

# Dissecting stellar populations with manifold learning

## I. Validation of the method on a synthetic Milky Way-like galaxy

Neitzel, A. W.<sup>1,2</sup>, Campante, T. L.<sup>1,2</sup>, Bossini, D.<sup>3,4</sup>, Miglio, A.<sup>5,6</sup>

<sup>1</sup> Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, 4150-762 Porto, Portugal  
e-mail: andreasneitzel@astro.up.pt

<sup>2</sup> Departamento de Física e Astronomia, Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre, s/n, 4169-007 Porto, Portugal

<sup>3</sup> Dipartimento di Fisica e Astronomia Galileo Galilei, Università di Padova, Vicolo dell'Osservatorio 3, 35122 Padova, Italy

<sup>4</sup> INAF – Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, 35122 Padova, Italy

<sup>5</sup> Department of Physics and Astronomy, University of Bologna, Via P. Gobetti 93/2, 40129 Bologna, Italy

<sup>6</sup> INAF – Osservatorio di Astrofisica e Scienza dello Spazio, Via P. Gobetti 93/3, 40129 Bologna, Italy

Received 30 July 2024 / Accepted 21 January 2025

### ABSTRACT

**Context.** Different stellar populations may be identified through differences in chemical, kinematic, and chronological properties, suggesting the interplay of various physical mechanisms that led to their origin and subsequent evolution. As such, the identification of stellar populations is key for gaining insight into the evolutionary history of the Milky Way galaxy. This task is complicated by the fact that stellar populations share significant overlap in their chrono-chemo-kinematic properties, hindering efforts to identify and define stellar populations.

**Aims.** Our goal is to offer a novel and effective methodology that can provide deeper insight into the nonlinear and nonparametric properties of the multidimensional physical parameters that define stellar populations.

**Methods.** For this purpose we explore the ability of manifold learning to differentiate stellar populations with minimal assumptions about their number and nature. Manifold learning is an unsupervised machine learning technique that seeks to intelligently identify and disentangle manifolds hidden within the input data. To test this method, we make use of *Gaia* DR3-like synthetic stellar samples generated from the FIRE-2 cosmological simulations. These represent red-giant stars constrained by asteroseismic data from TESS.

**Results.** We reduce the 5-dimensional input chrono-chemo-kinematic parameter space into 2-dimensional latent space embeddings generated by manifold learning. We then study these embeddings to assess how accurately they represent the original data and whether they contain meaningful information that can be used to discern stellar populations.

**Conclusions.** We conclude that manifold learning possesses promising abilities to differentiate stellar populations when considering realistic observational constraints.

**Key words.** Galaxy: structure – Galaxy: stellar content – Galaxy: evolution – methods: data analysis – asteroseismology – stars: oscillations

## 1. Introduction

Galactic archaeology is the study of the assembly and evolution history of the Milky Way (Freeman & Bland-Hawthorn 2002). The Milky Way, as we observe it, is not a monolithic structure, but instead an agglomeration of different components that differ in structural, kinematic, and chemical properties (Fuhrmann 1998; Bensby et al. 2003; Reddy et al. 2006; Bland-Hawthorn & Gerhard 2016). These include the thin and thick disk (Gilmore & Reid 1983), often identified as the low- $\alpha$  and high- $\alpha$  disk due to their dichotomy in  $[\alpha/\text{Fe}]$ – $[\text{Fe}/\text{H}]$  space (Adibekyan et al. 2012), the stellar halo (Helmi 2008) and the bulge (Shen & Li 2016). Other prominent stellar populations include stellar populations resulting from accretion events, such as streams from disrupted galaxies (e.g., Sagittarius Dwarf Spheroidal galaxy; Ibata et al. 1994; Vasiliev & Belokurov 2020) and the famous *Gaia*-Enceladus (Helmi et al. 2018), considered the last large merger event in the history of the Galaxy.

In addition to their differences in chemical and kinematic properties, dating these events is key. Stellar ages allow us to

better constrain the timeline of accreted events (Chaplin et al. 2020; Montalbán et al. 2021) and the formation of different stellar populations (Silva Aguirre et al. 2018). The current state of the art in stellar age estimation comes from asteroseismology, i.e., the study and observation of stellar oscillations (see, e.g., García & Ballot 2019). When global asteroseismic parameters are used, uncertainties in the age estimations of hydrogen-shell burning (RGB) and helium-core burning (HeCB) red-giant stars can reach up to 20–25% (Silva Aguirre et al. 2020; Mackereth et al. 2021; Stello et al. 2022; Campante et al. 2023). When the signal-to-noise ratio (SNR) and observation duration are sufficient, individual modes can also be used to double the precision in age (Metcalfe et al. 2014). This has been achieved in previous asteroseismic missions and is an expected goal of the upcoming Planetary Transits and Oscillations of Stars mission (PLATO; Miglio et al. 2017). By expanding our understanding of the Milky Way, we can impose more precise constraints on stellar nucleosynthesis and the timescales of formation of the various Galactic components (Matteucci 2021), and develop gas infall models that can provide a theoretical framework for

the evolution of the Galaxy (Noguchi 2018; Spitoni et al. 2019, 2023).

Scientific advances in Galactic archaeology have greatly benefited from large-scale surveys, particularly those including asteroseismic parameters (Anders et al. 2017; Miglio et al. 2021). The latest *Gaia* Data Release 3 (DR3; Lindgren et al. 2021; Gaia Collaboration et al. 2023) provides astrometry and broadband photometry for  $1.8 \times 10^9$  objects. Of these,  $3.38 \times 10^7$  contain radial velocities, providing a complete 6-dimensional picture of stellar kinematics. Spectroscopic surveys, such as Galactic Archaeology with High Efficiency and Resolution Multi-Element Spectrograph (GALAH; Buder et al. 2021), Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST; Cui et al. 2012), and Apache Point Observatory Galactic Evolution Experiment (APOGEE; Majewski et al. 2017), provide chemical abundances, with APOGEE DR17 providing spectroscopic measurements for 372,458 targets (Abdurro'uf et al. 2022). Future surveys collect from the 4-metre Multi-Object Spectroscopic Telescope (4MOST; de Jong et al. 2019), and the William Herschel Telescope (WHT) Enhanced Area Velocity Explorer (WEAVE; Dalton et al. 2014; Jin et al. 2024), are expected to significantly expand the volume of available data. Asteroseismic parameters have been obtained from surveys such as the Convection, Rotation and planetary Transits mission (CoRoT; Baglin et al. 2009), the *Kepler* and K2 missions (Borucki et al. 2010; Howell et al. 2014), and Transiting Exoplanets Survey Satellite (TESS; Ricker et al. 2015), with the PLATO mission scheduled for launch in 2026 (Rauer et al. 2014). The chronological, chemical, and kinematic data obtained from these surveys can be cross-matched to create stellar catalogs with complete chrono-chemo-kinematic parameters, which are expected to provide insight into the stellar populations of the Galaxy.

As the volume of data increases and new variables are included, the complexity of extracting valuable information from the data also increases. Insight into stellar populations may be hidden in the form of complex, nonlinear relations that span the multidimensional data. Such relations become harder to uncover as the number of parameters (i.e., dimensions) increases. A potential solution lies in data-driven approaches such as machine learning. Machine learning (ML) is the field of computer programs that are capable of autonomously solving complex problems without following explicit instructions. This is accomplished through statistical models used to draw inferences from patterns in the data. ML methodologies are gaining popularity in Galactic archaeology, with methods frequently tested in  $\Lambda$ CDM (Davis et al. 1985) cosmological zoom-in simulations, such as FIREbox (Feldmann et al. 2023) and Auriga (Grand et al. 2024). Ostdiek et al. (2020) trained deep learning neural networks on synthetic stars from the FIRE-2 simulations (Hopkins et al. 2018) to develop a catalog of accreted stars in the Milky Way. Necib et al. (2020) applied Gaussian Mixture Models (GMM) and Density-Based Spatial Clustering of Applications with Noise (DBSCAN; Ester et al. 1996) to the aforementioned catalog in the search for accreted structures. In Nikakhtar et al. (2021) and Stokholm et al. (2023), GMM was used to search for stellar populations. Sante et al. (2024), artificial neural networks, decision trees and dimensionality reduction techniques were applied on synthetic Milky Way-like galaxies from the ARTEMIS simulation (Font et al. 2020) to train a model capable of differentiating accreted from in situ stars. In Queiroz et al. (2023), the Bayesian isochrone-fitting code StarHorse was used to derive stellar ages, which, in conjunction with chemical abundances, permitted the construction

of a chrono-chemical stellar sample that was then analyzed using the t-distributed Stochastic Neighbor Embedding (t-SNE; van der Maaten 2013) dimensionality reduction technique and the Hierarchical DBSCAN (HDBSCAN; Campello et al. 2013) clustering algorithm in the search for stellar populations.

In this paper, we tackle the question of how to effectively uncover patterns and commonalities in the data, as well as 'disentangle' stellar populations embedded in high-dimensional data, through a technique referred to as manifold learning. For this purpose we make use of the Uniform Manifold Approximation and Projection for Dimension Reduction algorithm (UMAP; McInnes et al. 2018), a state-of-the-art non-parametric manifold learning dimensionality reduction algorithm. We explore the potential of UMAP to assist in identifying stellar populations while making minimum assumptions about the nature of the data. To validate the method, we test it on four different synthetic stellar samples with known ground truth.

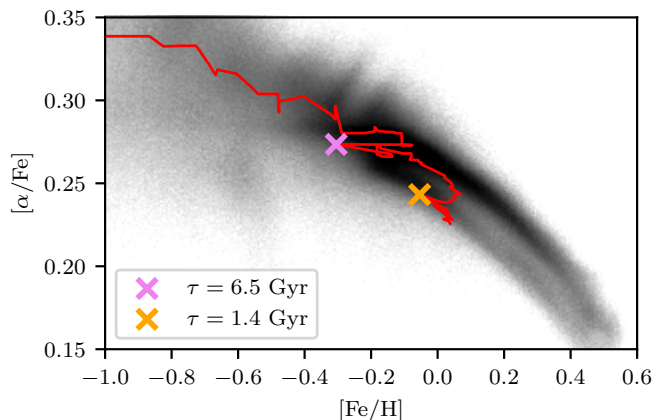
The structure of this work is as follows. Section 2 describes the construction of the synthetic stellar samples. Section 3 provides a brief review of the inner workings of UMAP and a detailed explanation of its application to the test samples. The results are presented in Sect. 4. Finally, we provide a summary of the results and our conclusions in Sect. 5. Further plots are provided in the Appendix.

## 2. Data

### 2.1. Synthetic Milky Way-like galaxy

Feedback In Realistic Environments (FIRE-1, later FIRE-2; Hopkins et al. 2014, 2018) is a numerical implementation of feedback mechanisms that regulate star formation in the GIZMO cosmological evolution code (Hopkins 2015). The *Latte* project (Wetzel et al. 2016; Hopkins et al. 2018) aims to replicate Milky Way-like galaxies in the FIRE-2 cosmological zoom-in simulations, resulting in three synthetic galaxies with Milky Way-like mass and morphology (m12i, m12f and m12m). For each of these galaxies, Sanderson et al. (2020) selected three heliocentric-like local standards of rest (LSR) and implemented the Ananke framework, which decomposes the original stellar particles of the FIRE-2 simulation at redshift  $z = 0$  into individual stars and subsequently convolves their observable properties using *Gaia* DR2 uncertainties, thereby generating synthetic *Gaia* DR2-like surveys from the FIRE-2 simulations. This procedure was later updated with *Gaia* DR3 uncertainties in Nguyen et al. (2024). The full FIRE-2 public data release is provided in Wetzel et al. (2023). For this work, we use the m12f-1sr0 *Gaia* DR3-like synthetic catalog.

The properties of the synthetic galaxy m12f are described in Hopkins et al. (2018) and summarized below. It is a spiral galaxy with virial mass  $M^{\text{vir}} = 1.6 \times 10^{12} M_{\odot}$  (virial radius  $R^{\text{vir}} = 306$  kpc) and a stellar mass  $M = 8.0 \times 10^{10} M_{\odot}$ . These values are comparable to those of the Milky Way, for which  $M^{\text{vir}} \approx 10^{12} M_{\odot}$  and  $M \approx 5 \times 10^{10} M_{\odot}$  (Bland-Hawthorn & Gerhard 2016). In the cosmological simulation, the first 5.8 Gyr ( $8 < \tau < 13.8$  Gyr) of m12f were characterized by an initial period of gas infall and several minor accretion events into its primordial dark-matter halo. This was followed by a brief quiescent period ( $6.5 < \tau < 8$  Gyr), which allowed the galaxy to begin forming the disk before experiencing a major collision with another galaxy ( $\tau \approx 6.5$  Gyr). This event is dubbed the "first major accretion event". Around 5 Gyr later, m12f experienced what we call its "second major accretion event" ( $\tau \approx 1.4$  Gyr). Both these major accretion events contributed metal-poor gas,



**Fig. 1.** Chemical abundance scatter plot for the redshift  $z = 0$  snapshot of the m12f synthetic galaxy. The red curve traces the average  $[\alpha/\text{Fe}]$  and  $[\text{Fe}/\text{H}]$  positions in equally spaced intervals of  $\Delta\tau = 100$  Myr. The tendency is to evolve from the top left to the bottom right, i.e., from high  $[\alpha/\text{Fe}]$  and low  $[\text{Fe}/\text{H}]$  to low  $[\alpha/\text{Fe}]$  and high  $[\text{Fe}/\text{H}]$ .

which mixed with the in situ gas, leading to episodes of elevated star formation with distinct chemical signatures. This is due to the different chemical enrichment histories of their progenitor nebulae, consistent expectations following the *Gaia*-Enceladus merger (see e.g., Vincenzo et al. 2019; Giribaldi & Smiljanic 2023; Ciucă et al. 2024). This effect is shown in Fig. 1 and confirmed in Fig. 2, where the distance of formation ( $d_{\text{form}}$ ) of stellar particles is plotted. The rebound at  $\tau \approx 3$  Gyr in Fig. 2 does not constitute an accretion event, but rather a flyby of the galaxy in which it ventured close to m12f, though not close enough to cause a collision. Following the prescription of Nikakhtar et al. (2021), we define stars with  $d_{\text{form}} > 30$  kpc as accreted. We also observe a very narrow chemical dichotomy at high  $[\text{Fe}/\text{H}]$  with a  $\Delta[\alpha/\text{Fe}]$  separation of about 0.02 dex (see Fig. 1). This is much smaller than the chemical separation between the Milky Way’s low- $\alpha$  and high- $\alpha$  disks, which is about 0.2 dex (Vincenzo et al. 2021).

