Cosmic Time. *Annual Review of Astronomy and Astrophysics*, 57:227–303, Aug. 2019. doi: 10.1146/annurev-astro-091918-104430.

V. F. Ksoll, D. A. Gouliermis, R. S. Klessen, E. K. Grebel, E. Sabbi, J. Anderson, D. J. Lennon, M. Cignoni, G. de Marchi, L. J. Smith, M. Tosi, and R. P. van der Marel. Hubble Tarantula Treasury Project - VI. Identification of pre-main-sequence stars using machine-learning techniques. *Monthly Notices of the Royal Astronomical Society*, 479(2):2389–2414, Sept. 2018. doi: 10.1093/mnras/sty1317.

V. F. Ksoll, L. Ardizzone, R. Klessen, U. Koethe, E. Sabbi, M. Robberto, D. Gouliermis, C. Rother, P. Zeidler, and M. Gennaro. Stellar parameter determination from photometry using invertible neural networks. *Monthly Notices of the Royal Astronomical Society*, 499 (4):5447–5485, Nov. 2020. doi: 10.1093/mnras/staa2931.

V. F. Ksoll, D. Gouliermis, E. Sabbi, J. E. Ryon, M. Robberto, M. Gennaro, R. S. Klessen, U. Koethe, G. de Marchi, C. H. R. Chen, M. Cignoni, and A. E. Dolphin. Measuring Young Stars in Space and Time – I. The Photometric Catalog and Extinction Properties of N44. *Astronomical Journal*, in production, tbd. 2021a. preprint: arXiv:2012.00521.

V. F. Ksoll, D. Gouliermis, E. Sabbi, J. E. Ryon, M. Robberto, M. Gennaro, R. S. Klessen, U. Koethe, G. de Marchi, C. H. R. Chen, M. Cignoni, and A. E. Dolphin. Measuring Young Stars in Space and Time – II. The Pre-Main-Sequence Stellar Content of N44. *Astronomical Journal*, in production, tbd. 2021b. preprint: arXiv:2012.00524.

B. M. Lasker, J. Doggett, B. McLean, C. Sturch, S. Djorgovski, R. R. de Carvalho, and I. N. Reid. The Palomar–ST ScI Digitized Sky Survey (POSS–II): Preliminary Data Availability. In G. H. Jacoby and J. Barnes, editors, *Astronomical Data Analysis Software and Systems V*, volume 101 of *Astronomical Society of the Pacific Conference Series*, page 88, Jan. 1996.

A. K. Leroy, F. Walter, E. Brinks, F. Bigiel, W. J. G. de Blok, B. Madore, and M. D. Thornley. The Star Formation Efficiency in Nearby Galaxies: Measuring Where Gas Forms Stars Effectively. *Astronomical Journal*, 136(6):2782–2845, Dec. 2008. doi: 10.1088/0004-6256/136/6/2782.

D. Leverington. *Babylon to Voyager and Beyond: A History of Planetary Astronomy*. Cambridge University Press, Cambridge, UK, 2003. ISBN 0521808405.

P. B. Lucke and P. W. Hodge. A catalogue of stellar associations in the Large Magellanic Cloud. *Astronomical Journal*, 75:171–175, Mar. 1970. doi: 10.1086/110959.

M.-M. Mac Low and R. S. Klessen. Control of star formation by supersonic turbulence. *Reviews of Modern Physics*, 76(1):125–194, Jan. 2004. doi: $10.1103/\mathrm{RevModPhys.76.125}$.

P. Madau, H. C. Ferguson, M. E. Dickinson, M. Giavalisco, C. C. Steidel, and A. Fruchter. High-redshift galaxies in the Hubble Deep Field: colour selection and star formation history to z∼4. *Monthly Notices of the Royal Astronomical Society*, 283(4):1388–1404, Dec. 1996. doi: $10.1093/\mathrm{mnras}/283.4.1388$.

A. McBride, R. Lingg, M. Kounkel, K. Covey, and B. Hutchinson. Untangling the Galaxy III: Photometric Search for Pre-main Sequence Stars with Deep Learning. *arXiv e-prints*, art. arXiv:2012.10463, Dec. 2020.

G. J. McLachlan and D. A. Peel. *Finite mixture models*. Wiley series in probability and statistics : Applied probability and statistics section. Wiley, New York ; Weinheim [u.a.], 2000. ISBN 0-471-00626-2 and 978-0-471-00626-8.