The local standard of rest m12f-1sr0 is located at a vertical distance from the galactic plane of  $Z_{\text{Gal}} = 0$  pc and at a distance  $R_{\text{Gal}} = 8.2$  kpc from the galactic center, similar to estimates of the solar position relative to the center of the Milky Way (for a comprehensive literature review, see Bland-Hawthorn & Gerhard 2016, chap. 3.2). Its azimuthal velocity around the galactic center is  $V_{\text{LSR}} = 226$  km s $^{-1}$ , consistent with its neighborhood and comparable to the azimuthal velocity of the solar system’s local standard of rest in the Milky Way (see Bland-Hawthorn & Gerhard 2016, chaps. 3.4 and 6.4). The Ananke framework simulates a *Gaia* DR3-like survey, accounting for extinction from gas and the uncertainties of *Gaia* DR3, resulting in 4,265,816,043 observable stars, of which 138,536,719 have radial velocities (for more details, see Nguyen et al. 2024). Rather than using the entire survey, we will instead construct realistic observable samples, both in terms of the number of stars and selection function, as described in Sect. 2.2.

We now define the 5-dimensional input chrono-chemo-kinematic parameter space that will be adopted in this study, which includes the following stellar parameters:

- The stellar age,  $\tau$ , in Gyr.
- Metallicity,  $[\text{Fe}/\text{H}]$ .
- The abundance ratio of  $\alpha$ -process elements over iron,  $[\alpha/\text{Fe}]$ , which is defined in FIRE-2 as  $[\text{Mg}/\text{Fe}]$ .

- The azimuthal velocity of a star around the galactic center,  $V$ , in km s $^{-1}$ .
- The orthogonal cylindrical velocity components to the azimuthal velocity,  $U$  and  $W$ , combined into  $\sqrt{U^2 + W^2}$ , in km s $^{-1}$ .

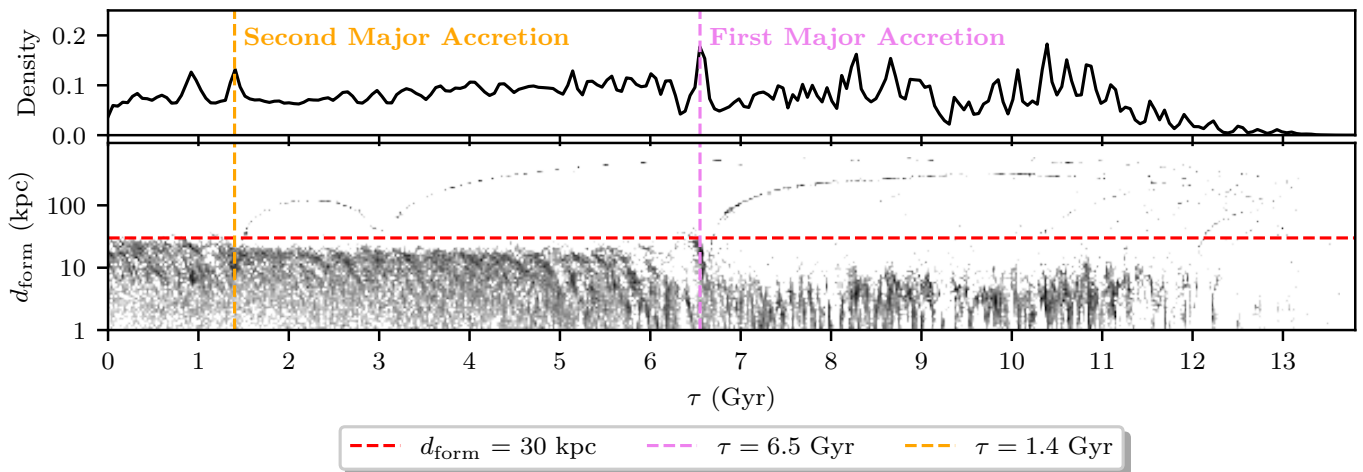
The stellar age and abundance ratios are provided by the simulations. We use  $[\text{Fe}/\text{H}]$  and  $[\alpha/\text{Fe}]$ , as these respectively trace type Ia supernovae (Ia SNe) and core-collapse supernovae (CC-SNe), serving as important galactic clocks (Arcones & Thielemann 2023). We do not consider the remaining elemental abundances available in the simulation. The seismic stellar samples (samples B, C and D; see Sect. 2.2) contain cool stars that do not possess sufficiently hot photospheric temperatures to excite helium. The surface  $[\text{C}/\text{N}]$  ratio is altered by processes occurring during stellar evolution (e.g., the first dredge-up) and hence does not qualify as a galactic clock. The remaining elements are  $\alpha$ -process elements, whose abundance ratios are highly (linearly) correlated with  $[\text{Mg}/\text{Fe}]$  (Pearson correlation coefficient  $\rho > 0.8$ ). For this reason, we opted to keep  $[\text{Mg}/\text{Fe}]$  as the only proxy for  $[\alpha/\text{Fe}]$ . Some FIRE-2 simulations do include r-process elements (van de Voort et al. 2015), but this is not the case for the Milky Way-like galaxies present in the *Latte* suite. Furthermore, to calculate the stellar velocities in a left-handed cylindrical coordinate system,  $\{U, V, W\}$ , we make use of the *Astropy* (Astropy Collaboration et al. 2022) package, using as input parameters the parallax, angular position, radial velocity and proper motions of the stars, as obtained from the Ananke *Gaia* DR3-like survey. The LSR is set at a distance of  $R_{\text{Gal}} = 8.2$  kpc from the galactic center and perfectly centered in the galactic plane ( $Z_{\text{Gal}} = 0$  pc).

## 2.2. Synthetic stellar samples

We generated four different *Gaia* DR3-like synthetic stellar samples from the m12f-1sr0 data set: samples A, B, C and D. Sample A will serve as the ground truth reference, being the most diverse sample, with 75,000 randomly selected stars in all evolutionary states within 3 kpc of the LSR observer. We select stars that have available radial velocities, that lie in the apparent *Gaia*-band magnitude range  $6 < G < 13$ , and further match the *Gaia* precision criteria in Gaia Collaboration et al. (2018a):

- Fractional parallax uncertainty  $< 0.1$ .
- Uncertainty in the *Gaia*  $G$  band  $< 0.022$  mag.
- Uncertainty in the *Gaia*  $G_{\text{BP}}$  band  $< 0.054$  mag.
- Uncertainty in the *Gaia*  $G_{\text{RP}}$  band  $< 0.054$  mag.
- Extinction  $A_G < 0.015$  mag.

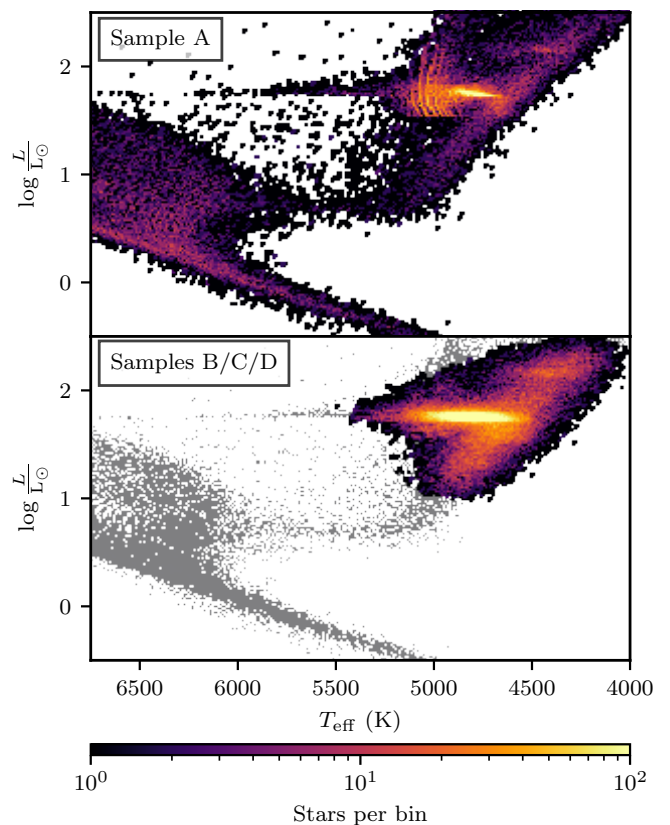
The remaining stellar samples were chosen to mimic red-giant stars, constrained by asteroseismic data from TESS. Red-giant stars are ubiquitous, visible at large distances due to their high luminosities, and they further possess deep convective envelopes that allow high-SNR detection of oscillation modes. These properties make them ideal targets for probing the Galactic structure across space and time. The TESS mission enabled the creation of a seismic red-giant star catalog containing the frequency of maximum oscillation power ( $\nu_{\text{max}}$ ) and asteroseismic mass estimates for 158,505 red giants, spanning a large sky area (Hon et al. 2021). Following the recommendations in Hon et al. (2021), we preserve only the catalog stars with  $\text{mass\_flag} = 1$  and  $\text{ruwe} \leq 1.40$ , where the former condition implies  $0.6 \leq M \leq 2.9 M_{\odot}$ , and the latter indicates an adopted astrometric solution likely to correspond to a single star. We further select stars with relative uncertainties below 10% in effective temperature ( $T_{\text{eff}}$ ),



**Fig. 2.** Redshift  $z = 0$  snapshot of the m12f synthetic galaxy considering only stellar particles located within 30 kpc from the galactic center at that particular instant in time. **Top:** Gaussian kernel density estimate for the age distribution of the m12f stellar particles. Spikes are noticeable at ages 6.5 and 1.4 Gyr. **Bottom:** Scatter plot of  $d_{\text{form}}$  as a function of age. Accreted galaxies are shown as curves of continuously generating stellar particles approaching from distances  $d > 100$  kpc, eventually colliding with m12f. These collisions coincide with spikes in star formation.

luminosity ( $L$ ), distance ( $d$ ), and the *Gaia*-band magnitude ( $G$ ). This left us with 137,404 seismic red-giant stars spanning almost the entire sky and covering a distance distribution with a mode of 800 pc.

To serve as a ground truth TESS-like sample, we generated sample B. Besides the same requirements as for sample A, the selection function of sample B is chosen to mimic the evolutionary states and physical parameters of the TESS seismic red-giant catalog. Firstly, to blur the otherwise discrete evolutionary track gridpoints in the Hertzsprung–Russell (HR) diagram, the  $T_{\text{eff}}$  and  $L$  values are perturbed according to Gaussian distributions with mean,  $\mu$ , equal to their true value and standard deviation,  $\sigma$ , equal to the median relative uncertainties found in the TESS seismic red-giant catalog (0.02 for  $T_{\text{eff}}$  and 0.04 for  $L$ ). Secondly, we replicated the TESS selection function by using kernel density estimators (KDE) and k-dimensional (KD) tree nearest neighbor algorithms. For the stellar mass,  $M$ , we constructed a 1-dimensional KDE and sampled stars to approximate the seismic TESS catalog’s mass distribution. We then wanted to perform independent 2-dimensional samplings of the  $T_{\text{eff}}-\nu_{\text{max}}$  diagram, the right-ascension and declination sky coverage, RA–DEC, and of the distance from the observer with the *Gaia* band,  $d-G$ . For computational reasons, we forego the use of KDEs in the 2-dimensional case, instead opting for the faster KD tree algorithm. This algorithm uses binary space partitions to efficiently find nearest neighbors of query points, allowing us to sample points within a volume of interest. We considered the sky position in the crossmatch because TESS observations are biased towards its southern and northern continuous viewing zones and the fact that TESS misses observations for parts of the sky. The combination of  $T_{\text{eff}}$ ,  $\nu_{\text{max}}$  (calculated via the scaling relation provided in [Stello et al. 2008](#)) and  $M$  ensures that the appropriate evolutionary states are selected. We discard a crossmatch in luminosity to avoid introducing bias due to reddening effects. The resulting crossmatch contains 76,168 stars, about half the size of the reference catalog. Figure 3 compares the parameter spaces of samples A and B, with the latter being shared with samples C and D, which we will now go on to describe. We provide a comparison of samples A, B and the TESS seismic red-giant catalog in Fig. B.1 of the Appendix.



**Fig. 3.** HR diagram of the synthetic stellar samples showing the main sequence, subgiant phase, red-giant branch and the red clump. **Top:** Sample A. **Bottom:** Samples B, C and D (color-coded) overlapping the sample A stars (in gray).

Finally, we generated samples C and D from sample B such that their chemical abundances and stellar ages are perturbed from their true values, with the kinematics being adopted from the error-convolved *Gaia* DR3-like values. The differences between samples C and D are as follows. Sample C will contain

stellar age and chemical abundance perturbations based on realistic observational uncertainties, whereas sample D will have enhanced-precision (i.e., optimistic) perturbations. Firstly, we calculate the stellar velocities in a left-handed cylindrical coordinate system,  $\{U, V, W\}$ , using the error-convolved parallax, angular position, radial velocity, and proper motions. These kinematics are shared by both samples. Secondly, we perturb the values for age by 20% for sample C (in line with typical asteroseismic stellar dating estimates for red-giant stars) and by 10% for sample D. Finally, we perturb the values of  $[\text{Fe}/\text{H}]$  and  $[\alpha/\text{Fe}]$  according to the median internal uncertainties of APOGEE DR17 for sample C (0.0095 and 0.0151 dex, respectively) and the 25th percentile uncertainties for sample D (0.0076 and 0.0123 dex, respectively). We note that a conservative uncertainty on the chemical abundances would be on the order of 0.1 dex, hence larger than the  $\sim 0.01$  dex adopted here. This reasoning comes from the fact that the chemical separation between the high- $\alpha$  and low- $\alpha$  disks is significantly smaller for m12f than observed for the Milky Way (see Fig. 1). As a result, a perturbation of 0.1 dex would cause a level of mixing in chemical space that would make the disks indistinguishable. We provide a comparison of samples B, C, and D in Fig. B.2 of the Appendix.

### 3. Methods

The identification and differentiation of stellar populations is a problem of stellar group membership assignment that considers commonalities in physical parameter space and plausible theories to explain their origin. Clustering algorithms such as K-Means (MacQueen 1967; Ikotun et al. 2023), DBSCAN (Ester et al. 1996), and HDBSCAN (Campello et al. 2013), as well as generative models such as Gaussian Mixture Models (GMM; Reynolds 2009), may provide a first step towards the assignment of group membership, however, it is crucial to consider their limitations. Generative model algorithms like GMM make parametric assumptions on the distributions that generate the data, which may pose a problem if the data are distributed according to a different parametric distribution or if the data do not follow any parametric distribution. Clustering algorithms do not make parametric assumptions about the data. However, there are nuances which may lead to some of them being more appropriate than others depending on the circumstances. Even when taking such nuances into account, the validity of the resulting clusters may be difficult to assess, particularly for large datasets where computational memory becomes a limiting factor. Furthermore, it can be shown that applying clustering algorithms to high-dimensional data may be suboptimal compared to clustering in lower-dimensional spaces, provided these lower-dimensional spaces are constructed to emphasize potentially meaningful local structure in the data. An example using HDBSCAN is provided in Appendix D. Given these considerations, we were motivated to pursue a strategy of dimensionality reduction.

Dimensionality reduction is the transformation of high-dimensional input data into a lower-dimensional output space. A popular, yet limited, unsupervised learning algorithm used for this purpose is Principal Component Analysis (PCA; Pearson 1901; Francis & Wills 1999), which identifies the principal components of the input dataset, defined as the eigenvectors of the covariance matrix. An  $N$ -dimensional input space results in  $N$  principal components, from which the top  $n < N$  are selected in order to reduce the dimensionality of the data. This method, popular due to its speed and deterministic nature, is limited to linear transformations of the data. Furthermore, feature importance in PCA is defined by a global measure of variance, which

is sensitive to noise and fails to capture structural nuance at a local level. For complex nonlinear cases, a more general approach is required.

Our methodology of choice must be capable of generalizing to non-linear and non-parametrically distributed data, in a manner that allows us to confidently assess the validity of the clusters found. One such algorithm is t-SNE (van der Maaten 2013), which works by determining the neighborhood relationships of the high-dimensional input data and then projecting them into a low-dimensional output space, known as a latent space embedding. This method is well-suited for nonlinear, nonparametrically distributed datasets and provides a visualization that helps make more confident decisions on how to segregate the data. It is important to note that the projection through the construction of a latent space embedding operates differently from standard projections, such as PCA. Whereas PCA is a deterministic algorithm limited to linear transformations that uses global variance as a metric for feature importance, latent space embedding iteratively evolves the dimensionally reduced space by prioritizing local neighborhood preservation. It is further capable of handling local non-linear relationships, emphasizing the presence of complex data relationships that can then be more easily clustered. This is the strategy we ultimately adopt, though using a different algorithm: UMAP.

We justify the choice of UMAP in the following subsection. A comparison of UMAP's effectiveness with PCA and t-SNE is provided in Appendix D.

#### 3.1. UMAP

The Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP; McInnes et al. 2018) algorithm is a method that combines aspects of manifold learning and topological data analysis for dimensionality reduction. Firstly, UMAP constructs a fuzzy topological representation of the input high-dimensional data, wherein data points are connected by weighted edges based on a similarity criterion that is quantified using a probability distribution that measures the likelihood of points being connected. Secondly, UMAP captures neighborhood relationships using a local approach to preserve the local structure of the data. Finally, a lower-dimensional latent space embedding is constructed by minimizing a cost function that balances the preservation of local neighborhood relationships and the overall global structure. This low-dimensional embedding is subsequently iteratively adjusted to achieve a configuration that best aligns with the underlying manifold. The final output is an embedding that contains information on both the local and global patterns and commonalities found in the data, depending on the choice of hyperparameters. The most important hyperparameters are `n_neighbors` and `min_dist`. The former constrains the size of the local neighborhood analyzed by UMAP when attempting to learn the manifold structure of the data, and the latter controls how tightly UMAP aggregates points together in the latent space embedding. Tuning these hyperparameters will determine whether the resulting embedding emphasizes local or global structure. An exaggerated emphasis on the finer structure can lead to noisy results, whereas emphasis on macroscopic structure can instead result in an oversmoothed, non-descriptive embedding. For these reasons, it is important to choose these hyperparameters carefully to balance out these effects.

In this paper, we applied the UMAP algorithm using the Python UMAP library on the synthetic samples A, B, C and D generated in Sect. 2.2. The input was input a 5-dimensional chrono-chemo-kinematic parameter space spanning the vari-

ables  $\{V, UW, [\text{Fe}/\text{H}], [\alpha/\text{Fe}], \tau\}$ , where  $UW = \sqrt{U^2 + W^2}$ . Since these variables span significantly different ranges of magnitudes, which may influence the algorithm’s output, each variable is scaled using the `StandardScaler` function in `Sklearn` (Pedregosa et al. 2011). Our choice of UMAP hyperparameters aimed to balance local structure with global structure. For this purpose, we varied `min_dist` and `n_neighbors` for sample A and visually inspected the resulting embedding, settling in the end on `min_dist = 0` and `n_neighbors = 15`. We found that meaningful embeddings were obtained after 1000 epochs. However, we conservatively set the number of epochs to 10,000 in all UMAP runs to ensure that the result has adequately approximated its final form. For simplicity, we use an Euclidean metric. Moreover, we reduce the 5D input data to a 2D representation for ease of visual inspection. An objective choice of the number of dimensions for the latent space embedding could be explored by calculating the intrinsic dimension, especially for input parameter spaces with a higher number of dimensions. Since ours has only 5 dimensions, we do not perform any such measure of intrinsic dimensionality.

### 3.2. Trustworthiness

An objective measure of the validity of an embedding is the trustworthiness score,  $T$  (Venna & Kaski 2001), which depends on the number of nearest neighbors,  $k$ . Note that  $k$  should not be confused with the `n_neighbors` hyperparameter. The trustworthiness score validates how closely the low-dimension embedding resembles the high-dimension input in terms of the preservation of neighborhood. In other words, an embedding is trustworthy if stars that are neighbors in the input parameter space remain neighbors in the embedding. The trustworthiness score is defined as:

$$T(k) = 1 - \frac{2}{Nk(2N - 3k - 1)} \sum_{i=1}^N \sum_{x_j \in U_k(x_i)} (r(x_i, x_j) - k), \quad (1)$$

where  $N$  is the number of stars,  $x_i$  is the coordinate of the star  $i$ , and  $U_k(x_i)$  is the set of  $k$  nearest neighbors of  $x_i$  in the embedding space that are not part of the  $k$  nearest neighbors of  $x_i$  in the input space. The term  $r(x_i, x_j) \in \mathbb{N}$  ranks the proximity of each coordinate  $x_j \in U_k(x_i)$  relative to the coordinate of the star,  $x_i$ , as measured in the input space. Thus,  $r(x_i, x_j) > k \forall x_j \in U_k(x_i)$ , meaning the term in the summation will be positive. If the  $k$  nearest neighbors of  $x_i$  in the embedding space are all within the  $k$  nearest neighbors of  $x_i$  in the input space, then the set  $U_k(x_i)$  is empty for that  $i$  and the step is skipped.

Equation (1) consists of two terms. The first term represents a perfect score of 1, while the second term is a penalty term, checking whether the neighborhood has been preserved for each star in the sample. If it has not, then it will decrease the score in proportion to the severity of the discrepancy, as determined by the rank term  $r(x_i, x_j)$ . This second term is normalized to constrain the trustworthiness score in the range  $T(k) \in [0, 1]$ , with values near 0 indicating heavy penalization. We calculate  $r(x_i, x_j)$  using the same metric as was used in UMAP. Since  $T$  varies as a function of the number of nearest neighbors,  $k$ , similar to the UMAP hyperparameter `n_neighbors`, we set  $k = \text{n\_neighbors}$ . We adopt the conservative heuristic that an embedding is considered trustworthy if the trustworthiness score is  $T(k = \text{n\_neighbors}) > 0.90$ .

Once the embedding passes the trustworthiness test, the next step is to select the learned structures. We look for evidence of differentiated structures, by which we mean topographically

interesting components that can be clearly distinguished from one another. For example, internally coherent components that are disjoint from one another may indicate different stellar populations with significantly different physical parameters. Conversely, an embedding is considered uninformative if it lacks topographically interesting features. Uninteresting features may include spherical shapes and homogeneity. Since these structures can vary in shape and size, we use an unsupervised clustering algorithm to objectively identify and select structures.

### 3.3. HDBSCAN

Though several unsupervised clustering algorithms exist, all of them possess their own specific limitations. To give two examples, the popular K-Means algorithm, while efficient, tends to produce clusters of similar size and lacks sensitivity to topographic features. On the other hand, Gaussian Mixture Models (GMM) are limited by the assumption that clusters follow Gaussian distributions. Both of these algorithms are unsuited for clustering the structures in the UMAP embeddings, since they may come in various shapes and sizes. Instead, we turn our interest to flexible, non-parametric clustering algorithms. One such algorithm is DBSCAN, which works as follows. Firstly, it selects a point and counts the number of neighbors within a radius given by the `eps` hyperparameter. Secondly, it checks whether the number of neighbors is equal to or greater than the `min_samples` hyperparameter. If true, the point is considered a core point, otherwise it is labeled as an outlier. Thirdly, core points that are within close proximity to each other are grouped into the same cluster. Through this method, DBSCAN is capable of capturing clusters of arbitrary shapes and sizes, mediated only by the density of points, with no prior parametric assumption. Additionally, DBSCAN does not require the user to specify the number of clusters, which is a mandatory hyperparameter in algorithms like K-Means and GMM.

One limitation that is not addressed by DBSCAN is a measure of the robustness of its clusters. In other words, we do not know how sensitive the results are to minor variations in the hyperparameters. This problem is addressed by HDBSCAN, which distinguishes itself from DBSCAN by replacing the `eps` distance with the mutual reachability distance. Whereas `eps` is a fixed distance and thus unable to capture clusters of different densities, the mutual reachability distance varies based on the local density information and can therefore identify clusters of varying densities. Firstly, HDBSCAN calculates this mutual reachability distance between all points. Secondly, the minimum spanning tree is created, where nodes represent points, and the edges are weighted according to the mutual reachability distance. Thirdly, the algorithm tracks the evolution of the clusters as the mutual reachability distance threshold increases. The parameter  $\lambda$ , defined as the reciprocal of the mutual reachability distance, is evaluated at each step. Clusters begin to group together, condensing the cluster tree until only a single cluster remains, encompassing all of the data. Throughout this process, some clusters will have persisted more robustly than others. This robustness is measured by the persistence score, defined as:

$$\text{Persistence} = \sum_{p \in \text{cluster}} (\lambda_p - \lambda_{\text{birth}}), \quad (2)$$

where  $\lambda_p$  represents the value of  $\lambda$  when the point is no longer connected to the cluster. This value is constrained between  $\lambda_{\text{birth}}$  and  $\lambda_{\text{death}}$ , which respectively define the values of  $\lambda$  when the cluster is born and dies.