A. F. McLeod, J. E. Dale, C. J. Evans, A. Ginsburg, J. M. D. Kruijssen, E. W. Pellegrini, S. K. Ramsay, and L. Testi. Feedback from massive stars at low metallicities: MUSE observations of N44 and N180 in the Large Magellanic Cloud. *Monthly Notices of the Royal Astronomical Society*, 486(4):5263–5288, July 2019. doi: $10.1093/\mathrm{mnras}/\mathrm{sty2696}$.

S. Mészáros, T. Masseron, D. A. García-Hernández, C. Allende Prieto, T. C. Beers, D. Bizyaev, D. Chojnowski, R. E. Cohen, K. Cunha, F. Dell'Agli, G. Ebelke, J. G. Fernández-Trincado, P. Frinchaboy, D. Geisler, S. Hasselquist, F. Hearty, J. Holtzman, J. Johnson, R. R. Lane, I. Lacerna, P. Longa-Peña, S. R. Majewski, S. L. Martell, D. Minniti, D. Nataf, D. L. Nidever, K. Pan, R. P. Schiavon, M. Shetrone, V. V. Smith, J. S. Sobeck, G. S. Stringfellow, L. Szigeti, B. Tang, J. C. Wilson, and O. Zamora. Homogeneous analysis of globular clusters from the APOGEE survey with the BACCHUS code - II. The Southern clusters and overview. *Monthly Notices of the Royal Astronomical Society*, 492 (2):1641–1670, Feb. 2020. doi: $10.1093/\mathrm{mnras}/\mathrm{stz3496}$.

G. R. Meurer, T. M. Heckman, M. D. Lehnert, C. Leitherer, and J. Lowenthal. The Panchromatic Starburst Intensity Limit at Low and High Redshift. *Astronomical Journal*, 114: 54–68, July 1997. doi: $10.1086/118452$.

A. A. Miller, J. S. Bloom, J. W. Richards, Y. S. Lee, D. L. Starr, N. R. Butler, S. Tokarz, N. Smith, and J. A. Eisner. A Machine-learning Method to Infer Fundamental Stellar Parameters from Photometric Light Curves. *Astrophysical Journal*, 798(2):122, Jan. 2015. doi: $10.1088/0004\text{-}637\mathrm{X}/798/2/122$.

M. Mirza and S. Osindero. Conditional Generative Adversarial Nets. *arXiv e-prints*, art. arXiv:1411.1784, Nov. 2014.

T. Mitchell. *Machine Learning*. McGraw-Hill International Editions. McGraw-Hill, 1997. ISBN 9780071154673. URL `https://books.google.de/books?id=EoYBngEACAAJ`.

P. C. Myers. Filamentary Structure of Star-forming Complexes. *Astrophysical Journal*, 700 (2):1609–1625, Aug. 2009. doi: $10.1088/0004\text{-}637\mathrm{X}/700/2/1609$.

D. Nardiello, M. Libralato, G. Piotto, J. Anderson, A. Bellini, A. Aparicio, L. R. Bedin, S. Cassisi, V. Granata, I. R. King, F. Lucertini, A. F. Marino, A. P. Milone, S. Ortolani, I. Platais, and R. P. van der Marel. The Hubble Space Telescope UV Legacy Survey of Galactic Globular Clusters - XVII. Public Catalogue Release. *Monthly Notices of the Royal Astronomical Society*, 481(3):3382–3393, Dec. 2018. doi: 10.1093/mnras/sty2515.

M. S. Oey and P. Massey. Triggered Star Formation and the Dynamics of a Superbubble in the LMC: The OB Association LH 47/48 in DEM 152. *Astrophysical Journal*, 452:210, Oct. 1995. doi: 10.1086/176292.

M. S. Oey, J. S. Parker, V. J. Mikles, and X. Zhang. H II Regions in Spiral Galaxies: Size Distribution, Luminosity Function, and New Isochrone Diagnostics of Density-Wave Kinematics. *Astronomical Journal*, 126(5):2317–2329, Nov. 2003. doi: 10.1086/378163.

S. S. R. Offner, P. C. Clark, P. Hennebelle, N. Bastian, M. R. Bate, P. F. Hopkins, E. Moraux, and A. P. Whitworth. The Origin and Universality of the Stellar Initial Mass Function. In H. Beuther, R. S. Klessen, C. P. Dullemond, and T. Henning, editors, *Protostars and Planets VI*, page 53, Jan. 2014. doi: 10.2458/azu_uapress_9780816531240-ch003.

R. Olney, M. Kounkel, C. Schillinger, M. T. Scoggins, Y. Yin, E. Howard, K. R. Covey, B. Hutchinson, and K. G. Stassun. APOGEE Net: Improving the Derived Spectral Parameters for Young Stars through Deep Learning. *Astronomical Journal*, 159(4):182, Apr. 2020. doi: 10.3847/1538-3881/ab7a97.

J. H. Oort. Outline of a theory on the origin and acceleration of interstellar clouds and O associations. *Bulletin Astronomical Institute of the Netherlands*, 12:177, Sept. 1954.

N. Panagia, R. Gilmozzi, F. Macchetto, H. M. Adorf, and R. P. Kirshner. Properties of the SN 1987A Circumstellar Ring and the Distance to the Large Magellanic Cloud. *Astrophysical Journal, Letters*, 380:L23, Oct. 1991. doi: 10.1086/186164.

G. Papamakarios and I. Murray. Fast ϵ-free Inference of Simulation Models with Bayesian Conditional Density Estimation. *arXiv e-prints*, art. arXiv:1605.06376, May 2016.

G. Papamakarios, T. Pavlakou, and I. Murray. Masked autoregressive flow for density estimation. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/file/6c1da886822c67822bcf3679d04369fa-Paper.pdf.

B. Paxton, L. Bildsten, A. Dotter, F. Herwig, P. Lesaffre, and F. Timmes. Modules for Experiments in Stellar Astrophysics (MESA). *Astrophysical Journal, Supplement*, 192(1): 3, Jan. 2011. doi: 10.1088/0067-0049/192/1/3.

B. Paxton, M. Cantiello, P. Arras, L. Bildsten, E. F. Brown, A. Dotter, C. Mankovich, M. H. Montgomery, D. Stello, F. X. Timmes, and R. Townsend. Modules for Experiments in Stellar Astrophysics (MESA): Planets, Oscillations, Rotation, and Massive Stars. *Astrophysical Journal, Supplement*, 208(1):4, Sept. 2013. doi: 10.1088/0067-0049/208/1/4.

B. Paxton, P. Marchant, J. Schwab, E. B. Bauer, L. Bildsten, M. Cantiello, L. Dessart, R. Farmer, H. Hu, N. Langer, R. H. D. Townsend, D. M. Townsley, and F. X. Timmes. Modules for Experiments in Stellar Astrophysics (MESA): Binaries, Pulsations, and Explosions. *Astrophysical Journal, Supplement*, 220(1):15, Sept. 2015. doi: 10.1088/0067-0049/220/1/15.

B. Paxton, J. Schwab, E. B. Bauer, L. Bildsten, S. Blinnikov, P. Duffell, R. Farmer, J. A. Goldberg, P. Marchant, E. Sorokina, A. Thoul, R. H. D. Townsend, and F. X. Timmes. Modules for Experiments in Stellar Astrophysics (MESA): Convective Boundaries, Element Diffusion, and Massive Star Explosions. *Astrophysical Journal, Supplement*, 234(2):34, Feb. 2018. doi: 10.3847/1538-4365/aaa5a8.

B. Paxton, R. Smolec, J. Schwab, A. Gautschy, L. Bildsten, M. Cantiello, A. Dotter, R. Farmer, J. A. Goldberg, A. S. Jermyn, S. M. Kanbur, P. Marchant, A. Thoul, R. H. D. Townsend, W. M. Wolf, M. Zhang, and F. X. Timmes. Modules for Experiments in Stellar Astrophysics (MESA): Pulsating Variable Stars, Rotation, Convective Boundaries, and Energy Conservation. *Astrophysical Journal, Supplement*, 243(1):10, July 2019. doi: 10.3847/1538-4365/ab2241.

Y. C. Pei, S. M. Fall, and M. G. Hauser. Cosmic Histories of Stars, Gas, Heavy Elements, and Dust in Galaxies. *Astrophysical Journal*, 522(2):604–626, Sept. 1999. doi: 10.1086/307674.

E. W. Pellegrini, M. S. Oey, P. F. Winkler, S. D. Points, R. C. Smith, A. E. Jaskot, and J. Zastrow. The Optical Depth of H II Regions in the Magellanic Clouds. *Astrophysical Journal*, 755(1):40, Aug. 2012. doi: 10.1088/0004-637X/755/1/40.

T. Peters, R. S. Klessen, M.-M. Mac Low, and R. Banerjee. Limiting Accretion onto Massive Stars by Fragmentation-induced Starvation. *Astrophysical Journal*, 725(1):134–145, Dec. 2010. doi: 10.1088/0004-637X/725/1/134.