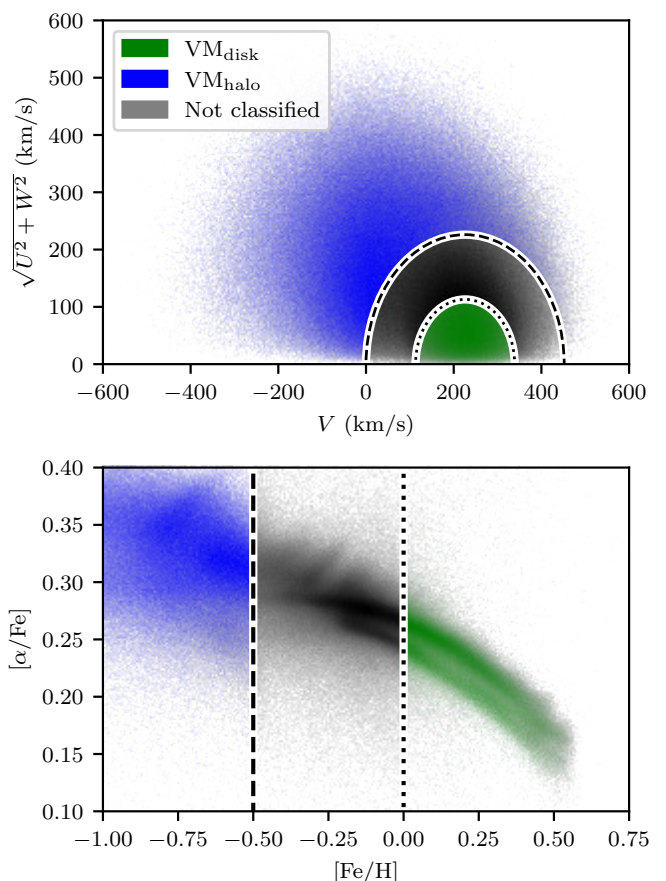
HDBSCAN automatically identifies the most persistent clusters, although these clusters also depend on HDBSCAN’s own hyperparameters, namely `min_cluster_size` and `min_samples`. To objectively select the optimal clusters, we conduct a hyperparameter optimization routine using the Python package `optuna`. Our loss function is the persistence score provided by the `hdbscan` package.

To mitigate the risk of clustering random noise, we must define a baseline persistence score to differentiate between statistically meaningful clusters and those that are not. We generated 1000 random 5-dimensional input spaces, each containing only noise, and computed their latent space embeddings. We then ran the HDBSCAN hyperparameter optimization procedure to obtain the optimal clusters. This gave us a distribution of persistence scores, and we define the baseline score as the 99.7th percentile ( $3\sigma$  equivalent) of this distribution. If the persistence scores from our samples exceed this baseline score, we regard the clusters as statistically meaningful; otherwise, we disregard them indistinguishable from random noise.

### 3.4. The VM criterion

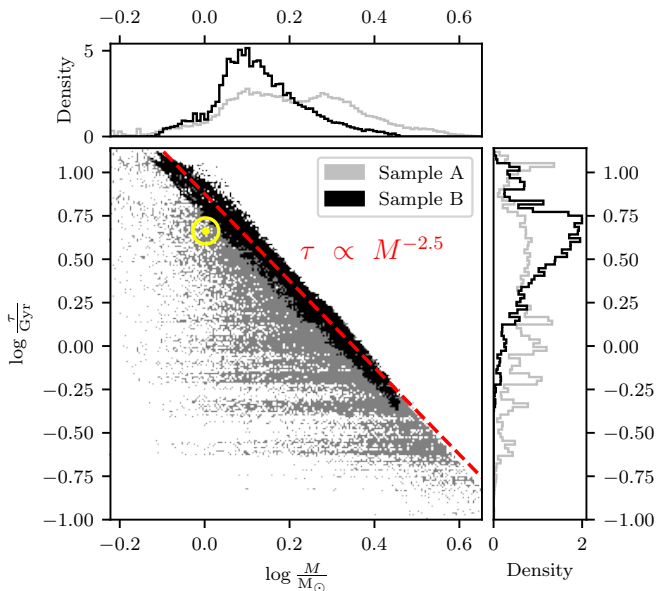
In order to validate the legitimacy of the embedding components as physically meaningful stellar populations, we study their physical parameters and seek to interpret them. Since the samples only probe the stars contained within a sphere of 3 kpc radius around the LSR, we expect our initial application of the embedding to distinguish between the disk and the halo, as the stars of each of these components distinguish themselves greatly in age, chemical content and kinematic profile. To achieve this, we employ a velocity and metallicity (VM) criterion, similar to the method used in [Ostdiek et al. \(2020\)](#). In our case, we want to define stars for which we have a high degree of confidence as being part of the halo, but not the disk, and vice versa. We call this  $VM_{\text{halo}}$  and  $VM_{\text{disk}}$ . We know that the azimuthal velocity for the LSR of `m12f-1sr0`, a disk star, is  $V_{\text{LSR}} = 226 \text{ km s}^{-1}$ , on par with the bulk of the stars in sample A. Let  $\boldsymbol{v}$  represent the 3D cylindrical velocity vector. Stars with  $|\boldsymbol{v} - \boldsymbol{V}_{\text{LSR}}| > |\boldsymbol{V}_{\text{LSR}}|$  and  $[\text{Fe}/\text{H}] < -0.50$  dex are defined as being part of the  $VM_{\text{halo}}$ , whereas stars with  $|\boldsymbol{v} - \boldsymbol{V}_{\text{LSR}}| < \frac{1}{2} \times |\boldsymbol{V}_{\text{LSR}}|$  and  $[\text{Fe}/\text{H}] > 0$  dex are defined as being part of the  $VM_{\text{disk}}$ , with everything in between being considered part of a transition zone that can include both halo and disk stars (see Fig. 4). These values were chosen based on an inspection of the physical parameters of the `m12f` galaxy. Figure A.3 shows the effects of this selection on the geometry of the corresponding stellar populations. This criterion is not intended to strictly define the halo nor the disk, but rather to distinguish two highly disparate populations that are strongly associated to the halo and the disk, while minimizing contamination.

We note that stars not classified as part of the  $VM_{\text{disk}}$  nor the  $VM_{\text{halo}}$  can still overlap with these components. For instance, they could assume disk-like kinematics but low metallicity. These “unclassified” stars thus span from halo stars to disk stars, and everything in between, including the parameter space in which both components of the  $[\alpha/\text{Fe}]$  dichotomy begin to take form. This selection ensures that the  $VM_{\text{disk}}$  and  $VM_{\text{halo}}$  are adequately separated in both chemo-kinematic space and the galaxy’s evolution, with the median age for the  $VM_{\text{disk}}$  being 2.4 Gyr, whereas the median age for the  $VM_{\text{halo}}$  is 10.4 Gyr. The percentage of stars contained in  $VM_{\text{disk}}$ ,  $VM_{\text{halo}}$  and  $VM_{\text{Not Classified}}$  for samples A, B, C and D is given in Table 1. We note that, by design, only a minority of the stars are in either the  $VM_{\text{disk}}$  or  $VM_{\text{halo}}$ . We also note that samples B, C,



**Fig. 4.** Kinematic and chemical parameters of the stellar particles in `m12f`. Dashed lines:  $|\boldsymbol{v} - \boldsymbol{V}_{\text{LSR}}| = |\boldsymbol{V}_{\text{LSR}}|$  and  $[\text{Fe}/\text{H}] = -0.50$  dex. Dotted lines:  $|\boldsymbol{v} - \boldsymbol{V}_{\text{LSR}}| = \frac{1}{2} \times |\boldsymbol{V}_{\text{LSR}}|$  and  $[\text{Fe}/\text{H}] = 0$  dex. Green represents the  $VM_{\text{disk}}$ , blue represents the  $VM_{\text{halo}}$ , and gray represents neither. **Top:** Toomre diagram. **Bottom:**  $[\alpha/\text{Fe}]$ – $[\text{Fe}/\text{H}]$  chemical plane.

and D are undersampled in both the  $VM_{\text{disk}}$  and  $VM_{\text{halo}}$  relative to sample A, which could affect the ability to resolve both the disk and the halo. This effect originates from the selection function applied to the seismic red-giant stars. The inferred age of red-giant stars is primarily determined by the time spent in the main sequence, which follows a scaling relation of  $\tau \propto M^{1-\eta}$ , where  $\eta$  is the exponent in the  $L$ – $M$  relation for main-sequence stars ([Kippenhahn et al. 2013](#); [Davies & Miglio 2016](#)). This results in a tight age–initial mass relation for red-giant stars, which will affect the age distribution of our sample. This effect is evident in Fig. 5, which compares the age and mass distributions of samples A and B. Sample A contains main-sequence stars that span a wide variety of masses and ages, whereas sample B (and, by extension, samples C and D) is constrained to masses  $0.6 < M < 2.9 M_{\odot}$  with a peak at  $M \approx 1.2 M_{\odot}$ , resulting in a significant undersampling of stars younger than  $\tau \approx 3$  Gyr.



**Fig. 5.** Age-mass distribution for sample A (silver) and sample B (black), showing the tight relationship for the seismic red-giant stars (sample B). The location of the Sun is identified by its usual symbol ( $\odot$ ), assuming an age of 4.6 Gyr. The red dashed line follows the  $\tau \propto M^{1-\eta}$  scaling, where  $\eta = 3.5$  (cf. Kippenhahn et al. 2013; Davies & Miglio 2016).

**Table 1.** Fraction of stars per VM component in each sample.

Sample	VM <sub>disk</sub> (%)	VM <sub>halo</sub> (%)	VM <sub>Not Classified</sub> (%)
A	24.4	6.3	69.3
B	15.1	3.9	81.0
C	15.2	3.9	80.9
D	15.2	3.9	81.0

## 4. Results

In this section, we perform two consecutive applications of UMAP on samples A, B, C and D. The first application of UMAP, henceforth referred to as the first run, generates the latent space embedding for the entire sample. If physically meaningful components are found within these embeddings, then we perform another application of UMAP, this time on each of the components, in an attempt to subdivide them into smaller constituent subcomponents. This is referred to as the second run.

### 4.1. First run

We applied the UMAP algorithm to samples A, B, C, and D, and for each sample we obtained 2D latent space embeddings. Figure 6 illustrates these embeddings, color-coded according to the distributions of their input parameters. These parameters do not follow a strictly monotonic gradient, especially for the kinematic parameters. This information is encoded in the assumed topology of the embeddings. We notice that old, low-metallicity stars tend to agglomerate together and self-segregate from the remaining stars. Despite this, there is a clear diversity in the chemical abundances and kinematics, being the sole region where we

identify retrograde stars ( $V < 0 \text{ km s}^{-1}$ ). To determine the validity of these embeddings, we compute their trustworthiness scores,  $T(k)$ , with  $k = \text{n\_neighbors}$  (see Sect. 3 for details). We find that the trustworthiness scores for samples A, B, C, and D are 0.988, 0.994, 0.969, and 0.978, respectively. Given that the embeddings satisfy  $T(k = \text{n\_neighbors}) > 0.90$  in all cases, we consider these embeddings to be informative and proceed with our analysis.

Figure 7 shows how there is a pronounced preference for ex situ stars (shown in red) to agglomerate within the aforementioned structure. This structure is fully disjoint for sample A. While not the case for samples B, C and D, there are nonetheless indications for the existence of a non-unified structure that appears to assume at least two distinct groups. We applied the HDBSCAN algorithm on these embeddings with a hyperparameter optimization routine, aiming to maximize the persistence score. We obtained two major components for all samples, denoted as the major component 1 (MC1) and the major component 2 (MC2). MC1 corresponds to the major component with the highest number of stars, and MC2 to the one with the lowest, which we color in green and blue, respectively. It is also worth noting that the vast majority of the accreted stars (defined as stars with  $d_{\text{form}} > 30 \text{ kpc}$ ) are located in MC2, with 97.6% to 98.7% of them found in this component, depending on the sample. At no point was  $d_{\text{form}}$ , the parameter used to define ex situ stars, included in the generation of the embeddings. The fraction of stars in MC2 classified as accreted ranges from 10.9% to 16.6%, depending on the sample.