G. Pietrzyński, D. Graczyk, W. Gieren, I. B. Thompson, B. Pilecki, A. Udalski, I. Soszyński, S. Kozłowski, P. Konorski, K. Suchomska, G. Bono, P. G. P. Moroni, S. Villanova, N. Nardetto, F. Bresolin, R. P. Kudritzki, J. Storm, A. Gallenne, R. Smolec, D. Minniti, M. Kubiak, M. K. Szymański, R. Poleski, Ł. Wyrzykowski, K. Ulaczyk, P. Pietrukowicz, M. Górski, and P. Karczmarek. An eclipsing-binary distance to the Large Magellanic Cloud accurate to two per cent. *Nature*, 495(7439):76–79, Mar. 2013. doi: 10.1038/nature11878.

G. Piotto, A. P. Milone, L. R. Bedin, J. Anderson, I. R. King, A. F. Marino, D. Nardiello, A. Aparicio, B. Barbuy, A. Bellini, T. M. Brown, S. Cassisi, A. M. Cool, A. Cunial, E. Dalessandro, F. D'Antona, F. R. Ferraro, S. Hidalgo, B. Lanzoni, M. Monelli, S. Ortolani, A. Renzini, M. Salaris, A. Sarajedini, R. P. van der Marel, E. Vesperini, and M. Zoccali. The Hubble Space Telescope UV Legacy Survey of Galactic Globular Clusters. I. Overview of the Project and Detection of Multiple Stellar Populations. *Astronomical Journal*, 149(3):91, Mar. 2015. doi: 10.1088/0004-6256/149/3/91.

J. C. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In *ADVANCES IN LARGE MARGIN CLASSIFIERS*, pages 61–74. MIT Press, 1999.

D. Rezende and S. Mohamed. Variational inference with normalizing flows. In F. Bach and D. Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 1530–1538, Lille, France, 07–09 Jul 2015. PMLR. URL `http://proceedings.mlr.press/v37/rezende15.html`.

M. Richter. *Inverse Probleme.* Springer eBook Collection. Springer Spektrum, Berlin, Heidelberg, 2015. ISBN 978-3-662-45811-2. doi: $10.1007/978\text{-}3\text{-}662\text{-}45811\text{-}2$. URL `http://dx.doi.org/10.1007/978-3-662-45811-2`.

O. Rippel and R. Prescott Adams. High-Dimensional Probability Estimation with Deep Density Models. *arXiv e-prints*, art. arXiv:1302.5125, Feb. 2013.

C. P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer texts in statistics. Springer, New York, NY, 2. ed. edition, 2004. ISBN 0-387-21239-6 and 978-0-387-21239-5. Literaturverz. S. [591] - 622.

A. W. Rodgers, C. T. Campbell, and J. B. Whiteoak. A catalogue of Hα-emission regions in the southern Milky Way. *Monthly Notices of the Royal Astronomical Society*, 121:103, Jan. 1960. doi: $10.1093/\mathrm{mnras}/121.1.103$.

A. L. Rosen, S. S. R. Offner, S. I. Sadavoy, A. Bhandare, E. Vázquez-Semadeni, and A. Ginsburg. Zooming in on Individual Star Formation: Low- and High-Mass Stars. *Space Science Reviews*, 216(4):62, May 2020. doi: $10.1007/\mathrm{s}11214\text{-}020\text{-}00688\text{-}5$.

J. Rousseau, N. Martin, L. Prévot, E. Rebeirot, A. Robin, and J. P. Brunet. Studies of the Large Magellanic Cloud stellar content: III. Spectral types and V magnitudes of 1822 members. *Astronomy and Astrophysics, Supplement*, 31:243, Feb. 1978.

E. Sabbi, J. Anderson, D. J. Lennon, R. P. van der Marel, A. Aloisi, M. L. Boyer, M. Cignoni, G. de Marchi, S. E. de Mink, C. J. Evans, I. Gallagher, J. S., K. Gordon, D. A. Gouliermis, E. K. Grebel, A. M. Koekemoer, S. S. Larsen, N. Panagia, J. E. Ryon, L. J. Smith, M. Tosi, and D. Zaritsky. Hubble Tarantula Treasury Project: Unraveling Tarantula's Web. I. Observational Overview and First Results. *Astronomical Journal*, 146(3):53, Sept. 2013. doi: $10.1088/0004\text{-}6256/146/3/53$.

E. Sabbi, D. J. Lennon, J. Anderson, M. Cignoni, R. P. van der Marel, D. Zaritsky, G. De Marchi, N. Panagia, D. A. Gouliermis, E. K. Grebel, I. Gallagher, J. S., L. J. Smith, H. Sana, A. Aloisi, M. Tosi, C. J. Evans, H. Arab, M. Boyer, S. E. de Mink, K. Gordon, A. M. Koekemoer, S. S. Larsen, J. E. Ryon, and P. Zeidler. Hubble Tarantula Treasury Project. III. Photometric Catalog and Resulting Constraints on the Progression of Star Formation in the 30 Doradus Region. *Astrophysical Journal, Supplement*, 222(1):11, Jan. 2016. doi: $10.3847/0067\text{-}0049/222/1/11$.

E. Sabbi, M. Gennaro, J. Anderson, V. Bajaj, N. Bastian, I. Gallagher, J. S., M. Gieles, D. J. Lennon, A. Nota, K. C. Sahu, and P. Zeidler. Time-domain Study of the Young Massive Cluster Westerlund 2 with the Hubble Space Telescope. I. *Astrophysical Journal*, 891(2):182, Mar. 2020. doi: $10.3847/1538\text{-}4357/\mathrm{ab}7372$.

E. E. Salpeter. The Luminosity Function and Stellar Evolution. *Astrophysical Journal*, 121:161, Jan. 1955. doi: $10.1086/145971$.

D. J. Schlegel, D. P. Finkbeiner, and M. Davis. Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds. *Astrophysical Journal*, 500(2):525–553, Jun 1998. doi: 10.1086/305772.

F. R. N. Schneider, N. Langer, A. de Koter, I. Brott, R. G. Izzard, and H. H. B. Lau. Bonnsai: a Bayesian tool for comparing stars with stellar evolution models. *Astronomy and Astrophysics*, 570:A66, Oct. 2014. doi: 10.1051/0004-6361/201424286.

F. R. N. Schneider, N. Castro, L. Fossati, N. Langer, and A. de Koter. BONNSAI: correlated stellar observables in Bayesian methods. *Astronomy and Astrophysics*, 598:A60, Feb. 2017. doi: 10.1051/0004-6361/201628409.

F. R. N. Schneider, O. H. Ramírez-Agudelo, F. Tramper, J. M. Bestenlehner, N. Castro, H. Sana, C. J. Evans, C. Sabín-Sanjulián, S. Simón-Díaz, N. Langer, L. Fossati, G. Gräfener, P. A. Crowther, S. E. de Mink, A. de Koter, M. Gieles, A. Herrero, R. G. Izzard, V. Kalari, R. S. Klessen, D. J. Lennon, L. Mahy, J. Maíz Apellániz, N. Markova, J. T. van Loon, J. S. Vink, and N. R. Walborn. The VLT-FLAMES Tarantula Survey. XXIX. Massive star formation in the local 30 Doradus starburst. *Astronomy and Astrophysics*, 618:A73, Oct. 2018. doi: 10.1051/0004-6361/201833433.

N. S. Schulz. *The formation and early evolution of stars*. Astronomy and astrophysics library. Springer, Berlin ; Heidelberg [u.a.], 2. ed. edition, 2012. ISBN 978-3-642-23925-0 and 3-642-23925-0. Literaturangaben.

A. E. Shapley, C. C. Steidel, M. Pettini, and K. L. Adelberger. Rest-Frame Ultraviolet Spectra of z~3 Lyman Break Galaxies. *Astrophysical Journal*, 588(1):65–89, May 2003. doi: 10.1086/373922.

K. Sharma, A. Kembhavi, A. Kembhavi, T. Sivarani, S. Abraham, and K. Vaghmare. Application of convolutional neural networks for stellar spectral classification. *Monthly Notices of the Royal Astronomical Society*, 491(2):2280–2300, Jan. 2020. doi: 10.1093/mnras/stz3100.

N. Siddharth, B. Paige, J.-W. van de Meent, A. Desmaison, N. D. Goodman, P. Kohli, F. Wood, and P. H. S. Torr. Learning Disentangled Representations with Semi-Supervised Deep Generative Models. *arXiv e-prints*, art. arXiv:1706.00400, June 2017.

R. A. Skibba, C. W. Engelbracht, G. Aniano, B. Babler, J.-P. Bernard, C. Bot, L. R. Carlson, M. Galametz, F. Galliano, K. Gordon, S. Hony, F. Israel, V. Lebouteiller, A. Li, S. Madden, M. Meixner, K. Misselt, E. Montiel, K. Okumura, P. Panuzzo, D. Paradis, J. Roman-Duval, M. Rubio, M. Sauvage, J. Seale, S. Srinivasan, and J. T. van Loon. The Spatial Distribution of Dust and Stellar Emission of the Magellanic Clouds. *Astrophysical Journal*, 761(1):42, Dec. 2012. doi: 10.1088/0004-637X/761/1/42.