The next step is to determine if the major components MC1 and MC2 correspond to physically meaningful stellar populations. In particular, we aim to verify if our assumption of disk and halo segregation is correct. If this assumption is correct, we would expect that MC1 and MC2 contain the disk and the halo separately, as defined by the VM criterion, i.e., VM<sub>disk</sub> and VM<sub>halo</sub> (see Sect. 3). To assess this, we construct a  $3 \times 3$  confusion matrix, with truth values defined as VM<sub>disk</sub>, VM<sub>halo</sub> and VM<sub>Not Classified</sub>. The predicted values are defined as MC1, MC2 and “other”, where “other” refers to outlier stars that were not assigned to any cluster by HDBSCAN. Since there is no predicted class analog for the truth class VM<sub>Not Classified</sub>, we exclude this from our evaluation metrics. We do not include, however, the predicted class “other”. For example, stars with truth value VM<sub>disk</sub> assigned a prediction label other than MC1 (i.e., MC2 or “other”) are defined as false negatives. The reverse situation is also true. Due to the size discrepancy between MC1 and MC2, the predictive power of MC1 dominates and skews the results. To mitigate this, we calculate the specificity and sensitivity, which, for a given truth class  $C_i$  (either VM<sub>disk</sub> or VM<sub>halo</sub>), are defined as:

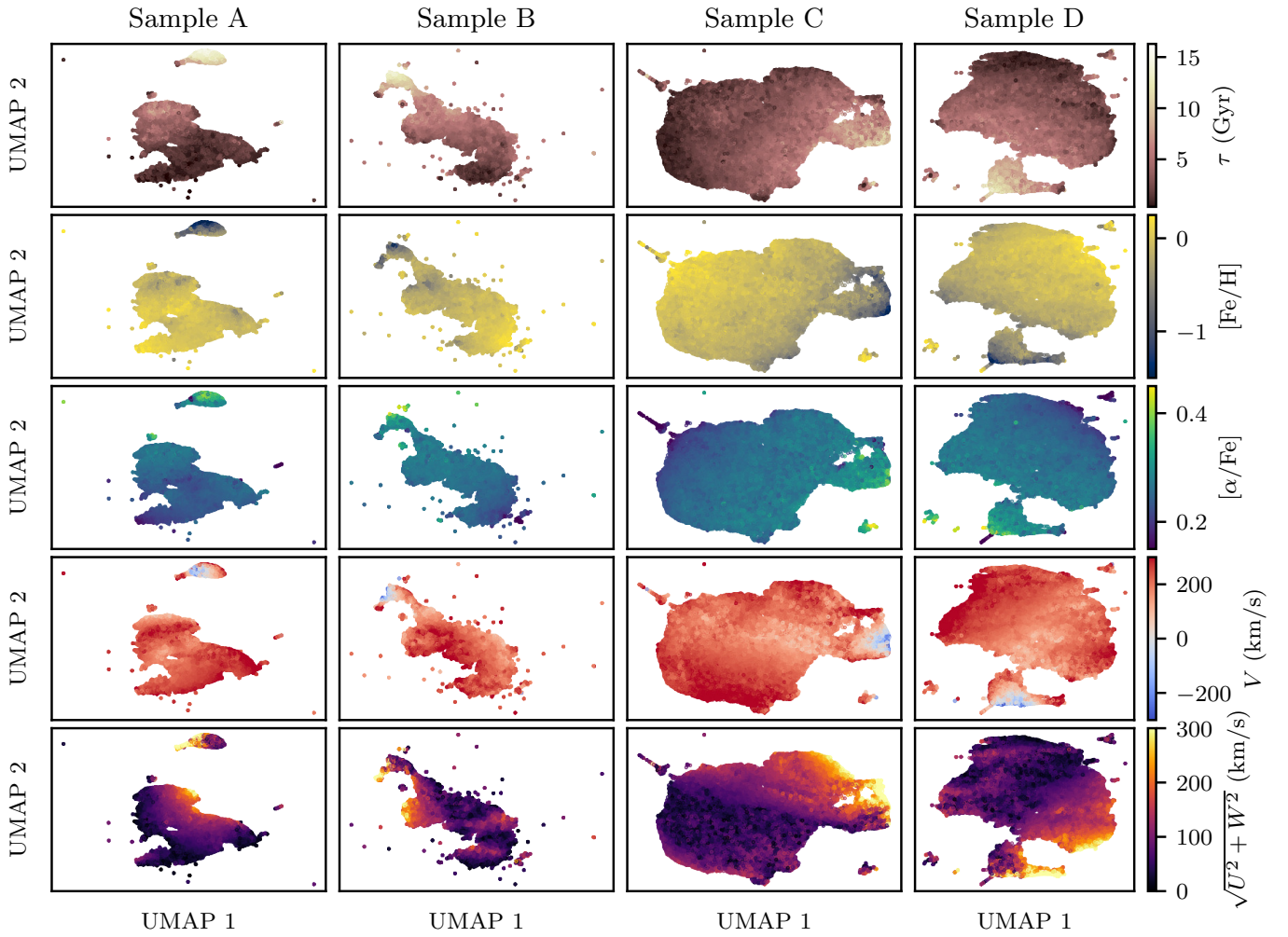
$$\text{Sensitivity}_{C_i} = \frac{\text{TP}_{C_i}}{\text{TP}_{C_i} + \text{FN}_{C_i}}, \quad (3)$$

$$\text{Specificity}_{C_i} = \frac{\text{TN}_{C_i}}{\text{TN}_{C_i} + \text{FP}_{C_i}}, \quad (4)$$

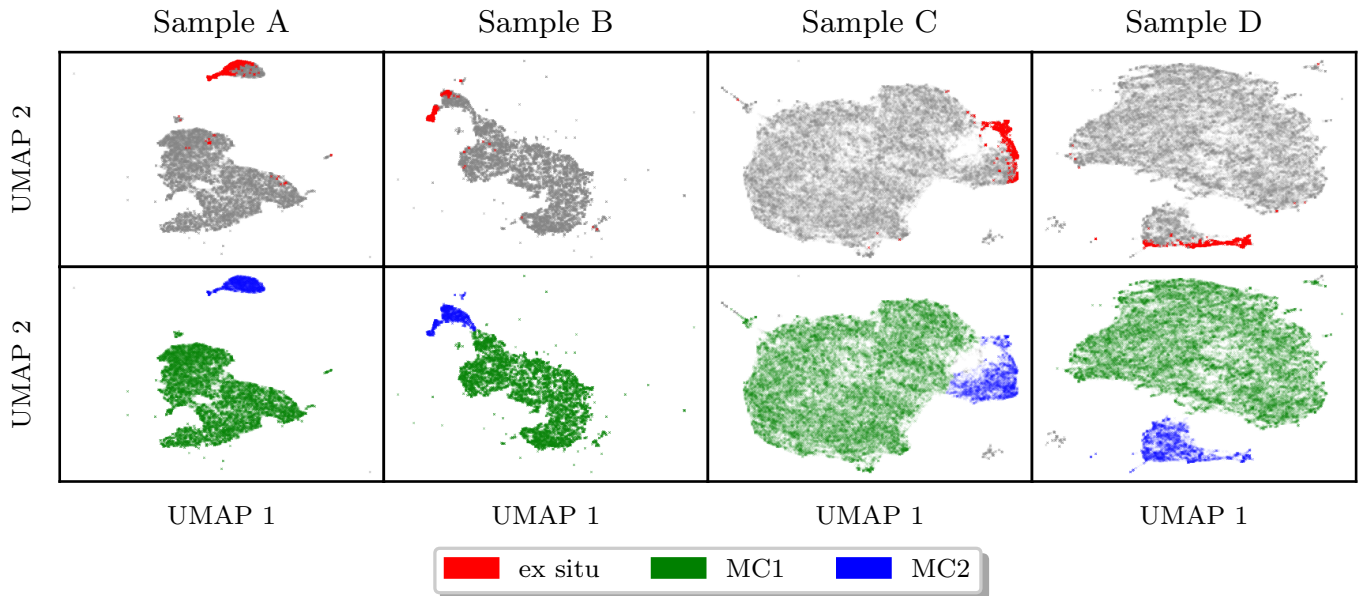
where TP, TN, FP and FN correspond to true positive, true negative, false positive and false negative, respectively. This allows us calculate the informedness score for each truth class:

$$\text{Informedness}_{C_i} = \text{Specificity}_{C_i} + \text{Sensitivity}_{C_i} - 1, \quad (5)$$

which can then be combined in an arithmetic mean to provide the total informedness. The total informedness ranges from  $-1$  to  $1$ , with values close to  $1$  indicating a good classification, values close to  $-1$  suggesting that the labels have been inverted, and values close to  $0$  indicating a poor classification.



**Fig. 6.** First run, 2D UMAP latent space embeddings for samples A, B, C and D, color-coded according to the input parameters. The specific orientation of the embeddings is not important, since there is no physical meaning to the coordinates of the latent space. What matters is the relative position of the points with respect to each other. As a result, the embeddings may appear flipped from one sample to another.



**Fig. 7. Top:** First run, 2D UMAP latent space embeddings for samples A, B, C and D. Red dots correspond to ex situ stars ( $d_{\text{form}} > 30$  kpc). **Bottom:** Selection of the major components 1 and 2 (MC1 and MC2) in green and blue, respectively.

**Table 2.** Scoring metrics for samples A, B, C and D.

Sample	$T(k = n\_neighbors)$	Informedness
A	0.988	0.989
B	0.994	0.966
C	0.969	0.966
D	0.978	0.973

**Notes.**  $T(k)$  is the trustworthiness score for  $k$  nearest neighbors. The informedness score is the arithmetic mean of the informedness scores for the disk and halo.

We find the informedness scores for samples A, B, C, and D to be, respectively, 0.989, 0.966, 0.966, and 0.973. These values indicate that MC1 and MC2 contain, respectively, the disk and the halo, with little cross-contamination. Our results are summarized in Table 2.

#### 4.2. Second run

A closer inspection of the latent space embeddings in Fig. 7 reveals further topographic features within MC1 and MC2. For samples A and B, MC1 appears to be either trying to dislodge or elongate itself, which is not the case for samples C and D. In all samples, we also notice that MC2 possesses a tail. These features hint at the existence of a more nuanced structure. Motivated by this, we performed a second run for MC1 and MC2, taken separately, for each of the samples. We ran UMAP 100 times, applying an HDBSCAN optimization routine to each embedding, and kept track of the persistence score and the number of clusters found. We retain the embeddings with the mode number of clusters and, from that pool, focus our attention on the embedding for which HDBSCAN returned the highest persistence score. These final embeddings, obtained for MC1 and MC2 in each of the samples, are displayed in Fig. 8. We note once again the tendency for ex situ stars to agglomerate together in the MC2 embeddings, potentially forming predominantly ex situ substructures. We show the corresponding distributions of the input parameters in Figs. C.1 and C.2 of the Appendix.

Similar to the first run, we compute the trustworthiness score  $T(k = n\_neighbors)$  for the second run MC1 and MC2 embeddings and compare them to the trustworthiness score of the first run embeddings. To assess the sensitivity of  $T$  as a function of  $k$ , we also compute the trustworthiness score in the vicinity of  $k = n\_neighbors$ . These results are plotted in Fig. 9, displaying a small downward trend as  $k$  is increased, but nonetheless satisfying  $T(k) > 0.90$  for all scenarios.

The HDBSCAN clustering of the MC1 and MC2 embeddings is shown in Fig. 10. While the algorithm initially converged to two clusters for MC1 from samples C and D, the persistence score did not surpass the baseline required to confidently assert them as statistically meaningful and were thus merged into one. This left us with two MC1 subcomponents for samples A and B, dubbed MC1a and MC1b, and just one for samples C and D, which we refer to as MC1a for consistency. As for MC2, HDBSCAN converged to three clusters for samples A and B, which we dub MC2a, MC2b, and MC2c. For samples C and D, we obtain only two subcomponents, which we name MC2a and MC2b for consistency. These subcomponents are named based

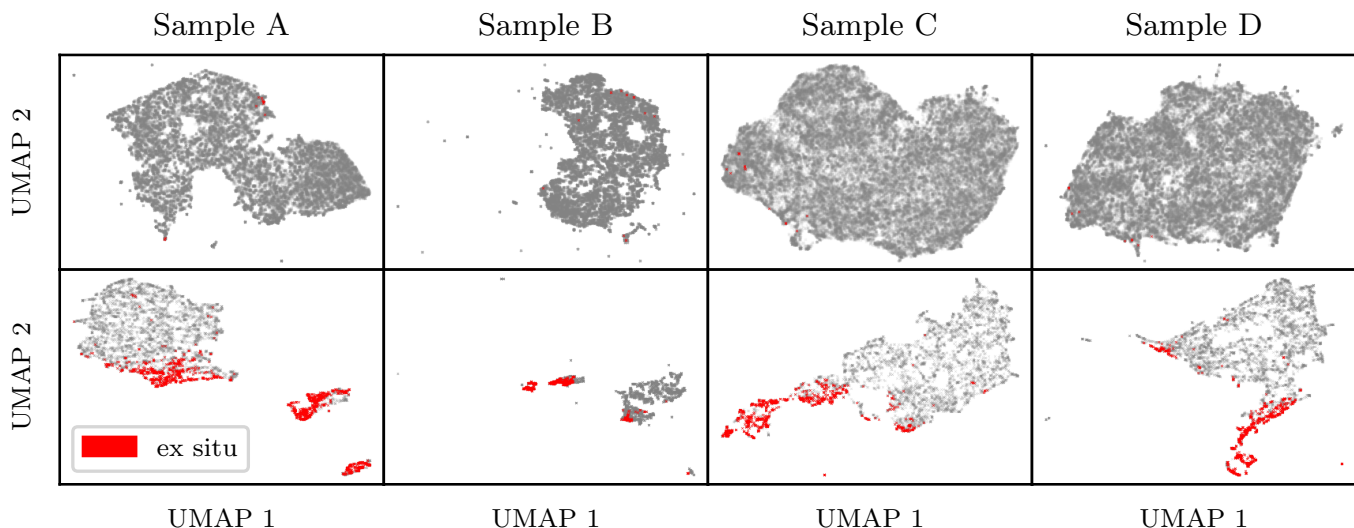
on the number of stars within them, and do not necessarily refer to the same population across samples.