M. F. Skrutskie, R. M. Cutri, R. Stiening, M. D. Weinberg, S. Schneider, J. M. Carpenter, C. Beichman, R. Capps, T. Chester, J. Elias, J. Huchra, J. Liebert, C. Lonsdale, D. G. Monet, S. Price, P. Seitzer, T. Jarrett, J. D. Kirkpatrick, J. E. Gizis, E. Howard, T. Evans, J. Fowler, L. Fullmer, R. Hurt, R. Light, E. L. Kopan, K. A. Marsh, H. L. McCallon, R. Tam, S. Van Dyk, and S. Wheelock. The Two Micron All Sky Survey (2MASS). *Astronomical Journal*, 131(2):1163–1183, Feb. 2006. doi: 10.1086/498708.

K. Sohn, H. Lee, and X. Yan. Learning structured output representation using deep conditional generative models. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL `https://proceedings.neurips.cc/paper/2015/file/8d55a249e6baa5c06772297520da2051-Paper.pdf`.

G. Srinivasan. *Life and death of the stars*. Undergraduate lecture notes in physics. Springer, Berlin ; Heidelberg, 2014. ISBN 3-642-45383-X and 978-3-642-45383-0. Literaturangaben.

S. W. Stahler and F. Palla. *The formation of stars*. Physics textbook. Wiley-VCH, Weinheim, [1. ed.] edition, 2004. ISBN 978-3-527-40559-6 and 3-527-40559-3.

T. P. Stecher and D. A. Williams. Photodestruction of Hydrogen Molecules in H I Regions. *Astrophysical Journal, Letters*, 149:L29, July 1967. doi: $10.1086/180047$.

M. Sunnåker, A. G. Busetto, E. Numminen, J. Corander, M. Foll, and C. Dessimoz. Approximate bayesian computation. *PLOS Computational Biology*, 9(1):1–10, 01 2013. doi: $10.1371/journal.pcbi.1002803$. URL `https://doi.org/10.1371/journal.pcbi.1002803`.

E. Tabak and C. Turner. A family of nonparametric density estimation algorithms. *Communications on Pure and Applied Mathematics*, 66(2):145–164, Feb. 2013. ISSN 0010-3640. doi: $10.1002/cpa.21423$. Copyright: Copyright 2012 Elsevier B.V., All rights reserved.

E. Tabak and E. Vanden-Eijnden. Density estimation by dual ascent of the log-likelihood. *Communications in Mathematical Sciences*, 8(1):217–233, 2010. ISSN 1539-6746. doi: $10.4310/CMS.2010.v8.n1.a11$. Copyright: Copyright 2017 Elsevier B.V., All rights reserved.

J. Tang, A. Bressan, P. Rosenfield, A. Slemer, P. Marigo, L. Girardi, and L. Bianchi. New PARSEC evolutionary tracks of massive stars at low metallicity: testing canonical stellar evolution in nearby star-forming dwarf galaxies. *Monthly Notices of the Royal Astronomical Society*, 445(4):4287–4305, Dec 2014. doi: $10.1093/mnras/stu2029$.

A. Tarantola. *Inverse problem theory*. SIAM, Society for Industrial and Applied Mathematics, Philadelphia, 2005. ISBN 0-89871-572-5 and 978-0-89871-572-9. Includes bibliographical references and index.

E. J. Tasker and J. Tan. Formation and Evolution of Giant Molecular Clouds in Disk Galaxies. In *American Astronomical Society Meeting Abstracts #213*, volume 213 of *American Astronomical Society Meeting Abstracts*, page 347.05, Jan. 2009.

J. M. Tomczak and M. Welling. Improving Variational Auto-Encoders using Householder Flow. *arXiv e-prints*, art. arXiv:1611.09630, Nov. 2016.

B. L. Trippe and R. E. Turner. Conditional Density Estimation with Bayesian Normalising Flows. *arXiv e-prints*, art. arXiv:1802.04908, Feb. 2018.

B. Uria, M.-A. Côté, K. Gregor, I. Murray, and H. Larochelle. Neural autoregressive distribution estimation. *Journal of Machine Learning Research*, 17(205):1–37, 2016. URL `http://jmlr.org/papers/v17/16-272.html`.