#### 4.3. Interpreting the subcomponents

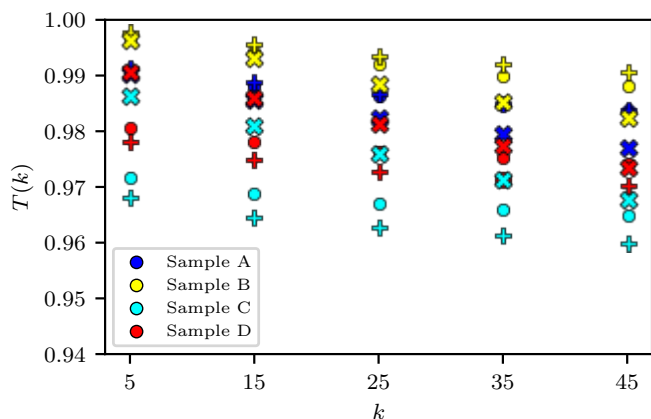
Whether the subcomponents found by HDBSCAN, applied to the UMAP latent space embeddings, correspond to genuinely distinct stellar populations remains to be seen. This would be straightforward if the stars had their population membership determined a priori. However, much like in a real scenario, we lack such direct labeling and must instead rely on inferences based on stellar parameters. Since these samples are based on synthetic data, we also have access to meta-parameters that can be used to validate our findings. One such useful meta-parameter is the distance of formation,  $d_{\text{form}}$ , which allows us to get a sense of the star formation history (SFH) and the merger history of the synthetic m12f galaxy (see Fig. 2). If our findings are meaningful, they should not only be sensibly segregated in the 5D input parameter space, but also probe different evolutionary phases of the galaxy, as revealed by the simulation’s meta-parameters.

We begin with an analysis of the input parameters. Figure 11 shows the distributions of the input parameters (Toomre diagram,  $[\alpha/\text{Fe}]$  vs  $[\text{Fe}/\text{H}]$ , and stellar ages vs  $[\text{Fe}/\text{H}]$ ) for all second run subcomponents of samples A, B, C and D. For a more detailed view, we provide Figs. C.3, C.4, and C.5 in the Appendix, respectively comparing the kinematic, chemical, and stellar age parameters for each sample. We note the tendency for the MC1 subcomponents to assume a narrow prograde distribution, in contrast to the wider kinematic distributions of the MC2 subcomponents, which include all retrograde stars. In chemical space, MC1 subcomponents assume a tight  $[\alpha/\text{Fe}]$ – $[\text{Fe}/\text{H}]$  correlation at the high-metallicity end, whereas the MC2 subcomponents span a wide range in  $[\alpha/\text{Fe}]$  at low metallicity. The  $\tau$ – $[\text{Fe}/\text{H}]$  distributions reveal what appear to be two distinct, parallel chemical evolutionary tracks, with the MC1 subcomponents consistently occupying the younger portion of the higher- $[\text{Fe}/\text{H}]$  track. These features are consistent with a distinction between the disk and halo as observed in the Milky Way, where the disk is comprised of young, metal-rich stars with a tight velocity dispersion around  $v \approx v_{\text{LSR}}$ , contrasting with the old, metal-poor stars of the halo, which are further characterized by a large velocity dispersion that encompasses stars with retrograde orbits (Helmi 2008; Bland-Hawthorn & Gerhard 2016; Gaia Collaboration et al. 2018b). The nuances then arise in the interpretation of each subcomponent. Since we have different results for each sample, we will go through them one by one.

For sample A, the subcomponents MC1a and MC1b appear to correspond to two chemically distinct disk populations, in line with the chemical dichotomy shown in Fig. 1. MC1a is young and occupies the lower- $[\alpha/\text{Fe}]$  track, in contrast to MC1b’s older age and higher  $[\alpha/\text{Fe}]$ . As for MC2, the subcomponent MC2a is the most populous, displaying a large velocity dispersion and the highest  $[\alpha/\text{Fe}]$ , and it further contains the oldest stars in the sample. In contrast, MC2b and MC2c both display predominantly large  $\sqrt{U^2 + W^2}$  velocities, with MC2c in particular being entirely prograde. MC2b and MC2c also display lower  $[\alpha/\text{Fe}]$  compared to MC2a, while following a lower- $[\text{Fe}/\text{H}]$  chemical evolutionary track. Features in the  $[\alpha/\text{Fe}]$ – $[\text{Fe}/\text{H}]$  plane are explained by differences in the rates and yields of the dominant nucleosynthesis processes, including type Ia SNe, CC-SNe, and stellar winds from asymptotic giant-branch stars. These, in turn, are dependent on the mass of the progenitor dark-matter halo, which influences star formation rates. The trends found here for



**Fig. 8.** 2D UMAP latent space embeddings for the major components MC1 (top row) and MC2 (bottom row) in each of the samples. Red indicates ex situ stars.



**Fig. 9.** Trustworthiness score,  $T$ , as a function of the number of nearest neighbors,  $k$ , for the UMAP embeddings obtained from samples A, B, C and D. Circles represent the first run embeddings, plus signs the second run MC1 embeddings and crosses the second run MC2 embeddings. All values satisfy  $T(k) > 0.90$ .

subcomponents MC2b and MC2c, compared to MC2a, are reminiscent of the observed trends resulting from the *Gaia*-Enceladus merger (Helmi et al. 2018). Indeed, the chemical evolutionary history of m12f appears to be explained by the accretion of at least one lower-mass galaxy (see, e.g., Vincenzo et al. 2019). The combined chrono-chemo-kinematic analysis strongly suggests that MC2a represents the primordial stellar halo, generated in situ, whereas the subcomponents MC2b and MC2c represent accreted populations.

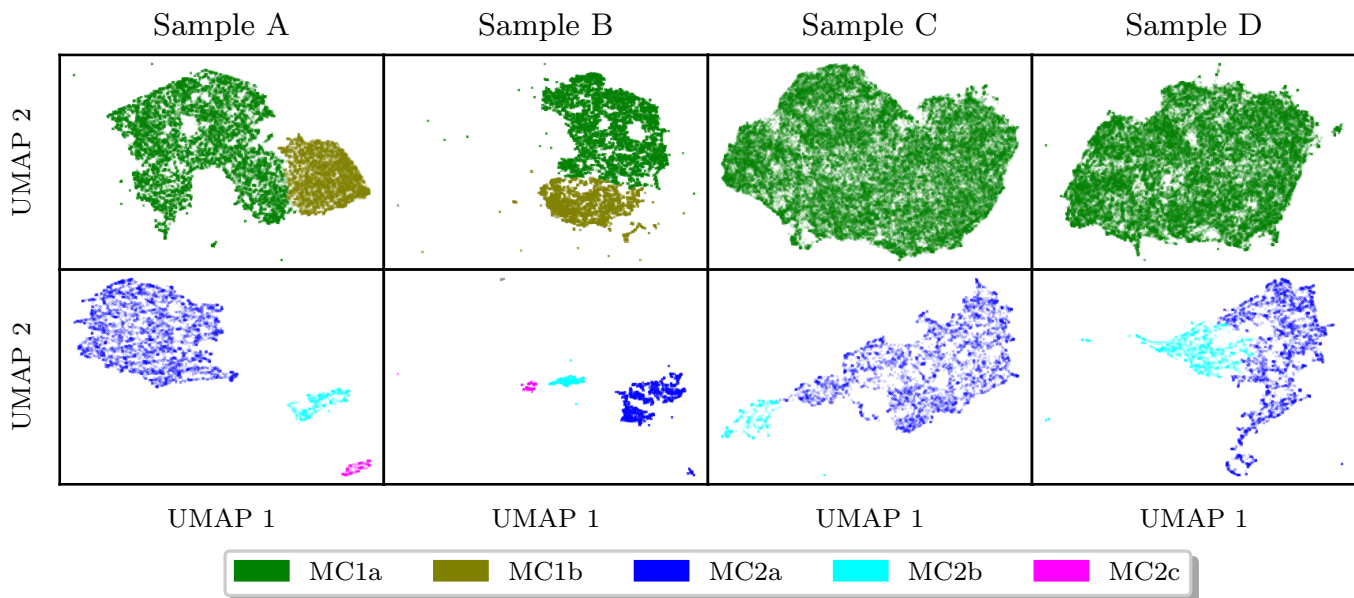
For sample B, the MC2a, MC2b, and MC2c subcomponents are identical to those found for sample A. Whereas for sample A, MC1a and MC1b respectively corresponded to the young and old disk populations, the reverse<sup>1</sup> is found for sample B. These subcomponents appear to represent the high- $\alpha$  and low- $\alpha$  disks,

<sup>1</sup> The labeling is ordered as a function of the number of stars in each subcomponent. For samples A and B, the subcomponents MC1a and MC1b have their labels switched, but otherwise correspond to the same populations.

similar to the MC1 subcomponents found for sample A. However, the partitioning of the disk was not possible for samples C and D. This is explained by the fact that samples C and D contain perturbed input parameters, leading to an inability to differentiate between the high- $\alpha$  and low- $\alpha$  disk subcomponents. As a result, the ability to distinguish the MC2 subcomponents for samples C and D is also affected. For sample C, the subcomponent MC2a appears to encompass the in situ halo and a significant portion of the accreted stars, while MC2b encompasses the remaining accreted stars. For sample D, we find that MC2a captures all accreted stars as well as the younger portion of the in situ halo, while MC2b captures the older portion of the halo.

To provide a fuller picture, we plot  $d_{\text{form}}$  as a function of age in Fig. 12. Let us begin by considering sample A. Focusing on the accreted stars ( $d_{\text{form}} > 30$  kpc), we observe that the subcomponents MC2b and MC2c trace the two major accretion events (cf. Fig. 2), although with some degree of contamination. Shifting our attention to the in situ stars ( $d_{\text{form}} < 30$  kpc), we identify the bulk of MC2a, which we associate with the primordial stellar halo. Leftward of the 8 Gyr mark, we notice the first stars cataloged as belonging to MC1b, with the other disk component (MC1a) emerging only in the past few Gyr. Interestingly, the two major accretion events traced by MC2b and MC2c are associated with bursts of in situ star formation, at  $\tau = 6.5$  and 1.4 Gyr, respectively (see Fig. 13). The origin of the low- $\alpha$  disk subcomponent, identified as MC1a, can be explained by the inflow of metal-poor gas brought into the disk by the second major accretion event (MC2c), which is mostly characterized by ages younger than  $\tau \approx 1.4$  Gyr.

The subcomponents found for sample B are largely analogous to the ones found for sample A, with MC2a, MC2b and MC2c representing the exact same populations. For MC1a and MC1b, however, the situation is not as straightforward. Firstly, the labels are switched compared to sample A, since now the older subcomponent (MC1a) contains more stars. Secondly, as shown in Fig. 13, the bulk of the younger subcomponent (MC1b) contains stars older than the time of the second major accretion. This means that, although this subcomponent contains the low- $\alpha$  disk, it also contains a portion of the high- $\alpha$  disk. Thus, there is an added difficulty in adequately segregating the two disk



**Fig. 10.** 2D UMAP latent space embeddings for the major components MC1 (top row) and MC2 (bottom row) in each of the samples, color-coded according to the identified subcomponents. **Top:** MC1, subdivided into MC1a (green) and MC1b (olive). **Bottom:** MC2, subdivided into MC2a (blue), MC2b (cyan) and MC2c (magenta).

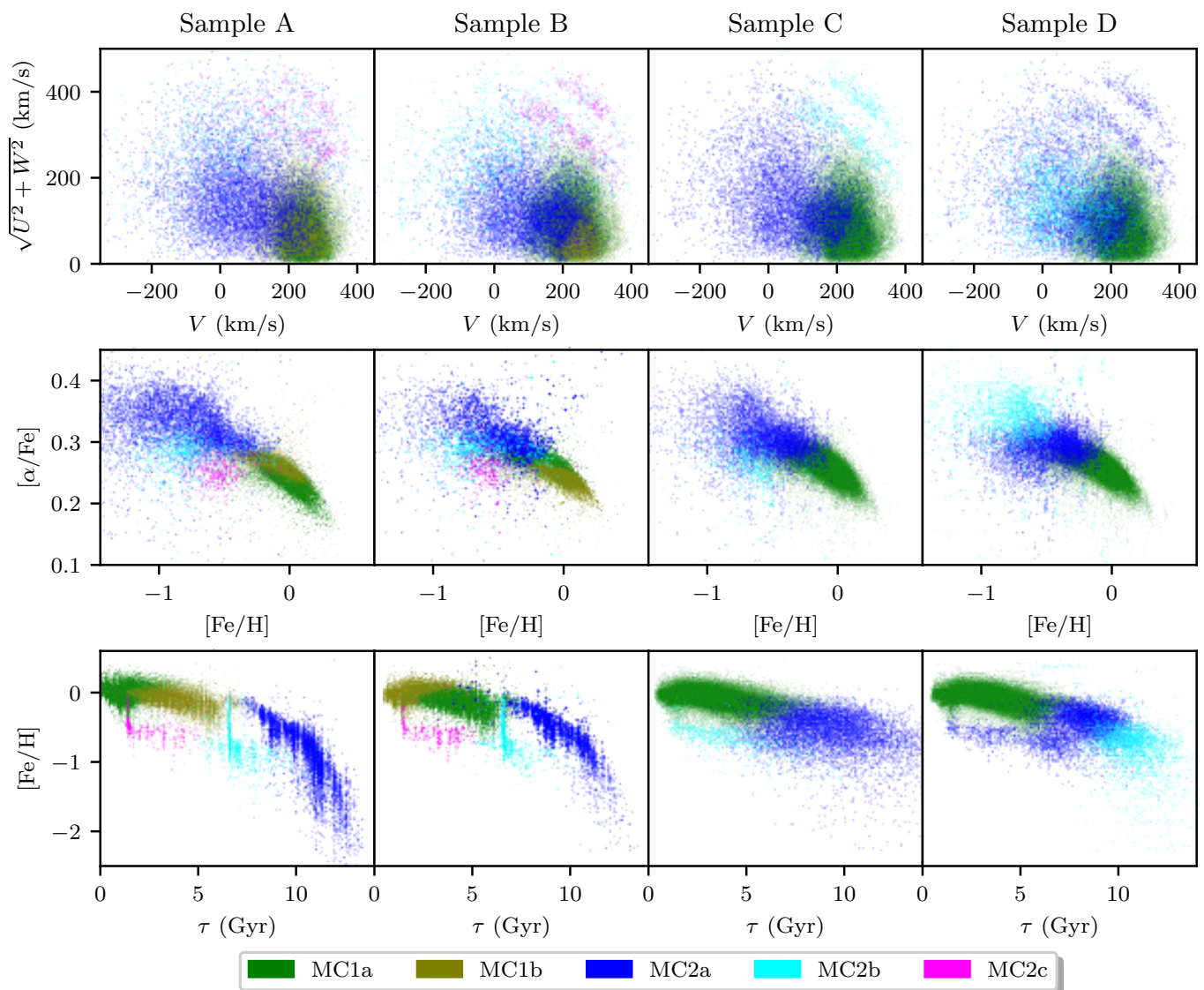
subcomponents, likely due to the undersampling of very young stars relative to sample A, which, in turn, is rooted in the selection of seismic red giants when building sample B (see Fig. 5). For sample C, there is no segregation of MC1 into subcomponents, whereas MC2 is segregated into two, rather than three, subcomponents. The subcomponent MC2a captures the in situ halo as well as the stars from the first accreted satellite galaxy, with MC2b representing the stars belonging to the second accreted galaxy. For sample D, MC1 is also not segregated into subcomponents, and the two subcomponents found for MC2 differ from those found in sample C. MC2a now corresponds to the full accreted population as well as the younger portion of the in situ halo, while MC2b corresponds to the older portion of the in situ halo. The detection of fewer subcomponents for samples C and D is explained by their perturbed input parameters, which blur the fine details that could have been used to analyze these subcomponents. The segregation of the low- $\alpha$  and high- $\alpha$  disks then becomes unfeasible because their separation in  $[\alpha/\text{Fe}]$  is of the order of the adopted uncertainties. The perturbations to the chemical parameters, along with the perturbations to the stellar age, also hinder the ability to adequately segregate the halo subpopulations. These effects are evident when comparing sample B with samples C and D, as shown in Fig. 11. Our interpretation of each identified subcomponent across the four samples is provided in Table 3.

## 5. Summary and conclusions

In this work, we used a cosmological zoom-in galaxy for with a known evolutionary history and generated four synthetic stellar samples (A, B, C, and D): sample A consists of 75,000 randomly selected stars in all evolutionary states, sample B consists of a TESS-like seismic red-giant selection containing 76,168 stars, and samples C and D are perturbed versions of sample B. Considering a 5D input parameter space that includes stellar ages, kinematics ( $V$  and  $\sqrt{U^2 + W^2}$ ) and chemical abundances ( $[\text{Fe}/\text{H}]$  and  $[\alpha/\text{Fe}]$ ), we applied the unsupervised mani-

fold learning algorithm UMAP to generate latent space embeddings. The HDBSCAN clustering algorithm was then applied multiple times to obtain the hyperparameters that maximized the persistence score, thereby retrieving the most persistent clusters. All embeddings were validated based on the trustworthiness score,  $T(k)$ , which consistently returned values satisfying  $T(k) > 0.90$ . Our findings are as follows:

1. A first run of the method demonstrated a robust separation of the disk and halo for all samples (see Fig. 7), confirmed by our test of the VM criterion, which returned informedness values greater than 0.95 (see Table 2). This shows that the inclusion of realistic uncertainties on the input parameters (i.e., sample C) allows the segregation of these two components using a purely automated and unsupervised approach. Moreover, we demonstrate that meaningful information can be inferred from the topology of the embeddings, even in the absence of purely fully structures.
2. We subsequently conducted a second run of the method on both the disk-like and halo-like components, identifying five subcomponents for the pristine samples A and B and three subcomponents for the perturbed samples C and D. For samples A and B, we were able to identify the high- $\alpha$  and low- $\alpha$  disks. The latter was generated from the inflow of chemically impoverished gas associated with the second major accretion event ( $\tau \approx 1.4$  Gyr; see Figs. 1 and 2). Since this population was formed relatively recently, an undersampling of the youngest stars affects the ability to resolve it. This effect is shown in Fig. 13, where the low- $\alpha$  disk is still able to be identified for sample B (MC1b), but at the expense of accuracy, being grouped together with in situ stars that precede the accretion event at  $\tau \approx 1.40$  Gyr. In the case of samples C and D, we were unable to differentiate between the two disk components. We attribute this to the very tight  $[\alpha/\text{Fe}]$  bimodality in the synthetic galaxy, which is eventually blurred as a result of the perturbation to the chemical abundances.
3. The halo-like component of samples A and B was segregated into three subcomponents: the in situ halo (MC2a), the first



**Fig. 11.** Distributions of the input parameters for samples A, B, C and D, color-coded according to the second run subcomponents. **Top:** Toomre diagram. **Middle:**  $[\alpha/\text{Fe}]$  versus  $[\text{Fe}/\text{H}]$ . **Bottom:** Stellar ages versus  $[\text{Fe}/\text{H}]$ .

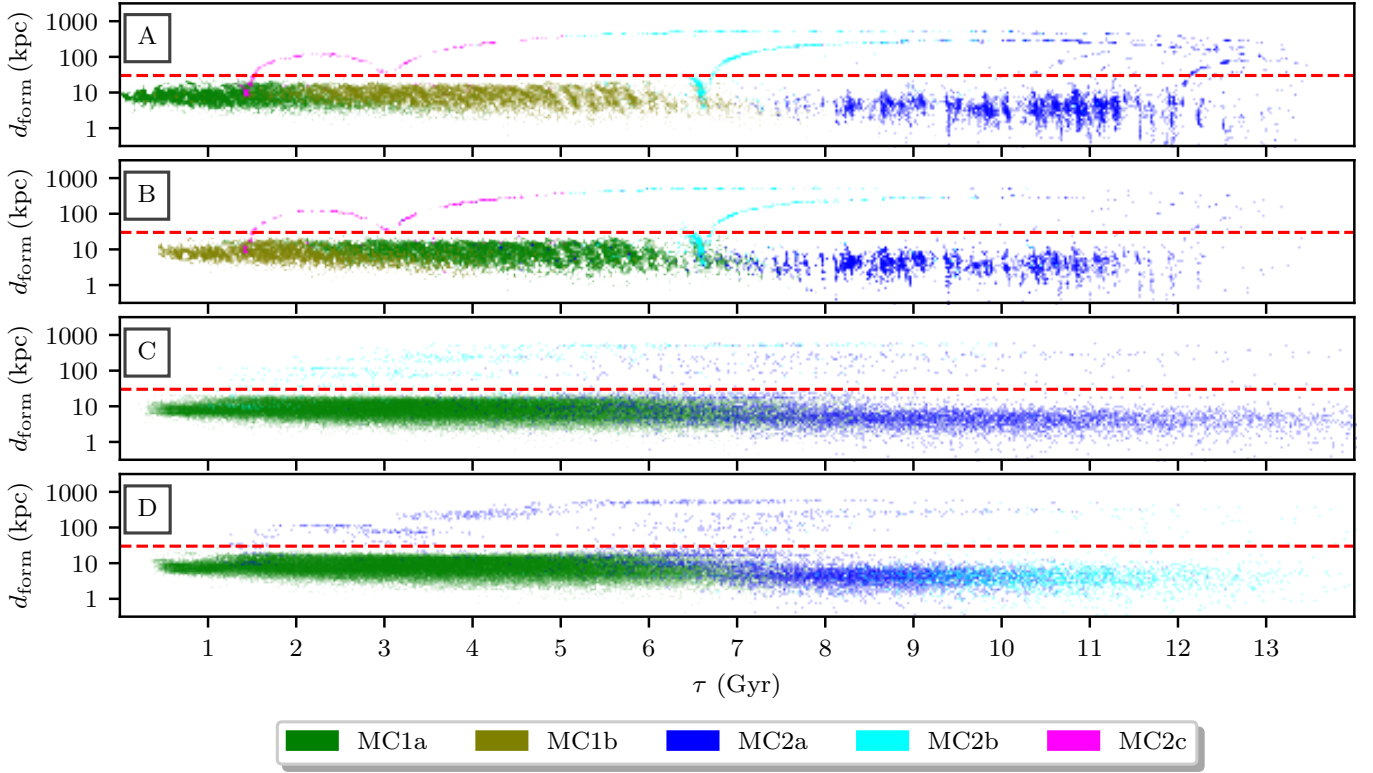
**Table 3.** Interpretation of each identified subcomponent across samples A, B, C and D.

	Sample A	Sample B	Sample C	Sample D
MC1a	Low- $\alpha$ disk	High- $\alpha$ disk	N/A	N/A
MC1b	High- $\alpha$ disk	Low- $\alpha$ disk	N/A	N/A
MC2a	in situ halo	in situ halo	in situ halo + first merger	Mergers + younger in situ halo
MC2b	First merger	First merger	Second merger	Older in situ halo
MC2c	Second merger	Second merger	N/A	N/A

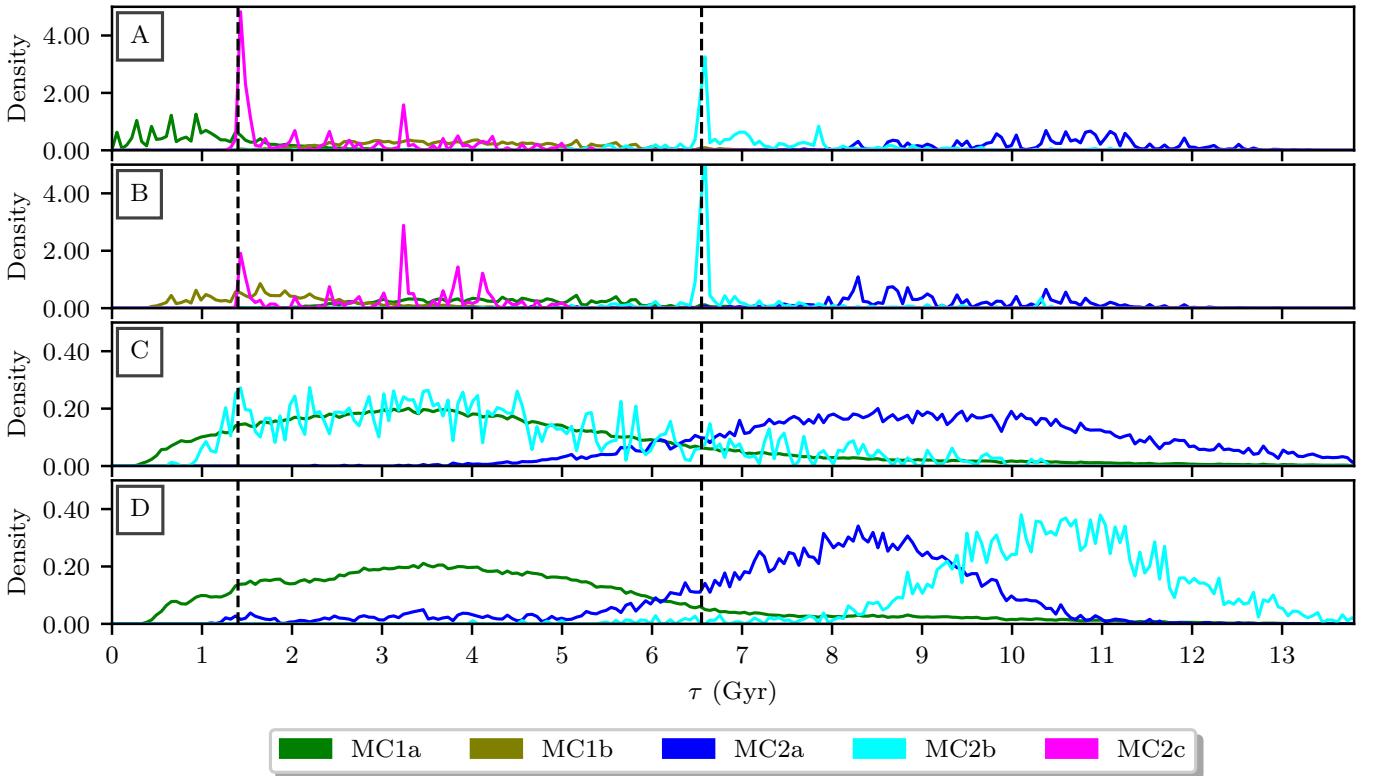
accreted population (MC2b), and the second accreted population (MC2c). For sample C, the second accreted population was segregated from the remainder of the halo. This loss in resolution is attributed to the perturbation of the input parameters. Similar or better results were expected for sample D, as it was perturbed using more optimistic uncertainties. However, sample D segregated the oldest halo stars ( $\tau > 10$  Gyr)

from the remainder of the halo, which includes both accreted populations and a sizable portion of the in situ halo.