D. Valls-Gabaud. Bayesian isochrone fitting and stellar ages. In *EAS Publications Series*, volume 65 of *EAS Publications Series*, pages 225–265, Nov. 2014. doi: 10.1051/eas/1465006.

E. F. van Dishoeck and J. H. Black. The Photodissociation and Chemistry of Interstellar CO. *Astrophysical Journal*, 334:771, Nov. 1988. doi: 10.1086/166877.

C. A. Vargas Álvarez, H. A. Kobulnicky, D. R. Bradley, S. J. Kannappan, M. A. Norris, R. J. Cool, and B. P. Miller. The Distance to the Massive Galactic Cluster Westerlund 2 from a Spectroscopic and HST Photometric Study. *Astronomical Journal*, 145(5):125, May 2013. doi: 10.1088/0004-6256/145/5/125.

E. Vázquez-Semadeni, J. Ballesteros-Paredes, and R. S. Klessen. A Holistic Scenario of Turbulent Molecular Cloud Evolution and Control of the Star Formation Efficiency: First Tests. *Astrophysical Journal, Letters*, 585(2):L131–L134, Mar. 2003. doi: 10.1086/374325.

E. Vázquez-Semadeni, A. Palau, J. Ballesteros-Paredes, G. C. Gómez, and M. Zamora-Avilés. Global hierarchical collapse in molecular clouds. Towards a comprehensive scenario. *Monthly Notices of the Royal Astronomical Society*, 490(3):3061–3097, Dec. 2019. doi: 10.1093/mnras/stz2736.

W. N. Venables and B. D. Ripley. *Modern applied statistics with S*. Statistics and computing. Springer, New York ; Berlin ; Heidelberg [u.a.], 4. ed. edition, 2002. ISBN 0-387-95457-0 and 978-0-387-95457-8. Previously published as: Modern applied statistics with S-PLUS, c1999 - Includes bibliographical references and index ; Hier auch später erschienene unveränderte Nachdrucke.

N. Vulic, P. Barmby, and S. C. Gallagher. Milky Way globular cluster metallicity and low-mass X-ray binaries: the red giant influence. *Monthly Notices of the Royal Astronomical Society*, 473(4):4900–4925, Feb 2018. doi: 10.1093/mnras/stx2626.

N. R. Walborn and J. C. Blades. Spectral Classification of the 30 Doradus Stellar Populations. *Astrophysical Journal, Supplement*, 112(2):457–485, Oct. 1997. doi: 10.1086/313043.

S. Walch, A. P. Whitworth, T. G. Bisbas, R. Wünsch, and D. A. Hubber. Clumps and triggered star formation in ionized molecular clouds. *Monthly Notices of the Royal Astronomical Society*, 435(2):917–927, Oct. 2013. doi: 10.1093/mnras/stt1115.

C. J. Wareing, S. A. E. G. Falle, and J. M. Pittard. Sheets, filaments, and clumps - high-resolution simulations of how the thermal instability can form molecular clouds. *Monthly Notices of the Royal Astronomical Society*, 485(4):4686–4702, June 2019. doi: 10.1093/mnras/stz768.

J. M. Will, D. J. Bomans, and A. Dieball. A photometric and spectroscopic study of the association LH47 in the superbubble N44 in the LMC. *Astronomy and Astrophysics, Supplement*, 123:455–471, June 1997. doi: 10.1051/aas:1997169.

T. Wong, A. Hughes, J. Ott, E. Muller, J. L. Pineda, J.-P. Bernard, Y.-H. Chu, Y. Fukui, R. A. Gruendl, C. Henkel, A. Kawamura, U. Klein, L. W. Looney, S. Maddison, Y. Mizuno, D. Paradis, J. Seale, and D. E. Welty. The Magellanic Mopra Assessment (MAGMA). I. The Molecular Cloud Population of the Large Magellanic Cloud. *Astrophysical Journal, Supplement*, 197(2):16, Dec. 2011. doi: 10.1088/0067-0049/197/2/16.

P. Zeidler, E. Sabbi, A. Nota, E. K. Grebel, M. Tosi, A. Z. Bonanos, A. Pasquali, C. Christian, S. E. de Mink, and L. Ubeda. A High-resolution Multiband Survey of Westerlund 2 with the Hubble Space Telescope. I. Is the Massive Star Cluster Double? *Astronomical Journal*, 150(3):78, Sep 2015. doi: 10.1088/0004-6256/150/3/78.