We have shown that the application of the unsupervised manifold learning algorithm UMAP, along with an automated HDBSCAN optimization routine, can provide key inferences on the nature and number of stellar populations without any such prior assumptions. This was achieved by intelligently projecting the high-dimensional input data into a lower-dimension, la-



**Fig. 12.** Distance of formation ( $d_{\text{form}}$ ) as a function of age for samples A, B, C and D, color-coded according to the second run subcomponents. The horizontal dashed line marks  $d_{\text{form}} = 30$  kpc.



**Fig. 13.** Kernel density estimation of the stellar ages for samples A, B, C and D, color-coded according to the second run subcomponents. Vertical dashed lines at  $\tau = 1.4$  and  $6.5$  Gyr mark the occurrence of each major accretion event.

tent space embedding. Various validation procedures show this to be a valid method to reliably and accurately depict the underlying non-linear and non-parametric structure hidden in the high-dimensional data. We have also shown that, by segregating the stars according to the embedded manifold structure, we can divide disk-like from halo-like structures for all four of our samples, including a TESS-like sky coverage of a red-giant sample with realistic *Gaia* DR3 kinematic, APOGEE DR17 chemical and seismic age uncertainties. We compared UMAP to other algorithms, highlighting the limitations of PCA, and showed that UMAP is capable of competing with, and even surpassing, t-SNE in terms of computational efficiency and its ability to discern stellar populations. These comparisons are provided in Appendix D.

We have also shown that, when adopting unperturbed (i.e., pristine) input parameters, the method can segregate subcomponents belonging to the disk and the halo. Although these results are not replicated for the perturbed samples, we emphasize that we made use of only five stellar parameters, these being  $\{V, UW, [\text{Fe}/\text{H}], [\alpha/\text{Fe}], \tau\}$ . However, we identify potential in the method to discern additional stellar populations, such as in situ vs ex situ stars, when dealing with realistic data sets. UMAP is expected to provide deeper insights into stellar populations through the inclusion of additional stellar parameters. Examples include angular momentum, potential energy, integrals of motion, orbital eccentricity, and other chemical abundance ratios, such as s-process elements ejected by the stellar winds of asymptotic giant-branch stars and r-process elements produced in neutron star mergers.

## Data availability

The Appendix is provided through the following URL: <https://doi.org/10.5281/zenodo.14749285>.

*Acknowledgements.* This work was supported by Fundação para a Ciência e a Tecnologia (FCT) through the research grants UIDB/04434/2020 (DOI: 10.54499/UIDB/04434/2020) and UIDP/04434/2020 (DOI: 10.54499/UIDP/04434/2020). T.L.C. is supported by FCT in the form of a work contract (CEECIND/00476/2018). D.B. acknowledges funding support by the Italian Ministerial Grant PRIN 2022, “Radiative opacities for astrophysical applications”, no. 2022NEXMP8, CUP C53D23001220006. A.M. acknowledges support from the European Research Council Consolidator Grant funding scheme (project ASTEROCHRONOMETRY, G.A. no. 772293, <http://www.asterochronometry.eu>).

## References

Abdurro’uf, Accetta, K., Aerts, C., et al. 2022, *ApJS*, 259, 35 1  
 Adibekyan, V. Z., Sousa, S. G., Santos, N. C., et al. 2012, *A&A*, 545, A32 1  
 Anders, F., Chiappini, C., Rodrigues, T. S., et al. 2017, *A&A*, 597, A30 1  
 Arcones, A. & Thielemann, F.-K. 2023, *A&A Rev.*, 31, 1 2.1  
 Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, *ApJ*, 935, 167 2.1  
 Baglin, A., Auvergne, M., Barge, P., et al. 2009, in *Transiting Planets*, ed. F. Pont, D. Sasselov, & M. J. Holman, Vol. 253, 71–81 1  
 Bensby, T., Feltzing, S., & Lundström, I. 2003, *A&A*, 410, 527 1  
 Bland-Hawthorn, J. & Gerhard, O. 2016, *ARA&A*, 54, 529 1, 2.1, 2.1, 4.3  
 Borucki, W. J., Koch, D., Basri, G., et al. 2010, *Science*, 327, 977 1  
 Buder, S., Sharma, S., Kos, J., et al. 2021, *MNRAS*, 506, 150 1  
 Campante, T. L., Li, T., Ong, J. M. J., et al. 2023, *AJ*, 165, 214 1  
 Campello, R. J. G. B., Moulavi, D., & Sander, J. 2013, in *Advances in Knowledge Discovery and Data Mining, Lecture notes in computer science* (Berlin, Heidelberg: Springer Berlin Heidelberg), 160–172 1, 3  
 Chaplin, W. J., Serenelli, A. M., Miglio, A., et al. 2020, *Nature Astronomy*, 4, 382 1  
 Ciucă, I., Kawata, D., Ting, Y.-S., et al. 2024, *MNRAS*, 528, L122 2.1  
 Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, *Research in Astronomy and Astrophysics*, 12, 1197 1

Dalton, G., Trager, S., Abrams, D. C., et al. 2014, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 9147, Ground-based and Airborne Instrumentation for Astronomy V, ed. S. K. Ramsay, I. S. McLean, & H. Takami, 91470L 1  
 Davies, G. R. & Miglio, A. 2016, *Astronomische Nachrichten*, 337, 774 3.4, 5  
 Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, *ApJ*, 292, 371 1  
 de Jong, R. S., Agertz, O., Berbel, A. A., et al. 2019, *The Messenger*, 175, 3 1  
 Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. 1996, in *Knowledge Discovery and Data Mining* 1, 3  
 Feldmann, R., Quataert, E., Faucher-Giguère, C.-A., et al. 2023, *MNRAS*, 522, 3831 1  
 Font, A. S., McCarthy, I. G., Poole-Mckenzie, R., et al. 2020, *MNRAS*, 498, 1765 1  
 Francis, P. J. & Wills, B. J. 1999, in *Astronomical Society of the Pacific Conference Series*, Vol. 162, Quasars and Cosmology, ed. G. Ferland & J. Baldwin, 363 3  
 Freeman, K. & Bland-Hawthorn, J. 2002, *ARA&A*, 40, 487 1  
 Fuhrmann, K. 1998, *A&A*, 338, 161 1  
 Gaia Collaboration, Babusiaux, C., van Leeuwen, F., et al. 2018a, *A&A*, 616, A10 2.2  
 Gaia Collaboration, Katz, D., Antoja, T., et al. 2018b, *A&A*, 616, A11 4.3  
 Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023, *A&A*, 674, A1 1  
 García, R. A. & Ballot, J. 2019, *Living Reviews in Solar Physics*, 16, 4 1  
 Gilmore, G. & Reid, N. 1983, *MNRAS*, 202, 1025 1  
 Giribaldi, R. E. & Smiljanic, R. 2023, *A&A*, 673, A18 2.1  
 Grand, R. J. J., Fragakoudi, F., Gómez, F. A., et al. 2024, *Monthly Notices of the Royal Astronomical Society*, 532, 1814 1  
 Helmi, A. 2008, *A&A Rev.*, 15, 145 1, 4.3  
 Helmi, A., Babusiaux, C., Koppelman, H. H., et al. 2018, *Nature*, 563, 85 1, 4.3  
 Hon, M., Huber, D., Kuszlewicz, J. S., et al. 2021, *ApJ*, 919, 131 2.2  
 Hopkins, P. F. 2015, *MNRAS*, 450, 53 2.1  
 Hopkins, P. F., Kereš, D., Oñorbe, J., et al. 2014, *MNRAS*, 445, 581 2.1  
 Hopkins, P. F., Wetzel, A., Kereš, D., et al. 2018, *MNRAS*, 480, 800 1, 2.1  
 Howell, S. B., Sobek, C., Haas, M., et al. 2014, *PASP*, 126, 398 1  
 Ibata, R. A., Gilmore, G., & Irwin, M. J. 1994, *Nature*, 370, 194 1  
 Ikkotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. 2023, *Information Sciences*, 622, 178 3  
 Jin, S., Trager, S. C., Dalton, G. B., et al. 2024, *MNRAS*, 530, 2688 1  
 Kippenhahn, R., Weigert, A., & Weiss, A. 2013, *Stellar Structure and Evolution* (Springer Berlin, Heidelberg) 3.4, 5  
 Lindegren, L., Klioner, S. A., Hernández, J., et al. 2021, *A&A*, 649, A2 1  
 Mackereth, J. T., Miglio, A., Elsworth, Y., et al. 2021, *MNRAS*, 502, 1947 1  
 MacQueen, J. 1967, in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1st edn., ed. L. M. L. C. J. Neyman, University of California Press, Berkeley, CA, 281–297 3  
 Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, *AJ*, 154, 94 1  
 Matteucci, F. 2021, *A&A Rev.*, 29, 5 1  
 McInnes, L., Healy, J., & Melville, J. 2018, *arXiv e-prints*, arXiv:1802.03426 1, 3.1  
 Metcalfe, T. S., Creevey, O. L., Doğan, G., et al. 2014, *ApJS*, 214, 27 1  
 Miglio, A., Chiappini, C., Mackereth, J. T., et al. 2021, *A&A*, 645, A85 1  
 Miglio, A., Chiappini, C., Mosser, B., et al. 2017, *Astronomische Nachrichten*, 338, 644 1  
 Montalbán, J., Mackereth, J. T., Miglio, A., et al. 2021, *Nature Astronomy*, 5, 640 1  
 Necib, L., Ostdiek, B., Lisanti, M., et al. 2020, *ApJ*, 903, 25 1  
 Nguyen, T., Ou, X., Panithanpaisal, N., et al. 2024, *ApJ*, 966, 108 2.1, 2.1  
 Nikakhtar, F., Sanderson, R. E., Wetzel, A., et al. 2021, *ApJ*, 921, 106 1, 2.1  
 Noguchi, M. 2018, *Nature*, 559, 585 1  
 Ostdiek, B., Necib, L., Cohen, T., et al. 2020, *A&A*, 636, A75 1, 3.4  
 Pearson, K. 1901, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2, 559 3  
 Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *Journal of machine learning research*, 12, 2825 3.1  
 Queiroz, A. B. A., Anders, F., Chiappini, C., et al. 2023, *A&A*, 673, A155 1  
 Rauer, H., Catala, C., Aerts, C., et al. 2014, *Experimental Astronomy*, 38, 249 1  
 Reddy, B. E., Lambert, D. L., & Allende Prieto, C. 2006, *MNRAS*, 367, 1329 1  
 Reynolds, D. 2009, *Gaussian Mixture Models*, ed. S. Z. Li & A. Jain (Boston, MA: Springer US), 659–663 3  
 Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2015, *Journal of Astronomical Telescopes, Instruments, and Systems*, 1, 014003 1  
 Sanderson, R. E., Wetzel, A., Loebman, S., et al. 2020, *ApJS*, 246, 6 2.1  
 Sante, A., Font, A. S., Ortega-Martorell, S., Olier, I., & McCarthy, I. G. 2024, *Monthly Notices of the Royal Astronomical Society*, 531, 4363 1  
 Shen, J. & Li, Z.-Y. 2016, in *Astrophysics and Space Science Library*, Vol. 418, Galactic Bulges, ed. E. Laurikainen, R. Peletier, & D. Gadotti, 233 1  
 Silva Aguirre, V., Bojsen-Hansen, M., Slumstrup, D., et al. 2018, *MNRAS*, 475, 5487 1  
 Silva Aguirre, V., Stello, D., Stokholm, A., et al. 2020, *ApJ*, 889, L34 1

- Spitoni, E., Recio-Blanco, A., de Laverny, P., et al. 2023, *A&A*, 670, A109 [1](#)
- Spitoni, E., Silva Aguirre, V., Matteucci, F., Calura, F., & Grisoni, V. 2019, *A&A*, 623, A60 [1](#)
- Stello, D., Bruntt, H., Preston, H., & Buzasi, D. 2008, *ApJ*, 674, L53 [2.2](#)
- Stello, D., Saunders, N., Grunblatt, S., et al. 2022, *MNRAS*, 512, 1677 [1](#)
- Stokholm, A., Aguirre Børsen-Koch, V., Stello, D., Hon, M., & Reyes, C. 2023, *MNRAS*, 524, 1634 [1](#)
- van de Voort, F., Quataert, E., Hopkins, P. F., Kereš, D., & Faucher-Giguère, C.-A. 2015, *MNRAS*, 447, 140 [2.1](#)
- van der Maaten, L. J. P. 2013, in *Proceedings of the International Conference on Learning Representations (ICLR)* [1](#), [3](#)
- Vasiliev, E. & Belokurov, V. 2020, *MNRAS*, 497, 4162 [1](#)
- Venna, J. & Kaski, S. 2001, in *Artificial Neural Networks — ICANN 2001*, ed. G. Dorffner, H. Bischof, & K. Hornik (Berlin, Heidelberg: Springer Berlin Heidelberg), 485–491 [3.2](#)
- Vincenzo, F., Spitoni, E., Calura, F., et al. 2019, *MNRAS*, 487, L47 [2.1](#), [4.3](#)
- Vincenzo, F., Weinberg, D. H., Miglio, A., Lane, R. R., & Roman-Lopes, A. 2021, *MNRAS*, 508, 5903 [2.1](#)
- Wetzel, A., Hayward, C. C., Sanderson, R. E., et al. 2023, *ApJS*, 265, 44 [2.1](#)
- Wetzel, A. R., Hopkins, P. F., Kim, J.-h., et al. 2016, *ApJ*, 827, L23 [2.1](#)