P. Zeidler, E. K. Grebel, A. Nota, E. Sabbi, A. Pasquali, M. Tosi, A. Z. Bonanos, and C. Christian. A High-resolution Multiband Survey of Westerlund 2 with the Hubble Space Telescope. II. Mass Accretion in the Pre-main-sequence Population. *Astronomical Journal*, 152(4):84, Oct. 2016. doi: 10.3847/0004-6256/152/4/84.

P. Zeidler, A. Nota, E. K. Grebel, E. Sabbi, A. Pasquali, M. Tosi, and C. Christian. A High-resolution Multiband Survey of Westerlund 2 with the Hubble Space Telescope. III. The Present-day Stellar Mass Function. *Astronomical Journal*, 153(3):122, Mar. 2017. doi: 10.3847/1538-3881/153/3/122.

P. Zeidler, E. Sabbi, A. Nota, and A. F. McLeod. The young massive star cluster Westerlund 2 observed with MUSE. III. A cluster in motion – the complex internal dynamics. *arXiv e-prints*, art. arXiv:2101.01694, Jan. 2021.

V. Zivkov, J. M. Oliveira, M. G. Petr-Gotzens, M.-R. L. Cioni, S. Rubele, J. T. van Loon, K. Bekki, F. Cusano, R. de Grijs, V. D. Ivanov, M. Marconi, F. Niederhofer, V. Ripepi, and N.-C. Sun. The VMC Survey. XXXII. Pre-main-sequence populations in the Large Magellanic Cloud. *Astronomy and Astrophysics*, 620:A143, Dec. 2018. doi: 10.1051/0004-6361/201833951.

Appendix

A

List of Abbreviations

ABC   Approximate Bayesian computation.

ACS   Advanced Camera for Surveys.

AGB   Asymptotic giant branch.

AI   Artificial intelligence.

AJ   The Astronomical Journal.

AURA   Association of Universities for Research in Astronomy.

cGAN   Conditional generative adversarial network.

CHEOPS   Characterising Exoplanet Satellite.

cINN   Conditional invertible neural network.

CMD   Colour-Magnitude Diagram.

CNM   Cold neutral medium.

CTTS   Classical T Tauri star.

cVAE   Conditional variational auto encoder.

ELT   Extremely Large Telescope.

EM   Expectation-Maximisation.

ESA   European Space Agency.

FoV   Field of view.

GHC	Global hierarchical collapse.
GMC	Giant molecular cloud.
GT	Gravoturbulent.
HIM	Hot ionised medium.
HRD	Hertzsprung-Russell Diagram.
HST	Hubble Space Telescope.
HTTP	Hubble Tarantula Treasury Project.
HUGS	HST UV Globular Cluster Survey.
IMF	Initial mass function.
INN	Invertible neural network.
IR	Infrared.
ISM	Interstellar medium.
KH	Kelvin-Helmholtz.
LMC	Large Magellanic Cloud.
LMS	Lower main-sequence.
MC	Molecular Cloud.
MCMC	Markove Chain Monte Carlo.
MESA	Modules for Experiments in Stellar Astrophysics.
MIST	MESA Isochrones and Stellar Tracks.
ML	Machine Learning.
MMD	Maximum Mean Discrepancy.
MNRAS	Monthly Notices of the Royal Astronomical Society.
MSE	Mean squared error.
MYSST	Measuring Young Stars in Space and Time.

NASA	National Aeronautics and Space Administration.
NNDE	Nearest neighbour density estimate.
Pan-STARRS	Panoramic Survey Telescope And Rapid Response System.
PARSEC	Padova and Trieste Stellar Evolution Code.
PMS	Pre-main-sequence.
RANSAC	Random Sample Consensus.
RC	Red Clump.
ReLU	Rectified Linear Unit.
RF	Random Forest.
RGB	Red giant branch.
SDSS	Sloan Digital Sky Survey.
SED	Spectral energy distribution.
SMC	Small Magellanic Cloud.
STScI	Space Telescope Science Institute.
SVM	Support Vector Machine.
TP	Thermal pulse.
TP-AGB	Thermal pulsing asymptotic giant branch.
UMS	Upper main-sequence.
UV	Ultraviolet.
UVIS	Ultraviolet and visible (light).
VLT	Very Large Telescope.
VMC	VISTA Survey of the Magellanic Clouds.
VRO	Vera C. Rubin Observatory.

Wd2	Westerlund 2.
WFC	Wide Field Channel.
WFC3	Wide Field Camera 3.
WIM	Warm ionised medium.
WNM	Warm neutral medium.
WTTS	Weak-lined T Tauri stars.
YSO	Young stellar object.
ZAMS	Zero-age-main-sequence.